Journal of Theoretical and Applied Information Technology
10th August 2014. Vol. 66 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN OPTIMIZED QOS BASED COST EFFECTIVE RESOURCE SCHEDULING IN CLOUD

S.P.JENO LOVESUM [1] DR.K.KRISHNAMOORTHY[2] BLESSED PRINCE. P[3]

[1]Assistant professor (SG), Department of Computer Science and Engineering, Karunya University, India

[2]Professor & Head Department of Computer Science and Engineering, Sudharsan College of Engineering, India

[3]Assistant professor (SG), Department of Information Technology, Karunya University, India

E-mail: [1]jenolovesum@gmail.com, [2]kkr_510@rediffmail.com, [3]blessedprince@gmail.com

ABSTRACT

Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers. Resource scheduling, which is a part of resource management is an important process that takes place in storage cloud which falls under IaaS cloud, so that the available resources may be properly allocated to the requesting tasks in a best fit manner so that no resources are wasted. As most of the scheduling techniques do not consider the minimum memory wastage in our work we have considered this factor as a QOS since most of the users of cloud resource are concerned about their cost and as memory usage plays a major role is calculating the cost we try to schedule the jobs among the available VMs so that minimum memory wastage is taken into account while allocating the VMs to complete the task. In this paper a QOS model for optimized task scheduling is used to minimize task completion time and task response time and minimum memory wastage. The task manager checks all the virtual machine and assigns the task to proper virtual machine which will have least memory wastage. Here the task is taken as storing a file in cloud storage. A QOS algorithm based on PSO is used to provide an optimal solution for resource scheduling. The simulation results show that the proposed method has the ability to find optimal trade-off solutions for task scheduling problems that represent the best possible compromises among the conflicting objectives.

Keywords:- *Cloud, Optimized, Scheduling, Storage Cloud, IAAS*

I. INTRODUCTION

In Cloud computing consumers are allowed to use the applications without installing and access their personal files at any computer with the help of internet. Cloud computing provides resources to users according to their demand The users request for available services according to their desired Quality of Service, and they are charged on a pay per use basis. One of the most challenging problems in Cloud computing is the workflow scheduling. The processing units in cloud environments are called as virtual machines. It should make sure that the tasks are not loaded heavily on one VM and some VMs do not remains idle and/or under loaded. Load balancing of non-pre-emptive independent tasks on virtual machines is an important aspect of task scheduling in clouds. Whenever certain VMs are overloaded and remaining VMs are under loaded with tasks for processing, the load has to be balanced to achieve optimal machine utilization. There are various algorithms designed for balancing the load among different tasks.

The main reason for scheduling is to optimize the usage of resources. One of the main factor that an user expects is that his task should be completed at the minimum cost for him at the same time his QOS also should not get disturbed. Considering this problem in this paper we have considered the QOS factors as task completion time, response time and minimum memory wastage.Most of the work in this area have concentrated on the completion time and response time but not much concentration is given to minimum memory wastage which is one of the important QOS that the user looks into when his task requires a memory usage. So in this paper we have considered the QOS minimum memory wastage as an additional QOS factor and we have carried out the scheduling. To optimize the method of scheduling we have used the PSO algorithm by incorporating the QOS factors of the user while finding the pbest and gbest values.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Local scheduling is used at the level of clusters, usually to balance load. Scheduling of tasks in cloud computing is an NP-hard optimization problem . the virtual machines should execute the tasks as early as possible and these VMs run in parallel. This leads to problems in scheduling of the customer tasks within the available resources. The scheduler should do the scheduling process efficiently in order to utilize the available resources fully. Different scheduling algorithms in are

- First Come First Serve Algorithm
- Round Robin algorithm
- Min–Min algorithm
- Max – Min algorithm

A proper scheduling policy attempts to assign these loads to available computing nodes so as to complete the processing of all loads in the shortest possible time. To improve the utilization of the processors, parallel computations require that processes be distributed to processors in such a way that the computational load is spread among the processors. Load balancing means shifting of tasks from one machine to another machine.

