International Journal of Advanced Technology in Engineering and Science www.ijates.com

Volume No.03, Special Issue No. 02, February 2015 ISSN (online): 2348 – 7550

A SURVEY ON CLOUD RESOURCE ALLOCATION STRATEGIES

K.Padmaveni[1], E.R. Naganathan[2]

[1]*Asst. Professor, CSE, Hindustan University, (India)*

[2]*Professor, CSE, Hindustan University, (India)*

ABSTRACT

Cloud computing delivers an efficient and elastic means for various services to meet the diverse and intensifying needs of users. It is possible for anyone to use the services from anywhere in the world. This way of outsourcing will decrease the infrastructure cost. This also supports for using product without difficulty of getting license. The resource allocation also has to be optimized for achieving maximum utilization of this technology. In this paper, various resource allocation strategies and their challenges are discussed. This would benefit both the researchers and cloud users to face challenges.

Index term: Resource Allocation Strategies

I INTRODUCTION

Cloud computing refers to both the applications delivered over the Internet as well as the hardware and systems software in the data centers. Cloud provides service in the form of virtualized hardware and software to the user. Cloud also includes the major functionalities of technologies like grid computing, utility computing and virtualization [1]. Cloud clients deploy their online applications on virtual machines(VMs) in cloud without violating the Service level agreement(SLA). Virtualization technology in data centers enables better server utilization and hence reduces the resource management task and cost of cloud maintenance It is the technology that maps the virtual machines to the Physical Machines (PMs). This mapping enables the PM to serve many users with different functional requirement (like applications) and non-functional requirements (like memory size etc.,).

Now a days the number of VMs deployed in cloud data center increases every day. The capacity of VM can be provisioned either statically or dynamically. In static provisioning, the average resource utilizations from data so far and the user defined requirements are used as input to the algorithm that will give a mapping of VMs to PMs. This needs the VM resource demand to be known in advance and the variability. In dynamic provisioning, it is possible to adapt to the varying workload and keep the load balanced

II DIFFERENT TECHNOLOGIES

2.1 Ant colony optimization algorithm for resource allocation

The cloud computing uses master/slave structure where the master node controls the slave node as in fig 1.

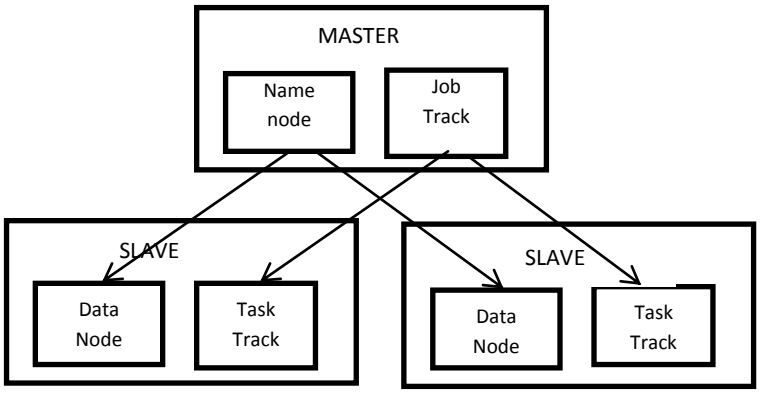

Fig 1:The structure of Hadoop cluster

The concept of ACO was stated by M.Dorigo et al,. The ACO shows the path of reaching the solution by reducing the search space in the following solution space.[2] They achieve this by the feedback mechanism called pheromone update. When the search begins, the query will be sent to the slave nodes. The slave nodes are considered as ants in behavior. The slave node that has more pheromone gets the larger probability of heading towards solution. Hence its path will be selected. Hua's paper[3] gives a detailed description of this.

