

A Survey On Load Balancing Methods and Algorithms in Cloud Computing

M. Lagwal^{1*}, N. Bhardwaj²

¹Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, INDIA

²Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, INDIA

Corresponding Author: monikalagwal67@gmail.com

www.ijcseonline.org

Received: 23/Mar/2017, Revised: 04/Apr/2017, Accepted: 20/Apr/2017, Published: 30/Apr/2017

Abstract—the IT trade has played a noteworthy role in computing. Due to this increasing of demands day by day the need of computing and storage is increasing quickly. Client-Users are demanding for services and resources at any time are provided. For this cloud computing requires load balancing techniques to control and handle overloaded demand and requires. Load balancing is one of techniques to control lots of requires at a time and help to utilization of resource and services. In load balancing various algorithm are provide to the user for satisfaction. In this paper, we have done the review of some load balancing algorithms based on different parameters in cloud computing.

Keywords— Cloud Computing, Load Balancing, Virtualization, Load balancingalgorithm , Ant-colonyalgorithm, Genetic algorithm

I. INTRODUCTION

Cloud computing is a technology based on internet, in which sharing of information is done through internet. Cloud computing providers have set up data-centers at different-different locations over geographical places on the web. Today this technology rising at a quick rate and providing high performance, efficiency and reduces the cost. The cloud computing gives large amount of information, services and storage spaces in cloud environment. Nowadays load balancing become very important feature in cloud computing. In load balancing the process are distributed workloads and computing resources in cloud computing environment among their nodes and handles all the requests at a time. This provides better performance during traffic in cloud system.

II. CLOUD COMPUTING CHARACTERISTICS

Cloud computing characteristics are as following [1-13]:

1. Service provide on demand - When any user want services and resources then the cloud provide services on demand.
2. Rapid Elasticity - Number of resources in cloud can be increase and decrease easily.
3. Resource pooling - Resources are allocated at different – different location according to user’s requirement.

4. Pay per use - According to the client’s utilization of computing resources is charges to paid [2].

III. SERVICES OF CLOUD COMPUTING

Cloud computing provides following services:

1. Software as a service (SAAS)

SAAS provide those services where user can access software and applications over the internet.
Example- salesforce.com, Google-docs, Google-mail, face book.

2. Platform as a service(PAAS)

PAAS accommodate all the services and resources to the customers that are required for developing an applications. It gives all the services throught internet. There is no need to install or download the software at user end.

Example - Google App Engine.

3. Infrastructure as a service(IAAS)

IAAS provide hardware as a service and services offers as storage, operating system, network, and virtualization. In IAAS there no need to purchase any software and hardware. These services are accommodating on the users demand.

Example - Amazon EC2.

Table-1 Services of Cloud-Computing

Software as a Service (SaaS)	Platform as a Service (PaaS)	Infrastructure as a Service (IaaS)
1. Communication (Emails etc.) 2. Productivity Tools(Office)	1. Application Development 2. Database Management 3. Security Services	1. Management 2. Network 3. Storage 4. Servers
Example – Oracle, IBM, Google Apps, etc.	Example- Amazon EC2, Microsoft Azure, etc.	Example- GoGrid, Flexiscale etc.

IV. CLOUD DEPLOYMENT MODELS

1. Public Cloud-This deployment model is available to everyone and can be accessible in the cloud environment. Public services are provided on the internet based on predefined policies.

Example- Drop box, Gmail, office-365, Google, Amazon, yahoo, Microsoft are providing public services through their private cloud [6].

2. Private Cloud - Private cloud model is used by a user or single organization. Private cloud solved the security issues because the private setup is implemented safely within the firewall. The private cloud model is available only to single user or the single organization. This organization can completely take care of the cloud to satisfy their requirements.

Example: Sun cloud, IBM, WINDOW AZURE.

3. Hybrid Cloud - The combination of public and private cloud.

Example: - Google Compute Engine.

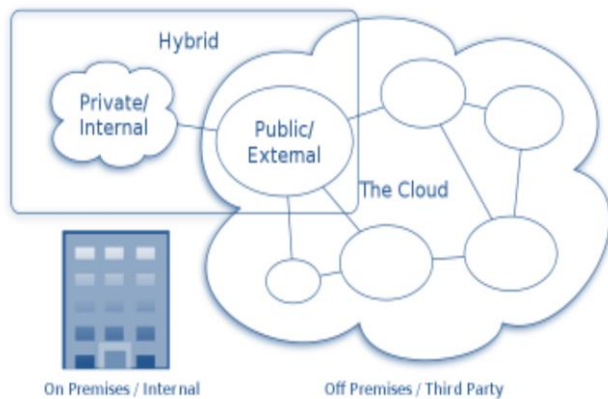


Fig.1. Types of Cloud

V. VIRTUALIZATION

Virtualization is the formation of a virtual sort rather than actual sort such as an operating system, a server, a storage device or network resources. Virtualization software helps us to run the several operating systems and multiple applications on same server or same node at the same time. In terms of cloud computing it can be referred as shared computing resources. Software and data are delivered on demand as services through the internet. Virtualization can exist without the cloud, but cloud computing cannot exist without virtualization. Cloud-computing uses virtualization to provide services to the client. On virtualization more than one system can run on the single computer so resource utilization can be increased.