As discussed, load balancing is the process of evenly distributing the job on the computational resources of the cloud. cloud, being a dynamic system, often leaves vacuum after some period of time as there might be some incoming jobs and some exiting the system. Processing unit in the cloud environment is virtual machine Load balancing strategy also deals with the migration of the jobs periodically from one node to another in order to balance the workload amongst the cloud nodes. The objective is derived on the basis of the load variation in each node in computational cloud environment.

2. RELATED WORKS

Abirami S.P. and Shalini Ramanathan [1] a linear scheduling algorithm for efficient resource scheduling has been discussed. It schedules the resources among the requestors and maximize the resource utility. Shortest time first resource allocation is better than the first come first serve. The algorithm improves the resource utilization and the response time. But ist is not applicable for real time applications.

Amit Nathan,Sanjay [2] the author talks about dynamic planning of task scheduling to achieve maximum resource utilization is used. It uses four policies for resource scheduling. The advantage of this method is maximizing the rate of resource utilization. Disadvantages are requiring more preemption and it increase the overall overhead.

Bo Yin, Ying Wang, et.al [3] authors consider the issue of how to manage and arrange large-scale jobs submitted to cloud in order to optimize resource allocation and reduce cost. A multi dimensional scheduling algorithm is used. But in this performance bottleneck may occur and scheduling process of resource allocation becomes difficult.

Fetahi Wuhib, Rolf Stadler,Mike [4] Dynamic resource management gives the particular challenges in large-scale cloud environment. To obtain the efficient heuristic solution to the problem such as to minimize the adaption cost for resource allocation and resource utility using gossip protocol. Scalability and adaptability are considered. Demand Sharing is based on heuristic algorithm and it refers to sharing of memory demand of the process while demand exceeds the capacity of virtual machine.

G.Sireesha, L.Bharathi [5] the authors explains a Parallel data processing which is one of the major issue for Infrastructure-as-a-Service (IaaS) clouds. A new processing framework is designed for exploiting the dynamic resource allocation offered by IaaS clouds for both task scheduling and execution. Parallelization and Scheduling Strategies are used to construct an execution Graph

In [6] the hybrid resource management architecture to perform location aware VM placement and dynamic resource utilization management is used. This paper considers the utility function to find out which PM is appropriate for a new VM or migration, the provider evaluates each PM using a utility function.

In [7] The proposed model efficiently reallocates the resources. Job scheduling system plays a very important role in how to meet Cloud users job's QoS requirements and use the cloud resources efficiently in accost effective way.

Shikharesh Mujumdar [8], Match making and scheduling is used to represent the resource allocation in cloud computing. Match making is the method of allocating jobs associated with user requests to resources designated from the resource pool. Scheduling is used to determining the order in which jobs mapped to a selected resource that is to be executed.

After doing a comparative analysis in the literature survey, the algorithms and techniques

Journal of Theoretical and Applied Information Technology
10th August 2014. Vol. 66 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used reveals the different ways of task scheduling and load balancing approaches. But most of the algorithms and methods only consider throughput, completion time, response time, scalability, deadline but they do not include minimum memory wastage as QOS factor. This paper proposes the new efficient task scheduling system that uses PSO algorithm along with QOS to find the best fit resource to allocate the task so that wastage of memory space in the VM is reduced. As in our work we have focused on storage cloud which involves memory space we try to reduce the wastage of memory space in the VM by finding the best resource using the PSO algorithm along with QOS and allocating the task to that resource. Most algorithms in the literature survey, used for scheduling and load balancing considers make span, response time, throughput, deadline, priority as their performance metric or the QOS factors. But in this optimized method of scheduling we have considered the wastage of memory space as the QOS parameter for scheduling.

