

Assorted Load Balancing Algorithms in Cloud Computing: A Survey

Priyanka Singh
P.S.I.T.
Kanpur, U.P. (208020)
A.K.T.U. Lucknow

Palak Baaga
P.S.I.T.
Kanpur, U.P.(208020)
A.K.T.U. Lucknow

Saurabh Gupta
P.S.I.T.
Kanpur, U.P.(208020)
A.K.T.U. Lucknow

ABSTRACT

The hidden meaning of Cloud Computing is entering or storing or sharing the data on the internet. In accustomed scheme, cloud computing is spreading and growing day by day and become most important factor in industries both government or private industries. An extreme concern in cloud computing paradigm is proper load balancing over the attainable resources. Assorted Load balancing algorithm and mechanism has been developed to make efficient utilization of attainable resources and improve the overall performance.

In this paper our aim is to explore assorted algorithms to find out the solution of load balancing in cloud computing and after that compare them on various parameter. Also discuss the merit and demerits of the algorithms.

Keywords

Cloud computing, load balancing, resources, server, node.

1. INTRODUCTION

At this time, cloud computing accommodates reliable and extensible way to contain the data and file, that relate virtualization, distributing and web services or storing.

The principle of cloud computing is to provide high quality services with less time and low cost. The definition of cloud computing provided by National Institute of Standards and Technology (NIST) says that: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, data storage, software applications and other computing services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1]".

The following figure1 shows the scenario of the cloud computing: Various devices such as tablet, PCs, laptops connect and access the data at any given point of time from a cloud. The main objectives of cloud are to reduce cost, enhance response

Time, provide better performance and quality services, hence Cloud is also called a pool of service [2].



Figure1. A Cloud Computing Scenario [2].

Cloud Computing services has been divided into three types which are as follows:

1.1 Infrastructure as a service (IaaS)

IaaS permits organizations to outsource computing equipment and resources such as storage, server, networking as well as services, such as load balancing and content delivery networks. IaaS is a cloud service model. Users could deploy and control all the things in virtual machines except the real hardware where virtual machines are deployed in.[3]

Currently, IaaS providers are: Amazon AWS, windows Azure, Rackspace Open Cloud, IBM Smart Cloud Enterprise, HP Enterprise Converged Infrastructure and many more.

1.2 Platform as a service (PaaS)

PaaS present a kind of platform as a service where OS, DB, Runtime Environment, web server are already given. Users could only deploy and control their own software and the runtime needed [3].

Currently, PaaS providers are: Google App

Engine, Red Hat Open Shift, Engine Yard, AppFog, Windows Azure Cloud Services and many more.

1.3 Software as a service (SaaS)

SaaS present a cloud service model that transfers on demand applications that are presented and managed by the service provider.

SaaS providers develop oneself in all things from productivity and, network monitoring and customer services, project management to security.

Currently, SaaS providers are: Google Apps, Microsoft Office 365.

Following figure2 shows the scenario of the cloud service models.

