A REVIEW OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Harish Sharma[1], Pradeep Semwal[2]

[1, 2] SGRR University, Dehradun, India

ABSTRACT - Cloud Computing is performing computing using the internet facility. Computing is performed as on demand of the user. The Cloud Computing Load Balancing algorithms can be applied in Static, Dynamic and Centralized environment. The paper compares and summarizes some of the load balancing strategies in cloud computing environment. The paper also discusses advantage and disadvantage for different Load Balancing Algorithm in Cloud Computing Environment.

Keywords: *Cloud Computing, Load Balancing, Static Load Balancing, Dynamic Load Balancing*

I. INTRODUCTION

Cloud Computing [1] provides pool of resources to its end-user that can be reallocated to different purposes within short time frames. Cloud computing power is delivery of computing services from computing power to computing infrastructure, applications etc. to end users as and when it is needed. The cloud computing combines the set of hardware, software, networks, storage, services and interfaces that combine to deliver aspect of computing as a service.

There are different reasons for using cloud computing.

- No hardware, software is required.
- Operating System independent
- Dynamic allocation
- Movement of programs
- Scalability
- Pay as you use
- No commitments
- Massive, Web-scale abstracted infrastructure.

Cloud Computing uses requirement based hardware as its base. The hardware can be replaced at any time without affecting the cloud. It uses a commodity – based software container system. For example, a service should be able to be moved from one cloud provider to any other cloud provider with no effect on service. Cloud Computing requires virtualization, abstraction layer for hardware and software. The some services provided by the Cloud Computing are:

- Software-as-a-service (SaaS)
- Platform-as-a-Service (PaaS)
- Infrastructure-as-a-Service (IaaS)

II. LOAD BALANCING

Load Balancing in Cloud Computing is needed to handle networked or distributed system in such a way that task the allocated to all processor for the efficient utilization of the resources [2]. For simplifying load balancing globally in cloud computing a proper scheduling algorithm is needed. Load Balancing is distribution of the load instruction to the multiple nodes across the network. The responsibility of load balancing algorithmic program is to pick next task in such a fashion to attenuate the execution time, utilization of resources at the datacenters. Many load balancing algorithm have been used in Cloud Computing for balancing load across the network.

Load Balancing Algorithm aims at:

1. Maximize the throughput of network.
2. Minimize the system overhead.
3. Ensure the reliability in the service.
4. Scalability of the system
5. Resources are efficiently utilized
6. Fault Tolerance

Cloud Computing Load Balancing can be divided into :
- Static load balancing
- Dynamic Load Balancing

Static Load Balancing

Static load balancing [3] uses a prior information about all the characteristics of the task, the computing resources, and communication network, memory are known and provided. Static load balancing algorithm are non-preemptive types.

Dynamic Load Balancing

In Dynamic Load Balancing no prior information about jobs are known. In dynamic load balancing jobs move from overloaded node to the node which is under loaded. The

dynamic load balancing algorithms can be centralized or distributed depending on whether the responsibility for the task of global dynamic scheduling should physically reside in a single processor (centralized) or the work involved in making decisions should be physically distributed among processors dynamic load balancing algorithm can be distributed and non-distributed. [4][5]. ClousSim [6] toolkit can be used for simulating the load balancing algorithm and measuring its effectiveness. [7]

III. LOAD BALANCING ALGORITHMS

Load Balancing is important and challenging issue in Cloud Computing. Load Balancing in Cloud Computing helps in efficient use of resource utilization, less response time, equal distribution of load and less power consumption, provide services to achieve complete resource utilization. Several algorithms has been proposed for load balancing in Cloud Computing few of them has been discussed below:

Round Robin Algorithm

Round Robin [8] is the simplest algorithm work in a circular fashion. With this nature of the algorithmic rule computer hardware allocates a time quanta or time slice to execute a task on each node. Once a VM is assigned a task it moves to the end of list. Round Robin provides higher performance than FCFS. If time slice is simply too huge then round robin behaves likes FCFS and if time slice is simply too short then it'll increase context modification inside the round robin algorithmic. At some point of time some nodes may remain idle while other heavily loaded.

Weighted Round Robin

It is the changed version of Round Robin. Task are assigned according to the capacity of the VM the higher VM will get more number of task. Server can be allocated weight, an integer value which represent processing power of VM. [10].