3. SYSTEM ARCHITECTURE

As in our work we have focused on storage cloud which involves memory space we try to reduce the wastage of memory space in the VM by finding the best resource using the PSO algorithm and allocating the task to that resource. Most algorithms used for scheduling and load balancing considers make span, response time, throughput, deadline, priority as their performance metric or the QOS factors. But in this optimized method of scheduling we have considered the wastage of memory space as the QOS parameter for scheduling. This scheduling method will assign to the virtual machine in best fit manner. i.e., task manager will check all the virtual machine and assigns the task to proper virtual machine which will have least memory wastage.

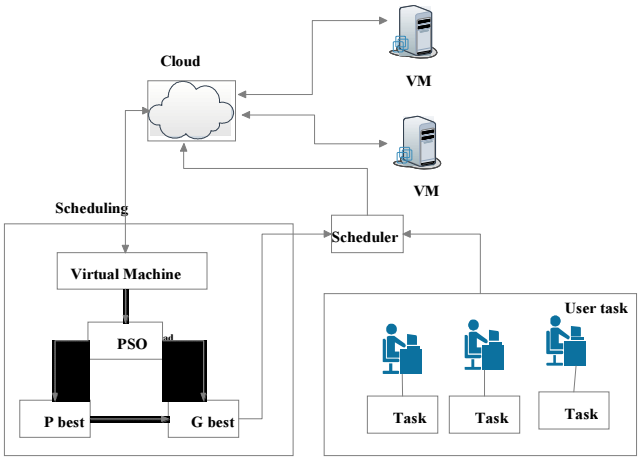

Fig.3.1. Overall Architecture Of The Proposed Method

User sends their task request (for e.g., storing a file in the storage cloud) to the cloud server. The task manager in the cloud server will decide which virtual machine to allocate to store that task. Task manager will select the virtual machine based on the particle swarm optimization algorithm so that the best fit value can be found. Our aim is to balance the load when there is an overloaded virtual machine. First step is to upload the file and cloud server will accept the request and it will transfer that request to the task manager where the VM scheduler performs the scheduling. Main function of VM scheduler is to perform the load balancing. We find the overloaded virtual machine based on the threshold value. After finding the overloaded virtual machine, next step is to migrate the task from overloaded virtual machine to under loaded virtual machine.

4. PROPOSED LOAD BALANCING METHOD

Load balancing of non-pre-emptive independent tasks on virtual machines is an important aspect of task scheduling in clouds. When a VM is overloaded and other VMs are less loaded with tasks , the load should be balanced among the available VMs to achieve better CPU utilization. The optimized algorithm is as follows:

Begin

Initialize the position and velocity randomly. (1)

Initialize the inertia weight. Calculate fitness value for each particle.(2)

Calculate p-best and g-best for each particle.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Do

Update velocity of each particle using (1)

Update position of each particle using (2).

Calculate fitness value of each particle.

Update p-best for each particle if its current fitness value is better.

Update g-best for each particle. i.e, choose the position of the particle with the best fitness value among all particles as the g-best. Update the inertia weight using fuzzy inference rules.

While

Termination criterion is not violated.

End

The CPU utilization is calculated using the formula
CPUusage=filesize+bandwidth
(4.1)

Doing so requires modification of the VM which may not always be possible. Instead, prediction is done based on the past external behaviour of VMs.

Aim is to migrate the task from VM of overloaded to under loaded VM, which can reduce the server's overload.

Trust = min (load).
(4.2)

The under loaded VM is found by
Under loaded VM=average (load in other virtual machine)
(4.3)

When improved positions are being discovered the swarm are guided towards that position until a satisfactory solution is found. This algorithm executes periodically to evaluate the resource allocation status based on the predicted future resource demands of VMs. A server is a hot spot if the utilization of any of its resources is above a certain hot threshold. This indicates that the server is overloaded and hence some VMs running on it should be migrated away. We define a server as a cold spot if the utilizations of all its resources are below a cold threshold. This indicates that the server is mostly idle.

The physical machines provide a set of virtual machines which are configured dynamically according to user requests. When the limited physical machines are provided to users from a pool of resources, the provided resources have two types; one is the dedicated resources and the other is the undedicated resources to give some extra margin in case of sudden request. In this Cloud system environment, if a new user requests resources when all of the resources are already assigned, then the undedicated resources allocated to others are provided to the new users via dynamic reconfiguration. Here we calculate the trust model based on the historical information.

