International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(01), Nov 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

CRITICAL STUDY OF DIFFERENT LOAD BALANCING ALGORITHMS

Vinod Kumar, Shivam Kumra

[1,2] *Assistant Professor, Department of Computer Science*

Dev Samaj College for Women, Ferozepur City (India)

ABSTRACT

Cloud computing is next generation of computing and a developing computing paradigm in the modern industry, either may be government organizations or the public organizations. In simple words we can say that Cloud Computing is set of different servers that cater to need of different clients based on their demands. Clouds have very powerful data centers to handle large number of user's requests. Cloud platform provides dynamic pool of resources and virtualization. Load Balancing is required to properly manage the resources of the service contributor. Load balancing is a technique to distribute the workload among many virtual machines in a Server over the network to achieve optimal resource consumption, decrease in data processing time, decrease in average response time, and avoid overload.Through better load balancing in cloud, performance can be improved and better services are provided to user. Here in this paper we have discussed many different load balancing techniques used to solve the issue in cloud computing environment.

Keywords—Cloud computing; Load balancing; Simulation; Virtual Machine; Cloud

I. INTRODUCTION

Cloud computing is a internet based service provider in which users are allowed to access services on demand.Cloud computing is relatively a new software system technology, which allows dynamic resource allocation on consolidated resources using a combination of different techniques from parallel computing, distributed computing, as well as platform virtualization technologies [1].. Cloud computing has been a primary focus in both the research community and the industry over recent years because of its flexibility in software deployments, and of its elasticity capability on resource consolidation. The latest trends show that a large number of medium and large scale businesses are shifting to cloud. The service providers are increasing day by day and provides services at lower costs.

The main Objective of Cloud Computing is to shift the computational services from desktop to the internet that is moving computation, services offered by them and data off-site to an external, internal, location that is not visible to main contractor. Cloud Computing model is often referred as "pay-per-use model" because we pay amount as per our usage of resources [3].

Cloud computing implements virtualization technique in which a single system can be virtualized into number of virtual systems [6]. On receiving a request from a client Load balancing helps to decides which client will use the virtual machine and which virtual machines wait or will be assigned to different virtual machine. Load balancing of can be managed by using virtualization technology where we can remap Virtual Machines (VMs) and physical resources according to the change in load. Due to these benefits, virtualization technology is most often implemented in Cloud computing. In load balancing there are different challenges that needs to be handled like scalability, throughput, availability ,Virtual machine relocation, , fault tolerance, but main issue is the load balancing , it is the process of distributing the load among various nodes of a distributed system in order to minimize the communication delay and to minimize the resource utilization and also avoiding a situation when some of the machines have large amount of data and consuming excess time while others have huge amount of load while other nodes are doing nothing or idle with very little work.

II. TYPES OF CLOUD COMPUTING

Cloud computing provides three types of services :

- **Infrastructure as a Service (IaaS):** It provides access to fundamental resources within the cloud i.e. virtual machines, storage etc. In this users need not to buy required servers or network resources of their own. The users pay only for the time duration they are using the service. [7].
- **Platform as a Service (PaaS):** It helps to provide runtime environment to build an application. In this model,Cloud computing provides a way where resources are available and users can create the required applications by themselves.
- **Software as a Service (SaaS):** It allows the users to use software applications as a service from various cloud providers through the internet [9]. In this type elasticity makes a cloud application different from another application.

III. LOAD BALANCING

Cloud computing is one of the fastest adopted and implemented technology in various sectors. Many organizations these days are implementing and setting up clouds, due to flexible architecture of cloud which always results in the increase in number of users reaching cloud and ultimately improving performance. Although clouds are categorized as public, private and hybrid models but still there may be problem of reliability in these clouds [4][5]. Cloud computing has been used by most of the organization such as, social networking websites, online applications design by Google doc and Several clouds are also used for online software testing [14].

Load balancing is one of the most important aspect in cloud computing environment that can purposeful improve resource utilization, performance and save energy by properly assigning/reassigning computing resources to the incoming requests from users. Therefore how to schedule virtual machines (VMs) effectively by considering various parameters that can influence its decision becomes an important research point for cloud computing.