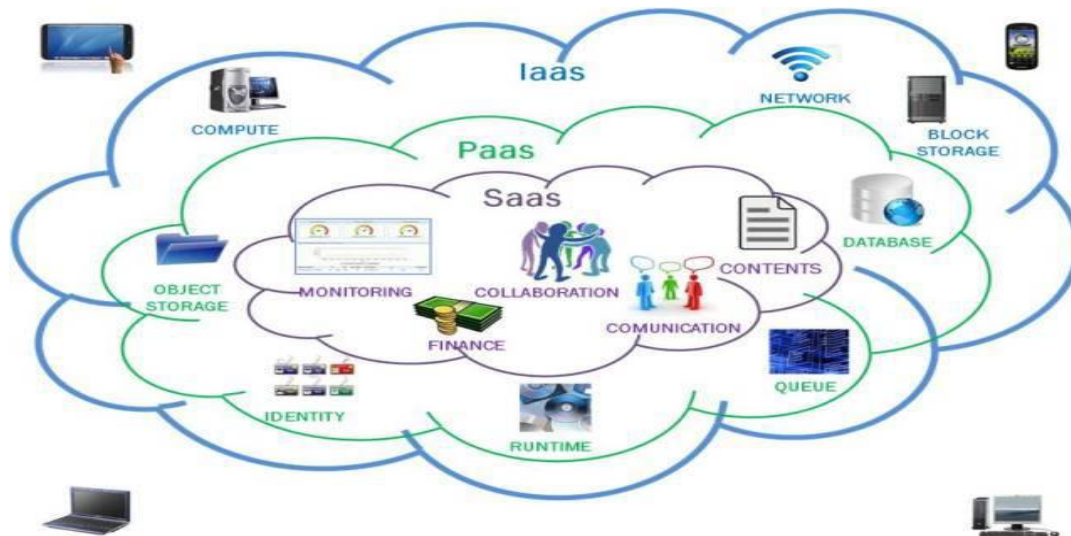


figure 2: cloud computing environment[4]

In this paper, we essentially focus on inclusive review of interactive load balancing algorithms in cloud computing. Address each and every problem in existing load balancing algorithm and compare them with each other.

In section 2 we discuss exactly what is load balancing and why we need load balancing and its advantage. In section 3 we discuss all the existing load balancing algorithm in cloud computing. In section 4 we discuss the summary table of algorithms. In section 5 we discuss the parameter table of all the algorithms. In section 6 we discuss the conclusion.

2. LOAD BALANCING

Load balancing is the mechanism of allocating and reallocating the load among available resources to improve the performance and maximise the throughput at minimum response time and minimum cost.

Load balancing has two meanings: first, it puts a large number of concurrent accesses or data traffic to multiple nodes respectively to reduce the time users waiting for response; second, it put the calculation from a single heavy load to the multiple nodes to improve the resource utilization of each node.[5]

2.1 Advantages of Load Balancing

Some major advantages of load balancing [6, 7] are as follows:

- a) Reduces waiting time.
- b) Minimizes response time.
- c) Maximizes utilization of Resources.
- d) Maximizes throughput.
- e) Improves reliability, and stability of the system.
- f) Accommodates future modification.
- g) Long starvation is avoided for small jobs.
- h) In load balancing overall system performance is enhance by improving the performance of each node.

In summary, the load balancing algorithms and mechanisms in cloud computing environment requires following conditions:

- a) a)When the local load is not very heavy so that it can make the local self organization to reduce the information exchange.
- b) Load balancing mechanism in cloud computing can be used in Distributed Environment.
- c) Load balancing mechanism in cloud computing should increase the throughput of the system at best without affecting the response time.
- d) Load balancing in cloud computing could be used in monitoring processor load, state and correct data movement.
- e) Exchanging load and state information between processors.

3. EXISTING LOAD BALANCING ALGORITHM IN CLOUD COMPUTING

Load Balancer develops the type of algorithms which is based on load balancing decision. These decisions are how to spread the load among nodes so that all users attain quick and better quality services. Load balancing is one of the demanding and challenging task in cloud computing.

Load balancing algorithm is primarily divided into two types: the static load balancing and dynamic load balancing. For the ease of understanding and better assimilation we discuss these algorithms one by one.

3.1 Static Load Balancing

In static algorithm the traffic is divided evenly among the servers. This algorithm requires a prior knowledge of system resources, the performance of the processors is determined at the beginning of the execution, therefore the decision of shifting of the load does not depend on the current state of system [8].

Advantage of static load balancing is: very easy and simple to understand and implement.

The major disadvantage of static load balancing is: at the time of execution task can not be moved to other system.

3.2 Algorithms that come under static load balancing

Some of the best algorithm of static load balancing in cloud computing are as follows:

3.2.1 Weighted Round Robin Algorithm

Weighted Round robin algorithm is the modified version of the round robin algorithm. In this algorithm, a weight is assigned to each and every server or node and distribution of the job is depends on the values of the weights. Larger capacities processors have the larger value of weight. In a situation where all weights become equal, servers will receive balanced traffic.