Dynamic Round Robin

This algorithmic works on reducing the energy consumption with the following steps.

i) If virtual machine has completed its execution and if there are more VM on this physical machine, physical machine will not accept any more VM. Such physical machines are referred to as to be in "retiring" state, i.e. once remainder of the virtual machines finishes their execution, then this physical machine will shut down.

ii) If a physical machine is in a retiring state for a long time then rather than waiting, all the running virtual machines cab be migrated to different physical machines. This waiting time threshold is termed "retirement threshold.

Throttled Load Balancer

The Throttled Load Balancer (TLB) maintains VM state in a record table every virtual machine can be in a busy/idle state. Once a request arrives it searches the table and if a match is found on the idea of size and availableness of the machine, then the request is accepted otherwise it came back and also the request is queued [10]. Current load of the VM is not considered during allocation.

Central Load Balancer

The algorithm central load balancer balances load among virtual machine with different hardware. Central load balancer maintains a table which contain VM id and states (busy/idle). It is essentially updated version of Throttled algorithmic program. Similar to the throttled it conjointly maintains a table containing the state of every VM at the side of their priority. The priority is calculated supported by processor speed and capability of memory [13]. The VM assignment policy is comparable thereto of throttled except that during this algorithmic program the VM with highest priority can get the primary preference. If it's busy then the VM with next highest priority is checked and therefore the method continues till a VM is found or the full table is searched. The algorithmic program with efficiency balances load in an exceedingly heterogeneous setting however this algorithm suffers from hindrance as all the requests can return to central load balancer. What is more the algorithmic program is predicated on the priority of VMs that is calculated in an exceedingly static method and isn't updated throughout job allocation.

Active Monitoring Load Balancing (AMLB)

AMLB [15] maintains a table of information concerning every VM and   the variety of requests presently allotted to that VM. When request arrives it checks for the least loaded Virtual Machine. Once letter of invitation to assign a replacement to VM arrives, it identifies the least loaded VM. If there exists one, the primary known is chosen. The load balancer returns the VM id to the info Center Controller. It passes the request to the VM known by that id and notifies the Active VM Load Balancer of the new allocation throughout allocation of VM solely importance is given on the present load of VM, its processing power isn't taken into account therefore, the waiting time of some jobs could increase violating the QoS demand.

VM Assigned Load Balancing

The algorithm is a changed version of Active Monitoring Load Balancing Algorithm. If a request for the new VM comes it checks for the new VM request in the VM Table if VM is available then a VM id is assigned and returned to the data center.

Shridhar G. Domanal et. al., [10] this algorithm will make use of all available VM can utilize all the VMs fully and properly not like the previous one wherever few VMs are full with several requests and rest can stay underneath used. however it's not clearly mentioned within the paper that how it happens. This algorithm won't use the VM if it's already allotted within the last cycle. However,   there's no logic behind it as a result of it should still be the smallest amount loaded VM having smart process speed. Thus a lot of tasks is appointed thereto. To search an evenly loaded VM can distribute the tasks equally only if there measure multiple VMs that equally loaded or succeeding least loaded VM contains a high process speed compare to the previous one. However the algorithm solely considers the load and if the VMs are equally loaded then the task is submitted to any of them regardless of the very fact that whether or not the VM is employed within the last iteration or not.

Load Balancing Min-Min

Min-Min [9][12] algorithm picks task with least execution time could also be a simple and fast formula capable of providing improved performance**.** Min-Min schedules the best tasks with the least execution time schedules and improve the overall produce span. Thus, smaller tasks will get dead first, whereas the larger tasks has to keeps on inside the waiting stage, which will finally lands up in poor machine usage. Min-Min illustrate minimum completion time for jobs that does not consider the previous load of the machine.

Load Balancing Improved Min-Min (LBIMM)

It starts by execution Min-Min algorithmic rule at the primary step. At the second step it chooses the tiniest size task from the heaviest loaded resource and calculates the completion time for that task on all different resources. Then the minimum completion time of that task is compared with the makespan made by Min-Min. If it's but makespan then the task is reassigned to the resource that manufacture it, and also the prepared time of each resources are updated. the method repeats till no different resources will manufacture less completion time for the tiniest task on the serious loaded resource than the makespan. so the overladen resources are freed and also the under loaded or idle resources are a lot of used. This makes LBIMM [10] to supply a schedule that improves load reconciliation and additionally reduces the completion time. however still it doesn't think about priority of employment whereas planning.