Existing Load Balancing Algorithms for Cloud Computing

To distribute workload among multiple network links among multiple virtual machines and to achieve maximum throughput, minimize response time. We use two algorithms to distribute the load.

A. Round Robin: In RR algorithm , the cloudlets or jobs are equally divided among all processors. Each cloudlet is assigned to the processor in a timely manner. The process distribution order is maintained locally independent of the allocations from isolated processors. Though the load distributions between processors are equal but the processing time for different processes are not same. So it may be possible that at any point of time some processors may be heavily loaded and others remain idle.In round robin scheduling the time slice play a vital role for scheduling, because if time slice is very large then round robin scheduling algorithm behave like FCFS scheduling. If the time quantum is small then context switching will be more and response time will be less.

METHODOLOGY

1. Vmload Balancer maintains an index of VMs and state of the Vms (busy/available).
2. Initially all Vms are available
3.The Data center controller receives the user requests/cloudlets.
4. The requests are allocated to Vms on the basis of their states known from the VM queue.
5. The roundrobinvmloadbalancer will allocate the time quantum for user request execution.
6. The Vmloadbalancer will decides the scheduling order based on arrival time and execution time.
7. After the execution of cloudlets, the VMs are de- allocated by the VmLoadBalancer.
8. The datacentercontroller checks for new /pending/waiting requests in queue.

B. Equally Spread Current Execution Algorithm(ESCE):

The load balancer tries to protect equal load to all the virtual machines connected with the data centre. In Equally spread current execution algorithm, the processes are handled with load priorities. It distributes the load to virtual machine by checking the load at current time and transfer of the load to that virtual machine which is lightly loaded and handles that request easily and result in less time taken , and give maximum possible throughput. In this technique the load balancer tries to divide the load into multiple virtual machines.

It maintain a index table containing list of virtual machine with current load. When all the virtual machines are currently loaded and when there is a request to the data centre to allocate the new VM, it scans the table for VM which is least loaded. If in case there are more than one VM is found than first come first serve algorithm is used and first identified VM is selected for handling the request of the client/node, the load balancer returns VM id to datacenter broker. The data centre communicates the demand to the VM identified by that id. After each allocation the index table is updated and When task is completed, it is informed to data centre which is further notified by the load balancer. The load balancer again updates the index table and result in decreasing the allocation count by one but in this there is an additional overhead of scanning the queue again and again.

METHODOLOGY

1. Initially all Vms are available
2. When a job is requested to datacenterbroker.
3. Count the active load on each VM
4. Return the id of those VM which is having least load.
5. The VMLoadBalancer will allocate the request to one of the VM.
6. If a VM is overloaded then the VMLoadBalancer will distribute some of its work to the VM having least work so that every VM is equally loaded.
7. The datacentercontroller receives the response to the request sent and then allocate the waiting requests from the job pool/queue to the available VM & so on.

C. Throttled Load Balancing Algorithm(TLB)

In TLB algorithm, an index table is maintained by load balancer which contains virtual machines as well as their states (Available or Busy). On receiving a request from client data centre firstly tries to find a suitable virtual machine (VM) to perform the requested task. The data centre broker ask the load balancer for allocation of the VM. The index table is scanned from top by the load balancer until the first available

VM is found or the index table is scanned fully. If the status of any VM is Available, then VM id is send to the data centre. The data centre then allocates the request to the VM identified using the throttled algorithm. Also, the data centre updates the index table and set the state of Vm to Busy. But during processing the request of client, if no VM is found, the load balancer returns -1 to the data centre [7][8]. The data centre queues the request of the client with it. When a certain VM completes its task, a request is sent to data centre to update its index table. The total execution time can be estimated in three phases. During first stage there is configuration of the virtual machines and they will be idle waiting for tasks, once tasks are allocated, the virtual machines in the cloud will start processing their assigned tasks, which is considered as the second phase, and finally during the third phase after completion of their dedicated tasks the virtual machines are de-allocated.