3.2.2 Map Reduce Algorithm

Map Reduce Algorithm has two general objectives – the first objective is to properly mapping the task and the second objective is to reducing the load after mapping. Three methods are included in this algorithm. The three methods are part, comp and group. In Map Reduce the part method is initiate the Mapping of tasks. In this step the request is partitioned into different parts. Then, the key of each part is saved into a hash key table. The comp method does the comparison between the parts. The next method is group method in which it groups the parts of similar entities using the Reduce tasks [9]. The main disadvantage of this algorithm is Overloading in the system.

3.2.3 Min-Min + Max-Min Algorithm

In Min-Min algorithm, among all the least time consuming task is searching the first step the task is arranging, according to that smallest time value on the machine. The running time for other tasks is also updated [10]. In this algorithm all the task are waiting in the queue. Firstly the smallest time consuming task is assigned to the processor and the largest time consuming task wish to wait for a particular time period. Major issue of this algorithm is the Starvation.

The terminology related to static load balancing for Min-Min is [11]

- Excepted Time of Compute (ETC) - The running time excepted for tasks in all the nodes are stored in ETC,
- Minimum Execution Time Algorithm (MET) – It finds the best job-processor-pair, were current load is not considered,
- Minimum Completion Time Algorithm (MCT) –It allocates the tasks based on the minimum completion time.

Max-Min algorithm is similar to Min-Min algorithm but max-min chooses the task with largest value and gives to the respective processor after assign the task, processor works according to updates .These assigned tasks remove from the queue [10] .Since it is a static algorithm, the time of every task is calculated in advance and performed in a correct manner.

3.2.4 Ant Colony Algorithm

This algorithm is developed to find out the best path between the food and ant colony, this is done to equally spread the load between all nodes. The local node is referred as the head node. When the load is tagged to the head node , ant with loaded node start moving forward to find out the node either node is overloaded or not. Whenever ant find out any loaded node it move further and if find out any overloaded node it move in reverse direction to displace the previously found node. At

that time when all task completes then the result is modified or updated to build a final result report.

3.3 Dynamic Load Balancing

Dynamic Load Balancing is very efficient and precise. Dynamic algorithm does not require the knowledge about the previous state of the node it based on the current state of the node because of this the over all performance of the system is increase.

It allows processes to be moved from an over utilized machine to an underutilized machine dynamically for faster execution [4]. Dynamic load balancing algorithm be based on the combination of knowledge based on all gathered information about the nodes and different properties of the selected nodes process and task on that node in public and private cloud [9]

3.4 Algorithms that come under dynamic load balancing

3.4.1 Throttled Load Balancing Algorithm

Throttled Load Balancing Algorithm absolutely established on Virtual Machine (VM). In this algorithm, the throttled load balancer search the acceptable VM to execute the client operation, this is done only when client requested the load balancer to search the acceptable VM. The searching of acceptable VM become an issue of this algorithm, the issue is delay in operation. The throttled load balancer keep up the state (Available/Busy) along an index table of Virtual Machines.

3.4.2 Modified Throttled Load Balancing Algorithm

Modified Throttled Algorithm is an advanced form of throttled algorithm. Modified throttled algorithm keep up the state (Available/Busy) along an index table of Virtual Machines. VM at first index is initially selected depending upon the state. If VM is available, then the request is assigned and if VM is not found then it return (-1) to the Data Centre Controller. When the next request arrives, the VM next to the already allocated VM is chooses [12]. This work is occur again an again till reaches to the size of index table.

3.4.3 Neighbour Aware Random Sampling Algorithm:

In this algorithm, each node attempt to forecast the current load of its neighbours from past information and the smallest loaded node is picked for Random walk. There are basically two methods to pick the neighbour node for Random Walk. First one is based on the Probability and the second one is based on the smallest loaded node first. Both the method is based on the principle to find the smallest loaded node and maintain the current information about each neighbour node.

This algorithm is based on the following way:

1. When a node receives a RequestToken, it gathers load information of every neighbour that has participated in that walk and updates the knowledge base with timestamp [13].
2. When a node A relays the AllotmentToken or job details to node B and B is neighbour of A, then the load of that job is added to the known load of B in A's knowledgebase[13].

3.4.4 Dual Direction Download Algorithm (DDFTP)

In this algorithm, divide the size m file into $m/2$ partitions. This algorithm is basically handle the network communication among the client and nodes. For selecting the accurate node, the algorithm consider attributes such as network load, node load, network speed while no current information of node is required [14]. This algorithm is developed for private as well as private cloud computing.