5.1. There are two types of virtualization are-

- a. Full-Virtualization- In this virtualization the whole applications of one machine are available over another machine. The real-machine is functionality available in virtual machine in virtual form.
- b. Para-Virtualization- In this virtualization the whole installation are not fully available, services are delivered in a partial manner.

5.2. Benefits of virtualization

- a. Sharing of resources helps in cost reduction.
- b. Hardware independent.
- c. Virtual machines can be transfer between different hosts.

VI. LOAD BALANCING

In the cloud computing environment, the load-balancing is a method that distributes the workload among their multiple nodes to avoid overload problems. It helps in managing the resources and services. It makes better utilization of resources and help in improving the performance of the system. Load balancing is the method to find out the overloaded node (cloud servers) and help to transfer the extra load to other nodes.

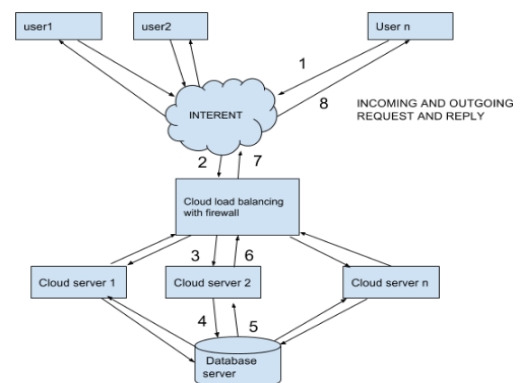


Fig.2. Load Balancing Mechanism in Cloud Computing

The following fig-2 step are as.

1. First a user sends a request for service or resource.
2. Then though internet request go to load balancing (CLB).
3. From CLB the request go to Database server through cloud server. After that Database server full field the request in good manner.
4. In this way incoming and outgoing request and reply will continues carry out with any failure.

In the algorithms of cloud load balancing, load balancing divided into two groups: Static load balancing and Dynamic load balancing algorithm.

A. Static load balancing algorithms

The load does not depend on the current state of the system but it requires knowledge about the application and resources of the system [7].

Static algorithm partition the traffic uniformly among the servers. By this approach, the traffic on the servers will be handling easy and subsequently it will make the circumstances better [3].

B. Dynamic load balancing algorithms

Dynamic algorithms are more flexible than the static algorithms and they does not reply on prior knowledge but depends on current state of the system [7]. A dynamic algorithms algorithm search throughout the whole network and selects the appropriate weights on server, and it prefers to the lightest server to balance the traffic. But selecting suitable server needs a valid communication within networks that lead to extra traffic being added on system [3].

VII. BENEFITS OF LOAD BALANCING

- i. Redundancy - It describes the process of running two or more, the same servers thus providing a guaranteed event that one server becomes occupied.
- ii. Scalability -Even though modest resources requirements are offered, scalability always is considered to find the correct host solution.
- iii. Resources Optimization -Through load balancing, one can optimize how traffic is circulated in the server cluster so, that it guarantees the best performance.
- iv. Security - In security, only one IP is exposed to the web browser with load balancing, which significantly reduces the amount of break points in case of attack [3].

VIII. DISBENEFITS OF LOAD BALANCING:

- a) More expensive.
- b) Single point of failure.
- c) Relatively lesser scalable.

IX. REVIEW OF ALGORITHM

1. Round-Robin Algorithm:

Round-Robin algorithm is a static load balancing algorithms. Round-Robin algorithm is the easy scheduling algorithms that based on the principle of time slices. In this algorithm, time is divided into multiple slices and each tasks or node is given a particular time interval. Each task's given a quantum and in this given quantum tasks has to perform its operations. Result: This method is very slow processes, because they have to share the same processors time with rest processes instead of finishing the tasks quickly.

Problem: a. if the Quantum too longs it can cause poor response time.

- b. No priority.

2. Min-Min Algorithms:

In Min-Min algorithms small tasks are executed first, which makes delay for bigger tasks for long period of time.

Result: No need for processor to wait for smaller tasks.

Problem: Starvation (for larger tasks).