METHODOLOGY

1.Throttled VmLoadBalancer maintains an index table of VMs and the state of the VM (BUSY/AVAILABLE). At the start allVM’s are available.

2. DataCenterBroker receives a new request.

3. DataCenterBroker queries the ThrottledVmLoadBalancer for the next allocation.

4. ThrottledVmLoadBalancer check the table from top until the first available VM is found.

5.If VM is found available then ThrottledVmLoadBalancer returns the VM id to the DataCenterBroker.

6. The DataCenterBroker sends the request to the VM identified by that id.

7.DataCenterBroker notifies the ThrottledVmLoadBalancer of the new allocation.

8.If all Vm’s are busy then ThrottledVmLoadBalancer returns

9. When the VM finishes processing the request, and the DataCenterBroker receives the response cloudlet, it notifies the ThrottledVmLoadBalancer of the VM de-allocation.

D. First Come First Serve

FCFS (First Come First Served), used in parallel task processing, is the simplest task ordering strategy. It chooses and processes the task according to the sequence in which they request the DataCenterBroker. With this scheme the user request which comes first to the datacenterbroker is allocated the virtual machine for execution first. The implementation of FCFS policy is easily managed with FIFO queue. The datacenterBroker searches for virtual machine which is in idle state or underloaded. Then the 1st request from the queue is removed and passed to one of the VM through the VMLoadBalancer

METHODOLOGY

1. FCFS VmloadBalancer maintains an index table of virtual machines & number of requests currently allocated to the VM. At start all have zero allocation.
2. The vmloadbalancer allocates the cloudlets/user requests to the available VMs on the basis of requests sent by the datacenterBroker.
3. The datacenterBroker stores the user requests in a queue on the basis of their arrival time.
4. The first request according to the arrival time is allocated to the VM which is under loaded or free by FCFS VmloadBalancer.

5.The FCFSVmLoadBalancer will execute the cloudlets and calculate the turnaround time, avg. waiting time and response time. After that it will display the result.

6.The datacenterBroker receives the response to the request sent and then allocate the waiting requests from the job pool/queue to the available VM & so on.

7Completion of their dedicated tasks the virtual machines are de-allocated.

E.MaxMin Load Balancing algorithm:

The Max-min algorithm is commonly used in distributed environment. When a job is requested to datacenterbroker then completion time is calculated for each task on the available resources[2]. This algorithm chooses large tasks to be executed firstly, which in turn small task delays for long time. This algorithm also works in two phases. First, the maximum expected completion time for all the tasks is calculated. The completion time for all the tasks is calculated on all the virtual machines. In the second phase, the task with the maximum expected completion time from makespan is selected and that tasks assigned to the corresponding resource. After the completion of the current task it is removed from the makespan and this process is repeated until all tasks are completed.

METHODOLOGY

1. MaxMin VmloadBalancer maintains an index table of virtual machines & number of requests currently allocated to the VM. At start all have zero allocation.
2. The vmloadbalancer allocates the cloudlets/user requests to the available VM.
3. Expected completion time is calculated for each VM

4. Task with maximum expected completion time is selected and is assigned to corresponding VM.
5. After completion of task the index table is update

IV. CONCLUSION AND FUTURE WORK

Cloud Computing has widely adopted by the many organizations, still there are some issues like Load Balancing, Virtual Machine Migration, Energy Management, etc. One of the concerned issues is the issue of load balancing, which distribute the load from heavily loaded to lightly loaded among all nodes to improve efficiency and user satisfaction. Using a proper load balancing algorithm, resource consumption can be kept to a minimum which will further reduce energy consumption . There are many existing load balancing techniques out of which few famous techniques are discussed in this paper that mainly focus on reducing associated overhead, service response time and improving performance etc. but none of the techniques has considered the energy consumption factors. Therefore, there is a need to develop an energy-efficient load balancing technique that can improve the performance of cloud computing by balancing the workload across all the nodes in the cloud along with maximum resource utilization.

