International Journal of Scientific and Engineering Research, Volume 4, Issue 12, December-2013
ISSN 2229-5518

Efficient Load Distribution of VM in Cloud Computing

Vartta Siyal, Dr. Naveen Choudhary, Dr. Dharm Singh

Abstract - As we are aware that cloud computing is becoming popular these days due to its high availability and applicability in current world scenarios. One of the critical performance issues with cloud computing is load balancing/ distribution. Load balancing is a methodology to distribute workload across multiple computers, or other resources over the network links to achieve optimal resource utilization, maximum throughput and minimum response time. Load Balancing is vital for cloud computing environment to enhance the job allocation strategies for efficient resource utilization. Many critical issues are required to be addressed for implementing efficient and effective load distribution techniques for cloud computing. Various load balancing techniques in the recent research literature are analysed in this paper. To improve the performance of cloud, a new VM load balancing algorithm has been proposed and implemented has done to achieve better response time.

Index Terms—Cloud Computing, Load Balancing, Round Robin, Load Distribution Algorithms, Cloud Sim, Active Monitoring Algorithm, Throttled Algorithm.

1. INTRODUCTION

Cloud computing comes into picture when you imagine about IT needs from a way to increase capacity or add capabilities on the run without investing in new infrastructure or to train new personnel or to license new software. Cloud Computing offer users to access distributed scalable, virtualized hardware or software infrastructure over the Internet.

Any technology that advances has its own pros and cons. Likewise cloud computing also has its own issues related to the load management, fault tolerance, data security etc. in cloud environment.

The performance of cloud computation system critically depends on several aspects, important one of which is load balancing/ distribution. The load balancing mechanism depends on the amount of task assigned to the system for a specific time period. This is the time where system has to manage and work according to the priority. Some load balancing algorithms which can be applicable in cloud computing environment are surveyed in the current paper. The algorithms may be static or dynamic in nature. [1]

Load balancing is an important prerequisite to utilize the full resources of cloud computing system. Load balancing mechanisms can be broadly categorized as sender initiated, receiver initiated or symmetric and dynamic or static. Physical resources can be split into a number of logical slices called Virtual Machines (VMs). [2]

2. LOAD BALANCING CATEGORIZATION

Load balancing is the process of distributing the load among various resources in a system. Thus load need to be distributed over the resources in cloud-based architecture in such a manner that each resource is approximately equally loaded at any point of time. Load balancing serves two important needs, primarily to promote availability of Cloud resources and secondarily to enhance performance of the system. [3].

In order to balance the requests for the resources, there are few major goals of load balancing algorithms as explained below:

a. Cost effectiveness: primary aim is to achieve an overall improvement in system performance at a reasonable cost.

b. Scalability and flexibility: the cloud computing system in which the algorithm is implemented may change in size or topology. So the algorithm must be scalable and flexible enough to allow such changes to be handled easily.

c. Priority: prioritization of the resources or jobs need to be done before hand through the algorithm itself for assigning better service to the important or high prioritized jobs in spite of equal service provision for all the jobs regardless of their origin. [4]

In cloud computing, proper load balancing aids in minimizing resource consumption, implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning etc. [5] For this various performance metrics as considered below are considered in existing load balancing techniques in cloud computing :

1. Throughput is used to calculate the no. of tasks whose execution has been completed.

2. Fault Tolerance is the ability of an algorithm to perform uniform load balancing in case of link failure. The load balancing should be a good fault-tolerant technique.

3. Migration time is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.

4. Performance is used to check the efficiency of the system. This has to be improved at a reasonable cost, e.g., reduce task response time while keeping acceptable delays.

5. Response Time is the amount of time taken to respond by a particular load balancing algorithm in a distributed environment.

6. Resource Utilization is used to check the utilization of resources.

7. Scalability is the ability of an algorithm to scale according to the requirement. [5]

There are 4 popular policies or strategies generally followed for load balancing as:

1. Transfer Policy: The part of the dynamic load balancing algorithm which selects a job for transferring from a local node to a remote node is referred to as transfer policy or transfer strategy.
2. Selection Policy: It specifies the processors involved in the load exchange (processor matching)
3. Location Policy: The part of the load balancing algorithm which selects a destination node for a

International Journal of Scientific and Engineering Research, Volume 4, Issue 12, December-2013
ISSN 2229-5518

transferred task is referred to as location policy or location strategy.

4. Information Policy: The part of the dynamic load balancing algorithm responsible for collecting information about the nodes in the system is referred to as information policy or information strategy.[6]

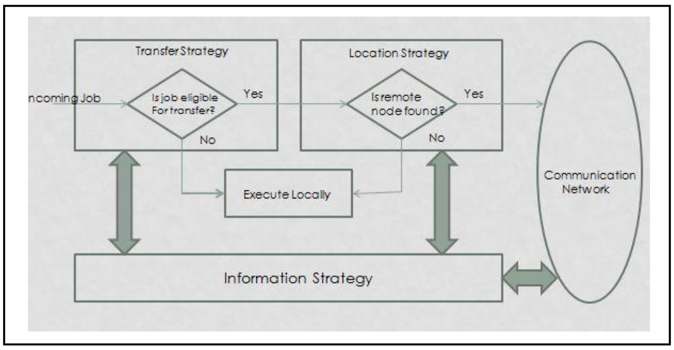

Fig. 1 Policies used in Load balancing techniques.