3. Max-Min Algorithms:

In max-min algorithms large tasks are executed first, which makes delay for smaller tasks.

Results: No need for processor to wait for longer tasks.

Problem: Starvation (for smaller tasks).

4. Opportunistic Load Balancing (OLB) :

In this algorithms each tasks is assigned randomly to the node and try to keep busy each and every node. But it doesn't consider the current workload on every node. Its handle unexecuted tasks easily. But this algorithm does not provide better results in load balancing terms.

5. Throttled Load Balancing Algorithm (TLB):

In this algorithm the load balancer maintains an index table for virtual machines as well as their states whether available or busy. First, the client makes a request to data centre to find a suitable virtual machine (VM) to perform the job. The data-centre asked for the load balancer to allocation of the available VM. The load balancer scans the whole index table and first find out the available VM. The data-centre acknowledges the load balancer about new allocation of VM with its ID and the data centre receives the index table accordingly, if appropriate VM is not found, the load balancer returns -1 value to the data centre.

The steps of TLB phase are

1. In the first phase, the formation of the virtual machines is done.
2. In second phase they will be idle waiting for the scheduler to plan and schedule the jobs in the queue; once jobs are allocated, the virtual machines in the cloud will start processing.

- 3. In third phase the cleanup or the destruction of virtual machines is done.
- 4. The results of computing model can be estimated without considering the VM formation and destruction time, it's estimated the total number of jobs executed within a time span. This algorithm will improve the performance by providing the services and resources on demand, which increased the rate of execution of jobs and thus reducing the rejection of number of jobs submitted.

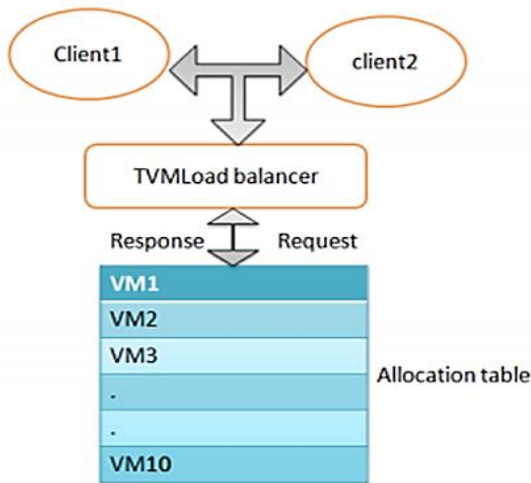


Fig.3. Throttled Algorithm

6. Ant Colony Optimization On Based Of Load Balancing Algorithm:

This algorithm is based on the techniques used by “the ants for searching their food”. The method from which ants find out the optimal path to reach their food, on the basis of this approach the workload among the nodes are distributed in well organized manner. Let’s take load balancing node in cloud computing as root head node, when ants start its movement from root node they start collecting information from the cloud node about the node is overload or not. If ants find out the node are not overload then they assigns the tasks to that node and start moving forward direction to next node. If ants find out that node is overload then its move backward direction. After reaching at target node all the information are collected and combined together to make a report about load balancing node.

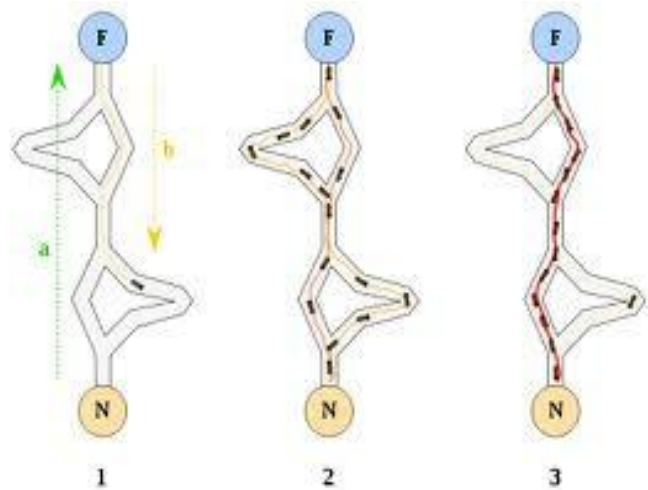


Fig.4. Ant-Colony Optimization Method

7. Weighted scheduling:

The weighted load balancing method is where server connections deals with performance of services and the weighted operation of service are providers and the results are shown as a load balancing priority weight [8].

$$\phi_i = \frac{\mu_i}{\lambda} \left(1 - \sqrt{\frac{\lambda}{\mu_i \beta}} \right)$$

$$\beta = \frac{1}{\lambda^2} \frac{[\sum_{i=1}^N \sqrt{\mu_i}]^2}{[\sum_{i=1}^N \mu_i]^2 - \frac{2}{\lambda} [\sum_{i=1}^N \mu_i] + 1}$$

From paper [8] Equation (1) and (2) are taken.