2.2 Threshold based scheduling algorithm

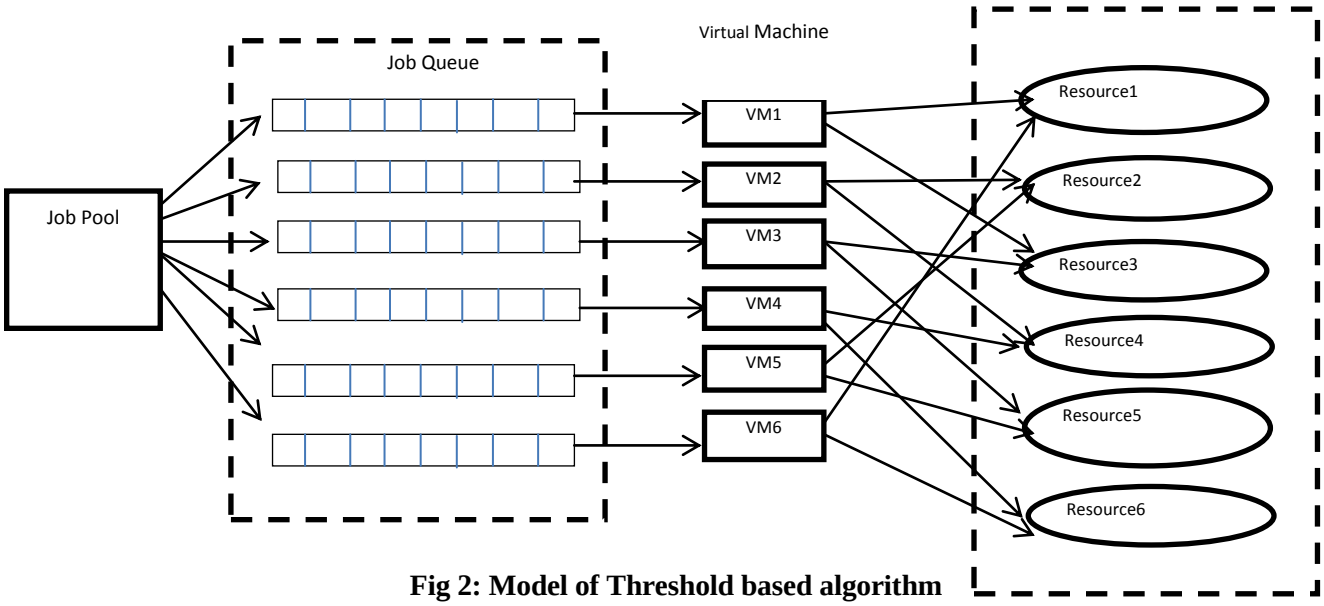

Fig 2: Model of Threshold based algorithm

2.3 Optimized Genetic Algorithm

The genetic algorithm was proposed by Holand. The reduction in search space and reaching the global optima without falling in local optima are the major features of Genetic algorithm. The algorithm is based on the genetic operators namely reproduction, crossover and mutation. The GA with improved performance works with dual fitness function. The flow of GA is shown in fig 3.

In GA, a solution is considered as a chromosome. The numbers that are part of each chromosome are the resource number allocated.Li's paper[4] compares Adaprive GA and dual fitness GA. It is stated that DFGA is superior to AGA.

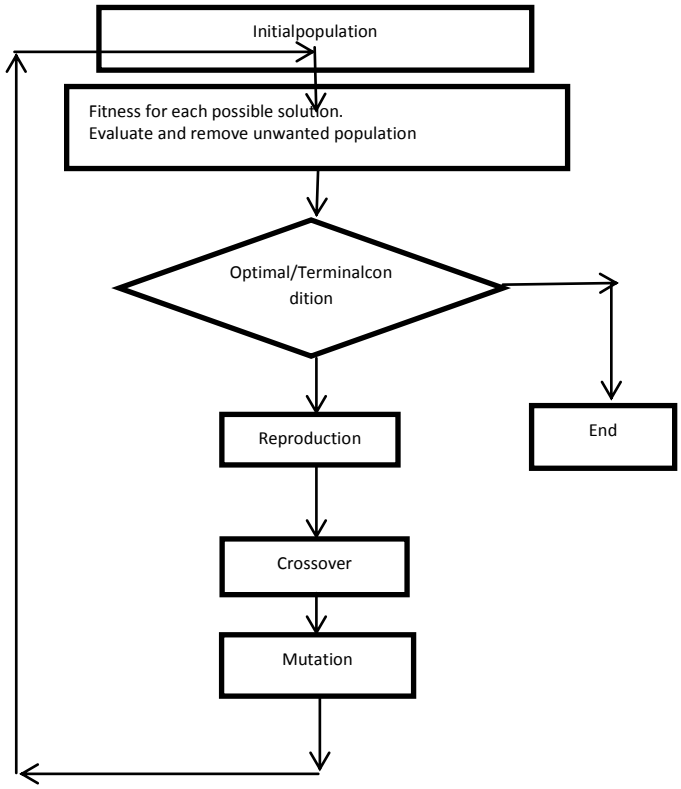

Fig 3: Flow of GA

2.4 Cloud Scheduling via migration of virtual machines

Wubin Li et al., have proposed a model for dynamic cloud scheduling via migration of virtual machines. They have investigated dynamic cloud scheduling use cases and proposed a linear integer programming model fir dynamic cloud scheduling via VM migration across multiple clouds. This model can be applied to handle both the infrastructure as well as services. They have not taken SLA violation into consideration.[5]