5. IMPLEMENTATION RESULTS AND DISCUSSION

For cloud storage we have taken CloudMe. It features a Cloud storage where users can store, access and share their data.

- Prediction
- Migration
- Particle swarm optimization
- Trust Allocation

Prediction

Predict the future resource needs of VMs. One solution is to look inside a VM for application level statistics, we make our prediction based on the past external behaviours of VMs.

PSO

Our approach is compared with the HBB load balancing model for better optimization of resource usage. A population of candidate solutions and particles are moved around in the search-space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known Virtual machine based HBB load. When improved positions are being discovered these will then come to guide the movements of the swarm. The process is repeated and by doing so it is hoped, but not guaranteed, that a satisfactory solution will eventually be discovered.

Migration

Our method executes periodically to evaluate the resource allocation status based on the predicted

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

future resource demands of VMs and whenever a VM is overloaded the load is distributed uniformly among the available VMs ,so that all VMs are equally loaded and no VMs are overloaded and other VMs are under loaded.

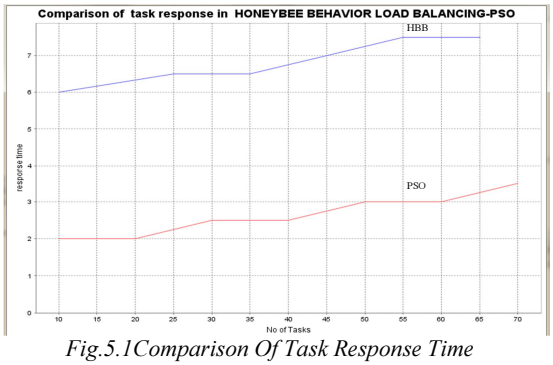

Fig.5.1Comparison Of Task Response Time

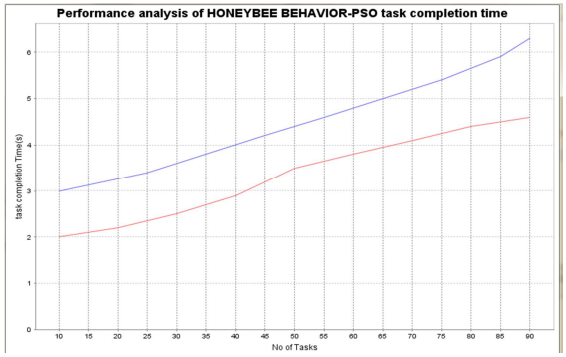

Fig.5.2 Comparison Of Task Completion Time

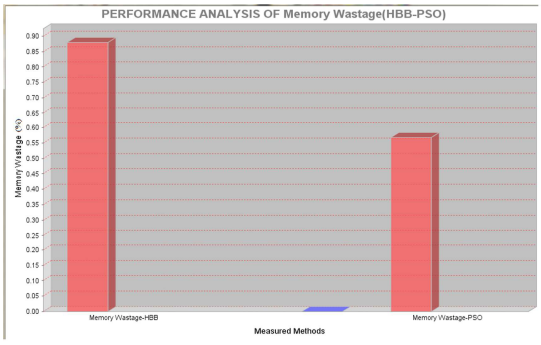

Fig.5.3.Performance Analysis In Terms Of Memory Wastage

Trust Allocation

The physical machines provide a set of virtual machines which are configured dynamically according to user requests. When the limited physical machines are provided to users from a pool of resources, the provided resources have two types; One is the dedicated resource and the other is the undedicated resource to give some extra margin in case of sudden request. In this Cloud system environment, if a new user requests resources when all of the resources are already assigned, then the undedicated resources allocated to others are provided to the new users via dynamic reconfiguration. Here we calculate the trust model based on the historical information. The proposed method is compared with HoneyBee load balancing algorithm in terms of makespan and response time and the QOS parameter memory space wastage.