3.4.5 Least Connection Algorithm

This algorithm is basically depends on the connection on a single node. The load balancer of this algorithm, firstly pick the node that have the smallest number of connection so load

can be transferred. The load balancer also maintains the set of connection on every node. Whenever a new connection is implemented the number of connection increases and whenever the old connection completes or time out action occur the number of connection decreases. Load balancer maintains all the increasing and decreasing connection related information [15].

This algorithm really work well when nodes having homogeneous capacities.

4. RESUME OR SUMMARY OF ALL THE ALGORITHMS

Following table shows the summary of the algorithms

Algorithms	How It Works	Advantages	Disadvantages
Weighted Round Robin Static Balancing Algorithm	Select the node on the bases of the weight.	Modified version of Round Robin. Work well for nodes Of different capacities.	Not work well if nodes have different execution time.
Map Reduce Static Load Balancing Algorithm	Based on the two method, first is to perform mapping and second is reduce the load.	Work Very efficiently The performance is very high.	Overloading is the main issue.
Min-Min Static Load Balancing Algorithm	Based on the least time consuming tasks. Complete all the tasks waiting in the queue.	Work very well for the task having least execution time.	Starvation is the main Issue.
Max-Min Static Balancing Algorithm	Based on the high time consuming tasks.	High efficiency and performance.	Waiting of high time consuming tasks is very long.
Ant Colony Static Load Balancing Algorithm	Based on to find the optimal path between food and ant colony.	All the information about nodes collected by ant is quite fast.	Over head on network is very high. Number of Ants does not doubtlessly define.

Throttled Dynamic Load Balancing Algorithm	Based on the Virtual Machines. All the Client tasks are performed by Virtual machines.	Very efficient and performance is really good.	Delay in operation due to long search to find acceptable Virtual Machines.
Modified Throttled Dynamic Load Balancing Algorithm	Bit Advance version of throttled algorithm It is also based o the Virtual Machines.	Performance is better then Throttled algorithm and more efficient.	Delay in operation.
Neighbour Aware Random Sampling Algorithm	Based on the neighbour nodes. The node is chooses on the bases of the random walk.	Communication delay is low as compare to other algorithms. Not centralized.	
Dual Direction Downloading Algorithm	Based on dividing the file into different partition and then each partition task	Download reliable files. Calculation is very fast. Reduce the communication delay.	High files storage requirements.
Least Connection Dynamic Load Balancing Algorithm	Based on the connection on every single node. Small connection node are picked	Work well for all node having same capacities.	Not work well for node having different time.

5. ANALYSIS OF ALGORITHM BASED ON THE FOLLOWING PARAMETERS OR METRICS

Now, we analyse all the discussed algorithm with the parameter such as: performance, complexity , scalability , fault tolerance , response time , throughput and many more.

5.1 Analysis of Static Load Balancing Algorithms

Algorithms	Weighted Round Robin	Map Reduce	Min-Min	Max-Min	Ant Colony
Metrics					
Performance	Fast	Fast	Fast	Fast	Slow
Complexity	Low	Low	Low	Low	No

Response Time	Fast	Fast	Fast	Fast	Slow
Fault Tolerance	No	No	No	No	No
Scalability	Low	Low	Low	Low	Low
Resource Utilization	High	High	High	High	High
Speed	N/A	Fast	Fast	Slow	Fast
Overhead	High	High	High	High	Low
Power Consumption	High	High	High	High	High
Waiting Time	High	High	High	High	High

5.2 Analysis of Dynamic Load Balancing

Algorithm Metrics	Throttled	Modified Throttled	Neighbour Aware Random Sampling	Dual Direction Downloading	Least Connection
Performance	Fast	Fast	Fast	Fast	Slow
Complexity	Low	Low	Low	High	High
Fault Tolerance	Yes	Yes	Yes	No	No
Scalability	High	High	High	High	Low
Resource Utilization	High	High	High	High	High
Speed	Fast	Fast	Fast	Slow	Slow
Overhead	Low	Low	Low	High	High
Power Consumption	Low	Low	Low	Low	Low
Waiting Time	Less	Less	Less	Less	Less

6. CONCLUSION

Cloud computing has the efficiency to properly balance the load using different algorithms. Load balancing is the main issue in cloud computing. In this paper, we discuss all the static and dynamic load balancing algorithm in cloud computing like as Weighted Round Robin, Map Reduce, Min-Min, Max - Min, Ant Colony, Throttled, Modified Throttled, Neighbour Aware Random Sampling, (DDFTP) Dual Direction Downloading and least connection. The main purpose of this paper is to summarize all assorted algorithm and compare all assorted algorithm with different parameter like as Response Time, Performance, Overhead, Waiting Time, Overhead, Scalability, Power Consumption, Speed.

There are a large number of future work can be done to properly balance the load in cloud computing. First is to develop the modified Throttled algorithm in the real world.

7. REFERENCES

- [1] Peter Mell, Timothy Grance, The NIST Definition of "Cloud computing", National Institute of Standards and Technology Computing Security Resource Center – www.csre.nist.gov.
- [2] Sidra Aslam, Munam Alisha, "Load Balancing Algorithm in Cloud Computing :A Survey of Modern Technology(NSEC2015)", of the IEEE Computer & Communication societies, IEEE2015.
- [3] Linlin Tanng, PingfeiRen, Jengshyang Pan, "An Improved K-Subset Algorithm For Load Balancing Problem In Cloud Computing", CCIS2014 of the IEEE2014.
- [4] Hamid Shoja, Hasein Nahid, Reza Azizi, "A Comparative Survey On Load Balancing Algorithms In Cloud Computing", ICCCNT2014 of the IEEE2014.
- [5] Haozheng Ren, Yihuanlan, Chaa yin, "The Load Balancing Algorithm In Cloud Computing Environment", 2nd International Conference On Computer Science and Network Technology of the IEEE2012.
- [6] S. F. El-Zoghdy and S. Ghoniemy, "A Survey of Load Balancing In High-Performance Distributed Computing Systems", International Journal of Advanced Computing Research, Volume 1, 2014.
- [7] Mr. Amitkumar S Manekar, Mr. Mukesh D Poundekar, Prof. Hitesh Gupta and Prof. Malti Nagle, "A Pragmatic Study and Analysis of Load Balancing Techniques In Parallel Computing", International Journal of Engineering Research and Application, Vol.2, Issue4, July-August 2012.
- [8] Nadeem Shah, Mohammed Farik, "Static Load Balancing Algorithm In Cloud Computing :Challenges and Solution", of the International Journal Of Scientific And Technology Research, Issue 10, October 2015.
- [9] Bhushan Ghutke, Urmila Shrawankar, "Pros And Cons Of Load Balancing Algorithm For Cloud Computing", 2014 International Conference On Information System And Computer Networks of the IEEE2014.
- [10] Sidra Aslam, Munam Alisha, "Load Balancing Algorithm in Cloud Computing :A Survey of Modern Technology(NSEC2015)", of the IEEE Computer & Communication societies, IEEE2015.
- [11] D.Saranya, L Sankara Maheswari, "Load Balancing Algorithms In Cloud Computing: A Review", International Journal Of Advanced Research In Computer Science And Software Engineering, Issue 7, July2015.
- [12] Priyadarashini Adyasha Pattanaik, Sharmistha Roy, Prasant Kumar Pattnaik, "Performance Study Of Some Dynamic Load Balancing Algorithms In Cloud Computing Environment", 2nd International Conference On Signal Processing And Integrated Networks (SPIN), of the IEEE2015.
- [13] Ariharan V, Sheeja S Manakattu, "Neighbour Aware Random Sampling (NARS) Algorithm For Load Balancing In Cloud Computing", of the IEEE2015.
- [14] Al-Jaroodi, J. and N. Mohamed. "DDFTP: Dual-Direction FTP," in proc. 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), IEEE, pp:504-503, May 2011.
- [15] Suriya and Prashanth, "Review of load balancing in cloud computing", IJCS, vol.10 issue.1, Jan 2013.