2.5 A VMCTune located in the cluster layer

Wenyu Zohu et al., have designed and implemented a load balancing schedule based on dynamic resource allocation policy for virtual machine cluster. Their algorithm monitors the real time resource utilization of Virtual machines and physical machines and instantly allocates resources virtual machines to the same physical machines in order to achieve load balancing. Then it tries for global load balancing . they have shown results in the improvement in utilization of physical machines as well.[6]

2.6 Generic Search

This is a schedule that is similar to the picking of balls and pins. If the goal is to use minimum number of bins, each bin us filled first. Once the bin is filled, the next bin is selected. Thus we can pack each bin to its capacity. [7].

If the goal is to use balance utilization across the set of bins, random allocation is done across given set of bins.

The balls are considered as the virtual machines and the bin is the physical machine. The bins are packed as complete binary tree with 2 power n nodes. This will have better packing.

This algorithm cannot determine any non existant node. Also there is no guarantee for minimal solution.[7]

2.7 Combined resource allocation

Magedanz et al.,have combined various cloud resource allocation, and has designed a multi cloud brokering mechanism for efficient utilization of cloud resources. According to them,the dynamic selection of optimal cloud service is more important. Different QoS constraints and dynamically changing workloads are considered. Multiple factors impacting multi cloud environment and their effects are evaluated. Also the metrics for assessing the performance of cloud systems for NGN media services are evaluated..[8]

2.8 Data Envelopment Analysis

Hsin-Hung Cho et al., have said that the overall cloud architecture has become imbalance due to the inclusion of various service classifications. They have proposed a Data Envelopment Analysis (DEA) to overcome the imbalance problem. Their analysis starts from the user's request and then to the DEA for evaluating the cloud parameters and thenfinding the resource allocation policy that is suitable for the users and the vendors.[9]

III CONCLUSION

This paper analyzes various resource allocation strategies under the cloud environment . Day by day new strategies are proposed for significant improve in the efficiency and use of resources. As the usage of cloud becomes wider and wider, the resource allocation is expected to further improve so that it improves the effectiveness in usage of cloud.

REFERENCES

[1]Stelios Sotiriadis, Nik Bessis, Fatos Xhafa, nick Antonopoulos "cloud virtual machine Scheduling" sixth international conference on complex, Intelligent and software Intensive ystem,2012, IEEE

[2]Li Huang, Hai-Shan Chen, ting-ting Hu "Survey on Resource Allocation Policy and job Scheduling algorithms of cloud computing", Journal of Software, Volume 8, No 2, Feb 2013 Academy Publisher

[3]Hua Xiayer, zheng Jun, Hu Wenxin, "Ant colony optimization algorithm for computing resource allocation based on cloud computing environment". Journal of east china normal university(Natural Science) 2010,

[4]Li Jian Feng, Peng Jian, "Task scheduling algorithm based on improved genetic algorithm in cloud computing" Journal of computer applications", 2011 pp 184-186

[5]Wubin Li, Johan Tordsson and Erik elmroth " Modeling for Dynamic cloud scheduling via migration of virtual machines Third IEEE International conference on cloud computing Technology and science, 2011

[6] Wenyu zohu, Shoubao yang, Jun Fang, Xianlong Niu, Hu song "VMCTune: A load balancing scheme for virtual machine cluster based on dynamic resource allocation 2010 ninth international conference on Grid and cloud computing

[7]Anshul Rai, Ranjitha Bagwan, Saikat Guha, Generalized resource allocation for the cloud SOCC 12, San Jose, CA, USA

[8]Magedanz, Schreiner,QOS aware multi-cloud brokering for NGN services:Tangible bebefits of elastic resource allocation mechanism. Communications and Electronics (ICCE), 2014 IEEE Fifth International Conference

[9]Hsin-Hung Cho, chi yuvan chen, Hao Chen Li, Shih TK .A fair cloud resource allocation using data envelopment analysis.Hetrogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference