



GNITED MINDS
Journals

*International Journal of
Information Technology
and Management*

*Vol. X, Issue No. XV,
May-2016, ISSN 2249-4510*

“SURVEY OF LOAD BALANCING IN CLOUD COMPUTING”

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

“Survey of Load Balancing in Cloud Computing”

Umesh Lilhore¹ Dr. Santosh Kumar²

¹Ph.D. Scholar MUIT Lucknow, Department of Computer Science & Engg. umeshlilhore@gmail.com

²Ph.D. Supervisor MUIT Lucknow, Department of Computer Science & Engg. sant7783@hotmail.com

Abstract – In now these days Cloud computing is a widely used technology over the internet. Cloud services and its consumers are growing rapidly. Cloud consumers are now demanding more services from cloud service provider. Cloud services are based on pay per use concepts. Performance of cloud computing depends on various factors; load balancing is one of them. In cloud computing proper load balancing are required for efficient execution of cloud services. So it attracts cloud researchers to develop more efficient load balancing method. In this research paper, we investigate the various loads balancing and task scheduling algorithms in Cloud Computing. We discuss and compare these existing methods provide a complete overview of the latest approaches in the field.

Keywords – Cloud Computing, Load Balancing and Types of Load Balancing Methods

----- X -----

1. INTRODUCTION

Cloud Computing is made up by aggregating two terms in the field of technology. First term is Cloud and the second term is computing. Cloud is a pool of heterogeneous resources. It is a mesh of huge infrastructure and has no relevance with its name “Cloud”. Infrastructure refers to both the applications delivered to end users as services over the Internet and the hardware and system software in datacenters that is responsible for providing those services.

In order to make efficient use of these resources and ensure their availability to the end users “Computing” is done based on certain criteria specified in SLA. Infrastructure in the Cloud is made available to the user's On-Demand basis in pay-as-you-say-manner. Computation in cloud is done with the aim to achieve maximum resource utilization with higher availability at minimized cost.

1.1 Cloud computing Vs cluster computing and Grid Computing-

Clusters (Sheeja and Jayalekshmi, 2014) are parallel and distributed systems, governed under the supervision of single administrative domain. The node (stand-alone computers) in the cluster integrates to form a single computing resource. Grid (Sheeja and Jayalekshmi, 2014) is aggregation of autonomous resources that are geographically distributed. The nodes in grid permit sharing and selection dynamically at runtime. Clouds (Sheeja and Jayalekshmi, 2014) (Domanal and Reddy, 2014) are not the combination of clusters and grid but are next generation to clusters

and grid. Similar to Cluster and Grid, Cloud is also a collection of parallel and distributed systems. Cloud is not a single domain. Unlike cluster and grid, cloud has multiple domains and the nodes of cloud are “Virtualized” (Ray and Sarkar, 2012).

2. CLOUD PERSPECTIVES

Cloud has different meaning to different stakeholders. There are three main stakeholders of cloud.

2.1 End users –These are the customers or consumers of cloud. They use the various services (Infrastructure/ Software/Platform) provided by the cloud.

2.2. Cloud Provider-Cloud provider can offer either public or private or hybrid cloud. They are responsible for building of the cloud.

2.3. Cloud developer-This entity lies between end user and cloud provider. Cloud developer has the responsibility of taking into consideration both the perspectives of the cloud. The developer of cloud must adhere to all the technical details of the cloud which are essential to meet the requirements of both, the cloud user as well as the cloud provider.

3. LOAD BALANCING IN CLOUD COMPUTING

Load balancing is a method for reassigning the entire workload to the separate nodes for a group of systems, inorder to make the time more efficient and

utilizes the resources in efficient manner (Sheeja and Jayalekshmi, 2014). Load balancing is the basic technique for scaling out an application server infrastructure. As request demand increases, new servers can be added to the resource pool, were the load balancer will directly begin sending traffic to the new server. There are two different kinds of algorithm in load balancing. They are static and dynamic algorithm.

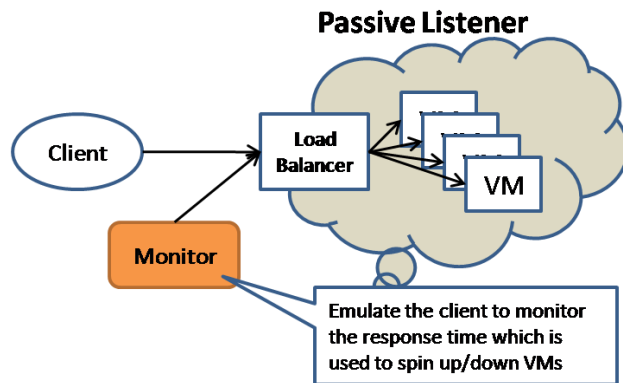


Figure 2 Load balancing

A. Static Algorithms- Static algorithms partition the traffic uniformly among the servers. In this approach, the traffic on the servers will be contemptibly easy and subsequently it will make the circumstances imperfectly (Domanal and Reddy, 2014).

B. Dynamic Algorithms- A dynamic algorithm selects the appropriate weights on server, by searching through the whole network and prefers the lightest server to balance the traffic. But selecting suitable server needs valid communication within networks that lead to extra traffic being added on system (Domanal and Reddy, 2014).

3.1 Benefits of Load Balancing-

- **Redundancy** It describes the process of running two or more, the same servers thus providing a guaranteed event that one server becomes occupied.
- **Scalability** Even though modest resources requirements are offered, scalability must always be considered for finding the correct host solution.
- **Resource Optimization-** Through load balancing, one can optimize how traffic is circulated to the server cluster so, that it guarantees the best performance.
- **Security** In security, only one IP is exposed to the web with load balancing, which significantly reduces the amount of break points in case of attack.

3.2 Resource allocation- Resource provisioning is the task of mapping of the resources to different entities of cloud on demand basis. Resources must be allocated in such a manner that no node in the cloud is overloaded and all the available resources in the cloud do not undergo any kind of wastage.

Mapping of resources to cloud entities is done at two levels-

3.2.1 VM mapping onto the Host-

Virtual machines reside on the host (physical servers). More than one instance of VM can be mapped onto a single host subject to its availability and capabilities.

3.2.2 Application or Task Mapping onto VM-

Applications or tasks are actually executed on VM. Each application requires certain amount of processing power for their completion.

3.3. Task Scheduling-

Task scheduling is done after the resources are allocated to all cloud entities. Scheduling defines the manner in which different entities are provisioned.

Task scheduling provides “Multiprogramming Capabilities” in cloud computing environment. Task scheduling can be done in two modes-

- a. Space shared b. Time shared

Both hosts and VM can be provisioned to users either in space shared mode or time shared mode.

Task	Sub-Category	Issues Resolved	Provider oriented	Customer oriented
Resource Allocation	At host level At VM level	Efficient Utilization Minimize	YES	YES
Task Scheduling	Space-Sharing Time-Sharing	Minimize overall response time	NO	YES

Table 2.2 Comparison between Resource Allocation and Task Scheduling

4. LOAD BALANCING METHODS

Following methods are uses in load balancing-

4.1 A. Round Robin Algorithm- Round Robin is one of the static load balancing algorithms, where preceding states are not taken into account. It is simple and uses the Round Robin Method for job allocation. It selects the first node at random and then allocates the job to all the other nodes

evenly in Round Robin Method. The main advantage of Round Robin is that it does not need any intercrosses communication. There is no prior information about the processors' running time, so that some tasks may get heavily loaded. To overcome this, weighted Round Robin algorithm is being proposed. Here each node assigned has a specific weight. Based on the nodes weight, they would get the requests. If all nodes are equal, then the node is indicated to traffic.

- B. **Opportunistic Load Balancing Algorithm-** It is also one of the static load balancing algorithms, which do not consider the present workload of the VM. It usually keeps each and every node busy. This deals with the unexecuted tasks quickly and in random order to the current node, where each one of task is assigned to the node randomly. This algorithm provides a load balancing schedule but does not produce a good result. The tasks are processed in a slow manner, where the current execution time of the node is not calculated.
- C. **Min-Min Load Balancing Algorithm** Min-Min is a static load balancing algorithm, where the parameters associated to the job are recognized in advance. In Min_Min algorithm, the execution and completion time of the unassigned waiting in queue are identified by the cloud manager. The jobs with minimum execution in time are being assigned first to the processors, so that the task is completed in time. But the tasks with maximum execution need to wait for a specific period of time. As such, all the all the tasks in the processor must be updated and the tasks in the queue must be removed. The task with minimum time execution performs better than the maximum time execution. The main disadvantage of this algorithm is that it leads to starvation.
- D. **Max-Min Load Balancing Algorithm** Max-Min is same as the Min-Min algorithm, where the maximum time jobs are selected. Only once the minimum time jobs are completed. When the minimum jobs are completed, the tasks that are in the queue are assigned to the processor. Execution time of all tasks is being updated to the processor. Since it is a static algorithm, the time of every task is calculated in advance and performed in a correct manner. An improved version of max-min algorithm was projected in (Soni, Kalra, 2014).
- E. **Ant Colony Optimization Based Load Balancing Algorithm** This algorithm is designed to seek out the optimal path among the food and colony of ant, based on its

actions. The main aim of this approach is to distribute the work load among the nodes in an efficient manner. The regional load balancing node is preferred as head node in Cloud Computing Service Provider (Babu, Krishna, 2013). The ants collect the information from the cloud node and assign the tasks to the particular node. Once the task is assigned to the head node, the ant moves in a forward direction with the overloaded node to the next node checking whether the node is overloaded or not. Once the job gets successful it is updated, then the result is reported based on the individual result of the ant. After receiving the individual result they are combined together to build the complete report. The solution set is updated automatically, when the ant updates the result for every movement. To prevent backward movement, the ant commits suicide when it reaches the target node.

- F. **Honeybee Foraging Load Balancing Algorithm** It is one of the dynamic load balancing algorithms, where they are designed based on the behaviour of honey bees. Honey bees have been classified into two types. They are: finders and reapers. The finder honeybee helps in finding the honey source. Once honey source is found, they do the waggle dance to indicate the quality and quantity of available honey. After that, the reapers gather the honey from the sources. Then, again they go for the waggle dance to specify the honey that is left. In load balancing, the servers are combined together as virtual servers, where each and every virtual server has a process queue. Once the request is received from the queue, it calculates the profit quality as the bee does in waggle dance (Babu, Krishna, 2013).

The server stays only when the profit is high, or else proceeds to forage by indicating that whether the state is loaded, overloaded, under loaded or balanced. Based on this, the current virtual machines are combined. It needs to maintain a separate queue for each and every node. Depending on the priority, the task is taken into concern, by removing the task that is waiting in the overloaded machine. The tasks removed are loaded into lightly loaded machine. Those tasks are known as scout bee for the next step (Babu, Krishna, 2013). The behaviour of honey bee in load balancing technique has stimulated to reduce the response time of virtual machine, which also reduces the waiting time. The main disadvantage of this algorithm is, it does not show any improvement in throughput.

- G. **Biased Random Sampling Load Balancing Algorithm** Biased Random Sampling is a

dynamic load balancing algorithm. Here, random sampling method is being used to achieve the load balancing across all the nodes. In this algorithm, all the servers are treated as nodes (Shaw, Singh, 2014). This method is represented in the form of virtual graph, constructed with the connectivity which represents the load on each node. Each node is taken as vertex in a directed graph. When a request is received from the client to the load balancer, the load balancer assigns the job to the node that has a minimum of one in-degree. Once a job is assigned to the node, the server starts executing the job, indicating the reduction in availability of free resources. After the completion of the job, the node gets incremented by one in-degree, indicating the increase in available resources. The addition and deletion of such processes are completed by the process of random sampling technique (Babu, Krishna, 2013).

Threshold value is used as a parameter that considers each and every process by representing the maximum walk length. The traversal is from one node to another node until finding a designation is known as a walk. After receiving the request from the load balancer, it compares the current node to the randomly selected node with the threshold value. If the threshold value is equal or greater than the current walk length, the node executes its job, or else it moves to another neighbour node that is randomly selected. The performance decreases as the number of servers increases.

H. Active Clustering load balancing Algorithm
Active Clustering is an improved method of random sampling. The concept of clustering is used in this algorithm. The main principle of this algorithm is grouping similar nodes together, and working based on those grouped nodes. Grouping of nodes helps the resources to increase the throughput efficiently. In this algorithm, a method called matchmaker is introduced (Shaw, Singh, 2014).

While an execution starts, the first node selects the neighbour node. The neighbour node is taken as match make node, which connects the neighbour node that is same as initial node. At last the match maker node gets disconnected. And this process is done iteratively to balance the load equally. The system performance is improved highly, by increasing the throughput. There is an efficient utilization of resources when there is an increase in throughput (Babu, Krishna, 2013).

Load Balancing Algorithm	Static or Dynamic	Job Allocation	Advantage	Disadvantage
Round-Robin	Static	Selects the First Node in Random Manner, and allocates other nodes in Round	Treats the entire server equally.	Any process is not known in advance.
Opportunistic Load Balancing Algorithm	Static	Based on the framework of the system.	Keeps every node busy.	The execution time is completed, but the node is still busy.
Min-Min Load Balancing Algorithm	Static	Identifies and completes the job waiting in queue.	Performs better small execution time.	Leads to Starvation.
AntColony Optimization Based Load Balancing Algorithm	Static	Based on actions of ants and seeking an optimal path in collecting their food	Distributes the workload among nodes in efficient and optimal job scheduling is	-
Honeybee Foraging Load Balancing Algorithm	Dynamic	Based on the behaviour of honeybees' and their approach in collecting honey	The response time and waiting time of the virtual machine are reduced.	There is a decrease in the throughput, when there is an increase in resources
Biased Random Sampling Load Balancing Algorithm	Dynamic	Based on Random Sampling Method.	Load balancing is achieved through all the nodes in the system.	Corrupts when load increases
Active Clustering load balancing Algorithm	Dynamic	Grouping nodes together.	Similar nodes are grouped together	Poor performance when there is an increase in variety of nodes

Table 3 comparison of load balancing algorithms in cloud computing

5. CONCLUSION

The load balancing of the current system is one of the greatest issues. Various techniques and algorithms are used to solve the problem. In this paper we survey various existing load balancing methods in different environments. A large number of parameters and different types of soft computing techniques can be included in the future for the better utilization and needs of the user. The various load balancing techniques are also being compared here.

REFERENCES

- Ajay Gulati, Ranjeev K. Chopra (2013). "Dynamic Round Robin for Load Balancing in a Cloud Computing", IJCSMC, Vol. 2, Issue. 6, pp. 274 – 278.
- B. Wickremasinghe, R. N. Calheiros, R. Buyya (2010). "Cloud analyst: A cloudsim-based visual modeller for analyzing cloud computing in: Proceedings of the 24th International Conference on Advanced Information Networking and Applications (AINA 2010), Perth, Australia."

- Brototi Mondal, Kousik Dasgupta and Paramartha Dutta (2012). "Load balancing in cloud computing using stochastic hill climbing a soft computing approach" in *Procedia Technology* 4, pp. 783-789, ELSEVIER C3IT-2012.
- Chandrakanta Korat and Piyush Gohel (2015). "A Novel Honey Bee Inspired Algorithm for Dynamic Load Balancing In Cloud Environment", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, pp. 6995 – 7000.
- Dhinesh Babu L.D., P. Venkata Krishna (2013). "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Published in *ELSEVIER Applied Soft Computing* 13, pp. 2292–2303.
- Dinesh Babu L. D., P. Venkata Krishna (2013). "Honey Bee inspired Load Balancing of tasks in cloud computing environments", *Applied soft computing* 13, pp. 2292-2303.
- Gulshan Soni, Mala Kalra (2014). "A Novel Approach for Load Balancing in Cloud Data Center", 978-1-4799-2572-8/14, IEEE 2014, PP 807-813.
- Mr. Manan D. Shah (2013). "Allocation Of Virtual Machines In Cloud Computing Using Load Balancing Algorithm" in *International Journal of Computer Science and Information Technology & Security (IJSITS)*, ISSN: 2249-9555 Vol. 3, No.1.
- Ms. Nitika, Ms. Shaveta, Mr. Gaurav Raj (2012). "Comparative Analysis of Load Balancing Algorithms in Cloud Computing" in *International Journal of Advanced Research in Computer Engineering & Technology* Volume 1, Issue 3.
- Rakesh Kumar Mishra , Sreenu Naik Bhukya (2014). "Service Broker Algorithm for Cloud-Analyst", *International Journal of Computer Science and Information Technologies*, Vol. 5 (3) pp. 3957-3962.
- Ranjan Kumar, G. Sahoo (2014). "Cloud Computing Simulation Using CloudSim", *International Journal of Engineering Trends and Technology (IJETT)* – Volume 8 Number 2, pp. 82-87.
- S. G. Domanal and G. R. M. Reddy (2014). "Optimal load balancing in cloud computing by efficient utilization of virtual machines," *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, pp. 1-4.
- Shridhar G. Domanal and G. Ram Mohana Reddy (2013). "Load Balancing in Cloud Computing Using Modified Throttled Algorithm", *IEEE, International Conference. CCEM 2013*. In press.
- Soumya Ray and Ajanta De Sarkar (2012). "Execution analysis of load balancing Algorithms in cloud computing Environment", *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, Vol. 2, No.5, pp. 78-98.
- Soumya Ray and Ajanta De Sarkar (2012). "Execution analysis of load balancing Algorithms in cloud computing Environment", *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, Vol.2, No.5, pp. 183-196.
- Subhadra Bose Shaw, Dr. A.K. Singh (2014). "A Survey on Scheduling and Load Balancing Techniques in Cloud Computing Environment", Published in *5th International Conference on Computer and Communication Technology (ICCCT)*, 978-1-4799-6758-2/14, IEEE 2014, PP 87-96.
- Tushar Desai, Jignesh Prajapati (2013). "A survey of various load balancing techniques and challenges in cloud computing", *International journal of scientific & technology research* volume 2, issue 11, pp. 158-162.
- Tushar Desai, Jignesh Prajapati (2013). A Survey of Various Load Balancing Techniques and Challenges in Cloud Computing", published in *International journal of scientific & technology research* volume 2, issue 11, pp. 158-162.
- Y. S. Sheeja and S. Jayalekshmi (2014). "Cost effective load balancing based on honey bee behaviour in cloud environment," *Computational Systems and Communications (ICCSC)*, 2014 *First International Conference on*, Trivandrum, pp. 214-219.