Max-Min Algorithm

In Max-Min [9][11] algorithm task with largest completion time of all available task is selected and executed in the node (VM) which produces minimum execution time for the selected task. Same procedure is repeated for the rest of the task.

User Priority Awarded Load Balance Improved Min-Min

User priority used by H. Chen et. al. [11] is included with LBIMM rule to develop PA-LBIMM. The algorithm works on two teams. All task are divided into two group G1 and G2 group is G1 is for high priority task group G2 is for ordinary task. The high priority task are executed first Min-Min algorithm. The ordinary priority task are scheduled next. Finally load balancing algorithm produces makespan of the executed task. [10]

Opportunistic Load Balancing (OLB)

Opportunistic load balancing (OLB) [10] [12] is a static load balancing algorithmic. OLB keeps all nodes busy thus doesn't think about the previous load. . However, OLB doesn't think about the execution time of the task in this node. This could

cause the task to be processed during a slower manner increasing the full completion time (makespan) and can cause some overhead since requests can be unfinished looking forward to nodes to be free.

Honey Bee Foraging

Honey Bee [16] is nature inspired distributed load balancing technique. It achieve load balancing through local server. Forager bees hunt for food supplies and when finding advertise this by waggle dance to gift quality of nectar or distance of food source from hive. Harvester bees then follow the foragers to the location of food to harvest it. [7]

In Honeybee Foraging servers are divided into virtual server having its own queue to maintain the server request and compute its processing time.

    (i)   Calculate profit for certain request. Profit can be set as needed. Generally calculating request takes the time of the waiting and CPU time.

    (ii)   It uses response time as a parameter.

    (iii)   The server stay only when the profit is high, or else proceeds forage by indicating that whether the state is loaded, overloaded or under loaded.

Exponential Smoothing Based Weighted Least Connection

Exponential Smoothing Forecast supported Weighted-Least association (ESBWLC) [3][14][15] – This algorithmic program ESBWLC builds the choice supported the expertise of the node's electronic equipment power, memory, variety of connections and therefore the quantity of space presently being employed. It takes advantage of all historical knowledge and distinguishes them through the smoothing issue to let recent knowledge create a larger impact on the prophetical worth than long

knowledge. ESBWLC then predicts that node is to be hand-picked supported exponential smoothing.

Weighted Active Monitoring Load Balancing

Jasmin James et. al [17] projected this technique that may be a combination of Weighted Round Robin and Active Monitoring Load leveling algorithmic program. Different weights are allotted to VMs counting on the accessible processing power of the VM. The next VM allocation is done by choosing among the smallest amount loaded VMs the tasks are allotted to the foremost powerful one in keeping with their weights. During this method it removes the shortcomings of Active Monitoring Load Balancing algorithmic program by not solely considering the load however additionally the process power of accessible VMs.

TABLE 1: Advantage and Disadvantages of Load Balancing Algorithm

S.no.	Algorithm	Description	Advantage	Disadvantage
1	Round Robin	Request is allocated for a fixed period of time	Equal distribution of work load	Process is not known in advance. For a larger task context switching increases.
2	Weighted Round Robin	According to processing capacity of VM weight is assigned.	Optimal resource utilization	Processing time not taken into consideration
3	Dynamic Round Robin	It maintains VM retiring state and VM threshold state	Cost of power consumption get reduced	Does not scale up for large data center
4	Throttled LB Algorithm	Maintain state of VM busy or idle	Evenly distribution of load	Does not consider current state of the VM
5	Central Load Balancer	Maintains list of all available VM and their state	Load is balanced centrally	Fixed priority
6	Active Monitoring Load Balancing	Least loaded VM is allocated with the request	Existing load is taken into consideration	VM processing power is not considered
7	VM Assign LB Algorithm	VM is allocated as and when available	Proper VM utilization	NA
8	Weighted Active Monitoring LB Algorithm	Weights are assigned to the VM according to their processing power.	Consider weight and processing power of VM	Complexity increases
9	Min-Min Algorithm	Select task with least execution time	Simple to execute	Does not consider existing load
10	Load Balancing Improved Min- Min LBIMM	Similar to Min- Min algorithm. From the all available task, the task with smallest completion time from the most heavily-loaded resource is calculated. If it is less than makespan of Min-Min then the task is reassigned to the resource that produces it. The same process is repeated.	Overall completion time is reduced	Does not consider priority
11	Max Min Algorithm	Job with higher execution time executed first.	Shorter makespan as compared to Min- Min	Shorter jobs have to wait
12	User Priority Awarded Improved Min- Min	Divides the task in two groups according to user priority VIP an ordinary.	Consider priority and makespan	No deadline
13	Opportunistic Load Balancing (OLB)	Uses static load balancing algorithm attempt to allocate selected job available VM	Keeps all available VM busy	Does take into account the previous load
14	Honey Bee Foraging	Distributed load balancing for self-organization	Well suited for heterogeneous environment	Increase in resource not increases efficiency
15	Weighted Least Connection	Assign task to the node having the least number of connection	Balances load efficiently	Processing speed not considered
16	Exponential Smoothing Forecast based WLC (ESBWLC)	Task are assigned according to the processing power and memory of the node.	Each is examined	Complex calculations

TABLE 2: Load Balancing Environment for Different Algorithms

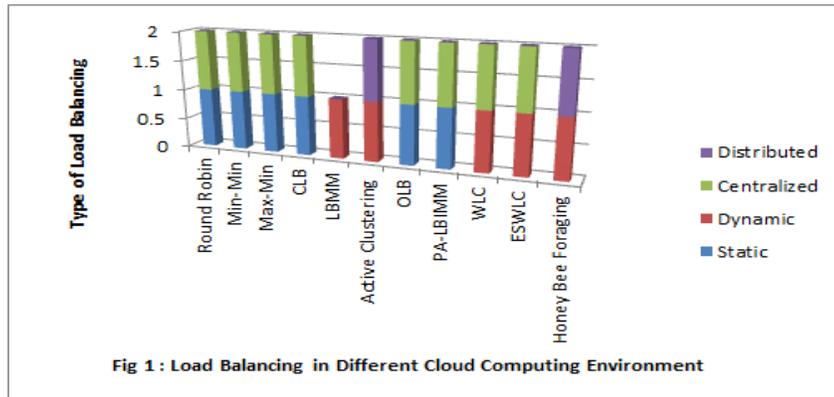| Algorithm	Static Load Balancing	Dynamic Load Balancing	Centralized Load Balancing	Distributed Load Balancing
Round Robin	True	False	True	False
Min-Min	True	False	True	False
Max-Min	True	False	True	False
CLB	True	False	True	False
LBMM	False	True	False	False
Active Clustering	False	True	False	True
OLB	True	False	True	False
PA-LBIMM	True	False	True	False
WLC	False	True	True	False
ESWLC	False	True	True	False
Honey Bee Foraging	False	True	False	True

Fig 1: Type of Load Balancing approach used in different algorithms

IV. LOAD BALANCING APPROACH

Virtual Machine (VM) migration [18] technique can be used for load balancing at the server. Dynamic movement of virtual machine in the virtualized server can give more options for load balancing.

A. Non Live VM Migration

It is the simplest migration technique which requires suspending and resuming the execution of VM before and after migration respectively.

B. Partial Migration

Partial Migration is the type of migration where only part of VM image is copied to the destination machine.

C. Local Area Live VM Migration [18]

This is most common type of VM migration it has two approaches for the VM migration Post-copy and Pre-copy.
(i) Post-copy migration transfers a VM 's memory contents after its processor state has been sent to the target machine.
(ii) Pre-copy migration first copies the memory state to the destination, through a repetitive process, after which its processor state is transferred to the target machine.

D. Live Storage Migration

Storage migration is required when Storage Area Network (SAN) is not present. Storage migration can be performed snapshotting, dirty block tracking and IO mirroring

E. Network Connection Migration

A VM may migrate to multiple network connections simultaneously. Virtualization technologies such as VMware and Xen provide live migration capabilities

V. CONCLUSION

The load balancing in cloud computing has great importance. Load balancing helps to improve the system performance. The paper summarizes different cloud computing algorithm and their advantages and disadvantages. The Existing load balancing techniques that have been discussed mainly focus on reducing associated response time, increasing throughput, reducing make-span, and improving performance by some other parameters as CPU, Memory and disk. The paper summarizes wit VM migration techniques used for load balancing in cloud computing.