Depending on who initiated the process, load balancing algorithms can be of three categories as given in [7]:

a. Sender Initiated: If the load balancing algorithm is initialised by the sender.
b. Receiver Initiated: If the load balancing algorithm is initiated by the receiver.
c. Symmetric: It is the combination of both sender initiated and receiver initiated.

3. POPULAR LOAD BALANCING TECHNIQUES

Some of the established and popular load balancing techniques are briefly discussed in this section for establishing comparison with the proposed load balancing methodology are presented in this section.

A. Active Monitoring Load Balancing Algorithm

Active Monitoring algorithm starts with selection of least loaded virtual machine (VM) to allocate the new job/task. For this purpose it uses the information of current request allocation of each VM. When a new request arrives it will check the first least loaded VM and assign that VM for the execution of task using Data Centre Controller. [2]

M. Sharma, P. Sharma, and S. Sharma [2] proposed an algorithm to find the expected response time of each resource (VM) and return the ID of virtual machine having minimum response time for allocation to the new request. According to them if we select a efficient virtual machine then it effect the overall performance of the cloud environment and also decrease the average response time. [2]

Since virtual machines are of heterogeneous platform, the expected response time can be found with the help of the following equation:[2]

Response Time = Final_time – Arr_time + TDelay

B. Round Robin Load Balancing Algortihm:

According to Nikita, Shaveta, and G. Raj [8] round robin algorithm is random sampling based method that selects the load randomly resulting in some heavily loaded servers or some lightly loaded servers. [8]

Round robin uses the time slicing mechanism. As name suggests that it works in the round robin manner or in circular fashion passing each new request to next node in line where each node is assigned with a time slice and each node has to wait for their turn. The time is divided and interval is allotted to each node. Each node is allotted with a time slice in which they have to perform their task. [1]

C. Throttled Load balancing Algorithm:

The Throttled algorithm starts by finding the suitable virtual machine for assigning a particular job. The job manager has a list of all virtual machines and using this list, it allocates the job to the appropriate machine. If no virtual machine is available to accept jobs then the job manager waits and takes the job in queue for fast processing. [1]

H. S. Mahalle, P. R. Kaveri, and V. Chavan [1] concluded that throttled load balancing algorithm reduces the cost of usage, so it works more efficiently in terms of cost for load balancing on cloud data centers. [1]

D. Equally Spread Current Execution Algorithm/Active Monitoring Algorithm

Equally spread current execution algorithm is a spread spectrum technique in which the load balancer spread the load of the job in hand into multiple virtual machines. It distributes the load randomly by checking the size and transfers the load to that virtual machine which is lightly loaded or handles that task easily and takes less time with maximum throughput. [8]

Equal Spread Current Execution Load algorithm dynamically allocates the resources to the job in queue leading to reduced cost for data transfer and virtual machine formation. The ESCE algorithm shows significant improvement in response time and the processing time. The ESCE equally spread the jobs, on various VMs leading to load balanced cloud computing environment and avoiding underutilization of any VMs. Due to this advantage, there is reduction in the virtual machine cost and the data transfer cost. [10]

4. PROPOSED LOAD BALANCING ALGORITHM

The paper briefly surveys various popular techniques for load distribution in cloud computing domain. The new algorithm has been proposed by modifying active monitoring algorithm to achieve better response time.

Virtual machine Usage is an important factor. Based on this factor we can efficiently balance the load. The VMUsage is considered in the terms of total time taken by the VM to execute a cloudlet, i.e.

Total time = end time – start time.

The start time is calculated every time a VM is allocated a new task according to clock value and end time is also noted when it deallocates. Thus it returns the total time taken by the VM to complete a task. Every time a task is executed, algorithm computes VMUsage and updates the previous value.

International Journal of Scientific and Engineering Research, Volume 4, Issue 12, December-2013
ISSN 2229-5518

The Proposed VM Load balancing algorithm firstly finds the VMUsage of each VM. Then next available VM is found based on the popular VM policy (VM having less VMUsage and least loaded from the current allocation count of VM) and returns the VMID to datacenter controller. Proposed algorithm finds the total time taken by each Virtual machine.

1. When a request for allocating a new VM arrives, Algorithm finds the most popular VM (efficient VM having least loaded, minimum VMUsage) and return the respective VMId to Datacenter Controller.
2. The new allocation is notified and algo updates the allocation table increasing the allocations count for That VM.
3. After finishing the processing of the request, the Datacenter Controller receives the Response. Datacenter controller notifies the efficient algorithm for the VM de-allocation.