REFERENCES

[1] http://research.ijcaonline.org/ncetct/number1/NCET CT4017.pdfthe-president’s-budget-making-cloud-computing-a-priority-for-thefuture as on Sep. 2012.

[2] Rajwinder Kaur, Pawan Luthra “ Load Balancing in Cloud System using Max Min and Min Min Algorithm”, National Conference on Emerging Trends in Computer Technology (NCETCT-2014)

[3] Eddy Caron , Luis Rodero-Merino “Auto-Scaling , Load Balancing And Monitoring In Commercial And Open-Source Clouds “ Research Report, January2012

[4] Bhavisha Kanani, Bhumi Maniyar,” Review on Max-Min Task scheduling Algorithm for Cloud Computing JETIR March 2015, Volume 2, Issue 3.

[5] Anthony T. Velte ,Toby J. Velte, Robert Elsenpeter, ”Cloud Computing: A Practical Approach ”, The McGraw-Hill Companies(2010), [Book]

[6] Saroj Hiranwal , Dr. K.C. Roy, ”Adaptive Round Robin Scheduling Using Shortest Burst Approach Based On Smart Time Slice” International Journal Of Computer Science And Communication July-December 2011 ,Vol. 2, No. 2 , Pp. 319-323

[7] Jaspreet Kaur “Comparison of load balancing algorithm in cloud”, june International Journal of Engineering Research and Applications 2012.

[8] Bhathiya Wickremasinghe ,Roderigo N. Calherios “Cloud Analyst: A Cloud-Sim-Based Visual Modeler For Analyzing Cloud Computing Environments and Applications”. Proc Of IEEE International Conference on Advance Information Networking And Applications ,2010.

[9] R. Buyya, R. Ranjan, and R. N. Calheiros, “Modeling And Simulation Of Scalable Cloud Computing Environments And The Cloudsim Toolkit: Challenges and Opportunities,” Proc. Of The 7th High Performance Computing and Simulation Conference (HPCS 09), IEEE Computer Society, June 2009.

[10] www.cloudbus.org/cloudsim.

[11] M. Randles, D. Lamb, and A. Taleb-Bendiab, A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing, IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, 2010, pp. 551–556.

[12] Foster, I; Yong Zhao; Raicu, I.; Lu, S. “Cloud Computing and Grid Computing 360-Degree Compared”, published in Grid Computing Environments Workshop, 2008. GCE '08 IEEE DOI 12-16 Nov. 2008.

[13] R. Buyya, C. S. Yeo, and S. Venugopal, “Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities”,Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC 2008, IEEE CS Press, Los Alamitos, CA, USA), Sept. 25-27, 2008, Dalian, China.

[14] Sun Microsystems, Inc.”Introduction to Cloud Computing Architecture” Whitepaper, Ist Edition, June 2009.

[15] F. Howell and R. Macnab, “SimJava: a discrete event simulation library for Java,” Proc. of the 1st International Conference on Web based Modeling and Simulation, SCS, Jan. 2008.

[16] R. Buyya, and M. Murshed, “GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing,”Concurrency and Computation: Practice and Experience, vol. 14, Nov. 2002, pp. 1175-1220.

[17] M. Armbrust , A. Fox, R. Griffith, A. D. Joseph, R. Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica,And M. Zaharia, “Above The Clouds: A Berkeley View Of Cloud Computing”, Eecs Department, University Of California , Berkeley,, February 2009,Technical Report No., Ucb/Eecs-2009-28, Pages 1-23.

[18] Rich Lee, Bingchiang Jeng ”Load Balancing Tactics In Cloud” International Conference On Cyber Enabled Distributed Computing And Knowledge Discovery, 2011

[19] A Survey on Open-source Cloud Computing Solutions Patrícia Takako Endo, Glauco Estácio Gonçalves, Judith Kelner.

[20] Zhong Xu, Rong Huang,(2009)“Performance Study of Load Balanacing Algorithms in Distributed Web Server Systems”, CS213 Parallel and Distributed Processing Project Report.