6. CONCLUSION AND FUTURE WORK

In this paper an optimized scheduling technique has been discussed and implemented. Resource management includes resource discovery, resource scheduling, resource allocation and resource monitoring. As a part of resource management, in this paper an optimized QoS based scheduling technique has been discussed and implemented. The QoS factor considered here is the memory space wastage in storage cloud which protects the cost of the cloud user and the cloud service provider. We try to reduce the wastage of memory space in the VM by finding the best resource using the PSO algorithm and allocating the task to that resource. Most algorithms used for scheduling and load balancing considers make span , response time, throughput, deadline, priority as their performance metric or the QOS factors but in this optimized method of scheduling we have considered the wastage of memory space as the QOS parameter for scheduling. This scheduling method will assign to the virtual machine in best fit manner. i.e., task manager will check all the virtual machine and assigns the task to proper virtual machine which will have least memory wastage. Also the task response time and task execution time has been compared with the Honey Bee algorithm. Simulation results show that the proposed method is more effective when compared with that of the Honey Bee. The QOS factor of the proposed method that is the memory space wastage in allocating the task to virtual machines also compared with the Honey bee algorithm. As a future work of this paper, we will be adopting the same technique for resource allocation so that a cost effective resource allocation can be achieved by both the user and the cloud service provider. since allocation and scheduling are part of resource management we will be incorporating both scheduling and allocation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES:

[1] Abirami S.P. and Shalini Ramanathan," Linear Scheduling Strategy for Resource Allocation in Cloud Environment", 2012.

[2] Amit Nathan,Sanjay chaudharya,gaurav soman., "Policy based resource allocation in Iaas cloud" , 2012.

[3] Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu.".A Multi-Dimensional resource allocation Algorithm in cloud computing", 2012.

[4] Fetahi Wuhib,Rolf Stadler,Mike Spreitzer,"Gossip protocol for Dynamic Resource Management in Large cloud environment", 2012.

[5] G.Sireesha, L.Bharathi.".Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", 2012.

[6] Gihun Jung and Kwang Mong Sim,".Location-Aware Dynamic Resource Allocation Model for Cloud Computing Environmen"t, 2012.

[7] Preeti Agrawal,Yogesh Rathore, "An Approach for Effective Resource Management in Cloud Computing", 2011.

[8] Shikharesh Mujumdar,"Resource management on cloud: Handling uncertainities in parameters and policies" ,2011.

[9] Jiayin Li, Meikang Qiu, Zhong Ming, Gang Quan, Xiao Qin, Zonghua Gu." 2012. Online optimization for scheduling preemptable tasks on IaaS cloud systems" Journal of Parallel and Distributed Computing, 72,666-677, Elsevier Publications.

[10] J.Akhani, S.Chaudhary, and G.Somani, , 2011. "Negotiation of Resource Allocation in IaaS Cloud,Proc. Fourth Annual ACM Bangalore Conference, Bangalore, India.

[11] M. Moradi, M.A. Dezfuli, M.H. Safavi, Department of Computer and IT, Engineering, Amirkabir University of Technology, Tehran, Iran, A New Time Optimizing Probabilistic Load Balancing Algorithm in Grid Computing IEEE 978-1-4244-6349-7/10/©2010.

[12] M. Randles, D. Lamb, A. Taleb-Bendiab, A comparative study into distributed load balancing algorithms for cloud computing, in: Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications Workshops, Perth, Australia, April, 2010, pp. 551–556.

[13] T Amudha, T T Dhivyaprabha,2011" QoS Priority Based Scheduling Algorithm and ProposedFramework for Task Scheduling in a Cloud Environment" IEEE International Conference on Recent Trends in Information Technology (ICRTIT).

[14] A.Nathani, S.Chaudhary, and G.Somani, Jun. 2011. "Policy based resource allocation in IaaS cloud",Future Generation Computer Systems

[15] K. Mukherjee, G. Sahoo, Mathematical model of cloud computing framework using fuzzy bee colony optimization technique, in: Proceedings of the 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, December 28–29, 2009, pp. 664–668.