5. IMPLEMENTATION AND RESULTS

The implementation of the proposed algorithm is done in the cloudsim [11] based simulation environment named Cloud Analyst [12]. To implement the proposed load balancing algorithm Java language is used. The parameters taken into account are as follows-

PARAMETERS	VALUES
VM Image size	10000
VM Memory	1024Mb
VM Bandwidth	1000
No. of Data Centres	3
Data Center (No. of Machines)	20
Data Centers (No. of VMs per Data Center)	(75, 50, 25)
User Grouping Factor	1000
Request Grouping Factor	100
Executable Instruction Length	250

FIGURE I. PARAMETER VALUE

The overall response time of the proposed algorithm and the cost factor is as shown in the table-

	Avg. (ms)	Min (ms)	Max(ms)
Overall Response Time	119.05	64.51	407.51
DataCenter Processing Time	22.64	8.01	25.76

Total VM Cost ($)	360.04	
Total Data Transfer Cost ($)	20.33	
Grand Total	380.37	

FIGURE II. RESULT DETAILS

The comparison of the existing popular VM load balancing algorithms with the proposed algorithm shows that the proposed algorithm achieves the better response time than the popular existing algorithms. The graphical representation of the comparison is as-

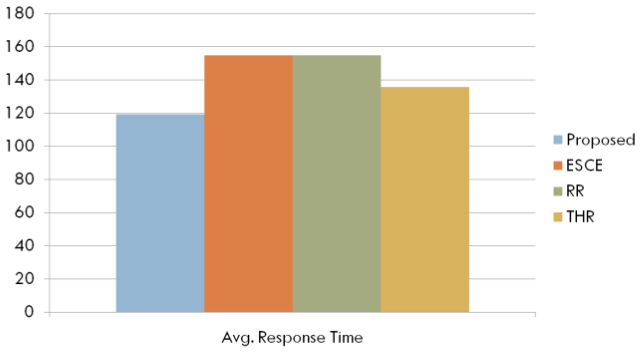

FIGURE III. COMAPRISON ANALYSIS

6. CONCLUSION

The paper briefly surveys various popular techniques for load distribution in cloud computing domain and proposed a new load balancing algorithm.

It is seen that the response time in the algorithm is directly dependent to the performance of the cloud computing scenario. The more efficient algorithm will provide higher performance.

The paper tries to present the proposed load balancing technique for cloud computing and analyze its benefits and drawbacks with respect to load balancing algorithms for the increasingly becoming popular cloud computing environment.

7. REFERENCES

[1] H. S. Mahalle, P. R. Kaveri, and V. Chavan, “Load balancing on cloud data centres,” IJARCSSE. vol. 3, pp 1-4, January 2013

[2] M. Sharma, P. Sharma, and S. Sharma, “Efficient load balancing algorithm in VM cloud environment,” IJCST. vol. 3, pp. 439-441, March 2012

[3] Z. Chaczko, V. Mahadevan, S. Aslanzadeh, and C. Mcdermid, “Availability and load balancing in cloud computing,” in International Conference on Computer and Software Modelling, IPCSIT, Singapore. vol. 14, pp. 134-140, 2011.

[4] S. Ray, and A. D. Sarkar, “Execution analysis of load balancing algorithms in cloud computing environment,” IJCCSA. Vol. 2, pp. 1-13, October 2012.

[5] N. Sran, and N. Kaur, “Comparative analysis of existing load balancing techniques in cloud computing,” IJESI. Vol. 2, pp. 60-63, January 2013.

[6] R. P. Padhy, G. P. Rao, “Load Balancing in cloud computing systems,” Thesis, National Institute of Technology, Rourkela, Orissa, India. May, 2011.

[7] A. M. Alakeel, “A guide to dynamic load balancing in distributed computer systems,” IJCSNS. Vol. 10, pp. 153- 160, June 2010.

[8] Nikita, Shaveta, and G. Raj, “Comparative analysis of load balancing algorithms in cloud computing,” IJARCET. Vol. 1, pp. 120-124, May 2012.

[9] G. Raj, and A. Nischal, “Efficient resource allocation in resource provisioning policies over resource cloud communication paradigm,” IJCCSA. Vol. 2, pp. 11-18, June 2012.

[10] J. Kaur,” Comparison of load balancing algorithms in a cloud,” IJERA. Vol. 2, pp. 1169-1173, May-June 2012.

[11] Rodrigo et. al., “CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms.”2010.

International Journal of Scientific and Engineering Research, Volume 4, Issue 12, December-2013
ISSN 2229-5518

[12] Bhathiya, Wickremasinghe. "Cloud Analyst: A Cloud Sim-based Visual Modeller for Analysing Cloud Computing Environments and Applications."

Vartta Siyal
M.Tech Student,
Department Of Computer Science,
College of Technology and Engineering,
Udaipur
Email Id- vtvartta@gmail.com

Dr. Naveen Choudhary
Head, Associate Professor
Department Of Computer Science,
College of Technology and Engineering,
Udaipur
Email Id- naveenc121@gmail.com

Dr. Dharm Singh
Assistant Professor
Department Of Computer Science,
College of Technology and Engineering,
Udaipur
Email Id- dharm@mpuat.ac.in