REFERENCES

[1] Peter Mell and Timothy Grance (September 2011). The NIST Definition of Cloud Computing (Technical report). National Institute of Standards and Technology: U.S. Department of Commerce. doi:10.6028/NIST.SP.800-145. Special publication 800-145.

[2] Raza Abbas Haidri, C.P. Katti, P.C. Saxena, "A Load Balancing Strategy for Cloud Computing Environment", 2014 IEEE International Conference on Signal Propagation and Computer Technology (ICSPCT), Ajmer, India, pp. 636-641.

[3] Ren, X., R. Lin and H. Zou, "A dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast", International Conference on. Cloud Computing and Intelligent Systems (CCIS), IEEE sept. 2011, pp: 220-224

[4] J. Prassanna, P. Ajit Jadhav, V. Neelanarayanan, "Towards an Analysis of Load Balancing Algorithm to Enhance Efficient Management of Cloud Data Centres", 3rd International Symposium on Big Data and Cloud Computing Challenges ISBCC – 2016, VIT University, Chhenai, India, pp 143-155, Springer International Publication, DOI : 10.1007/978-3-319-30438-2-13.

[5] Mohammadreza Mesbahi, Amir Masoud Rahmani, "Load Balancing in Cloud Computing: State of art Survey, MECS, pp. 64-78, March 2016

[6] R. N. Carlheiros, Rajiv Ranjan et al., "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", Wileyonlinelibrary.com, August 2010, DOI: 10.1002/spe.995

[7] M. Katyal, A. Mishra, "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment", International Journal of Distributed and Cloud Computing, Vol. 1, Issue 2, December 2013.

[8] S.B. Shaw, A.K. Singh, "A Survey on Scheduling and Load Balancing Technique in Cloud Computing Environment", 2014 International Conference on Computer and Communication Technology (ICCCT), Allahabad, 2014, pp. 87-95.

[9] Geethu Gopinath P P, Shriram K Vasudevan "An in depth analysis of load balancing technique in cloud computing environment", 2nd International Symposium on Big Data and Cloud Computing Challenges, VIT University, Chennai, India, ISBCC-2015, pp. 427-432, 2015.

[10] S.Domanal, G. Reddy, "Optimal Load Balancing in Cloud Computing by Efficient Utilization of Virtual Machines", IJATES, Vol 3, issue 2, 2014, pp.122-129

[11] Huankai Chen, Professor Frank Wang, Dr Na Helian, Gbola Akanmu, "User-Priority Guided Min-Min Scheduling Algorithm

For Load Balancing in Cloud Computing", IEEE, Bangalore, 2013, pp. 1-8

[12] S. Wang, K. Yan, W. Liao, and S. Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, September 2010, pages 108-113.

[13] Gulshan Soni, Mala Kalra, "A Novel Approach for Load Balancing in Cloud Data Center", 2014 IEEE International Advance Computing Conference (IACC)*, Gurgaon, India, 2014, pp. 807-812.

[14] Bhushan Ghutke, Urmila Shrawankar, "Pros and Cons of Load Balancing Algorithms for Cloud Computing", 2014 IEEE, International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, 2014, pp. 123-127.

[15] A. Kumar and M. Kalra, "Load balancing in cloud data center using modified active monitoring load balancer," 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Spring), Dehradun, 2016, pp.1-5

[16] Martin Randles, David Lamb, 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

[17] Jasmin James, Dr. Bhupendra Verma, "Efficient VM Load Balancing Algorithm for a Cloud Computing Environment", International Journal on Computer Science and Engineering (IJCSE) , Vol. 4 No. 09 Sep 2012 , pp. 1658-1663.

[18] Boutaba, R., Zhang, Q., & Zhani, M. F. (n.d.). Virtual Machine Migration in Cloud Computing Environments. In Communication Infrastructures for Cloud Computing (pp. 383–408). IGI Global. https://doi.org/10.4018/978-1-4666-4522-6.ch017