Table 2 WRR Dynamic algorithm symbols

Symbol	value
ϕ_i	Weight, which represents the shared load percentage
λ	ARRIVAL TIME
μ_i	FINISHED TIME
β	Assume value

Let’s see numerical problems on WRR Dynamic algorithm

Process	Arrived time (λ)	Finished time (μ)	Results (weight)
p1	2	3	1
p2	5	7	1
p3	4	8	1

Solution:

A. $\lambda = 2, \mu = 3$

$\phi 1 = 3/2 [1 - \sqrt{2}/3]$ here, $\beta = 6$

$= 3/2 [1 - \sqrt{1/9}]$

$= 3/2 [2/3]$

$= 1$

$\beta = 1/2 (\sqrt{3}) 2/ 1/2 [3] 2 - 2/2[3] + 1$

$\beta = 6$

B. $\lambda = 5, \mu = 7$

$\phi 2 = 7/5 [1 - \sqrt{5}/7]$ here, $\beta = 35/4$

$= 7/5 [12/7]$

$= 7/5 [5/7]$

$= 1$

$\beta = 1/5 (\sqrt{7}) 2/ 1/ (5) 2 [7] 2 2/5 [7] + 1$

$= 1/5 * 7 / 1/25 * 49 2/5 * 7 + 1$

$= 7/5 * 125/20$

$\beta = 35/4$

Results - Response time equal for every process.

X. GENETIC ALGORITHM

In GA, there are four steps are done to perform the genetic algorithm and there are selections, crossover, mutations and termination. Here's in this genetic algorithm we selects the virtual machine from random basis. We crossover the processors and VM's chromosome with each other and find out the fitness function as a result. Offspring are created by exchanges between processors and VM at crossover method. After that we allocate the offspring result to the processors. By this process we got the higher priority processors with jobs to execute it.

Let's takes numerical Problem on GA [13].

$F(x) = x^2$

SOLUTION:

STEP 1: Calculate the fitness function values using genetic algorithm method.

String no	Initial population	X value	Fitness f(x)=x ²	Prob p _i	p%	Expected court	Actual court
1	01100	12	144	0.1247	12.47	0.4987	1
2	11001	25	625	0.5411	54.11	2.1645	2
3	11011	5	25	0.0216	2.16	0.0866	0
4	10011	19	361	0.3126	31.26	1.2502	1
sum			1155	1.0000	100	4.0000	4
avg			288.75	0.2500	25	1.0000	1
max			625	0.5411	54.11	2.1645	2

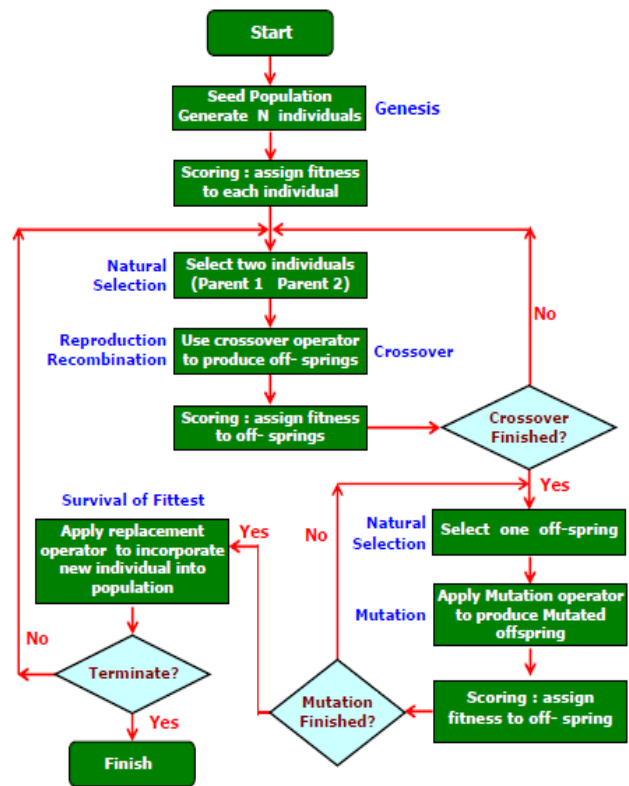


Fig.4. Genetic algorithm flow chart

Step 2: Crossover table

String no	Mating pool	Crossover	Offspring after crossover	X -values	Fitness value
1	01100	4	01101	13	169
2	11001	4	11000	24	576
3	11001	2	11011	27	729
4	10011	2	10001	17	289
sum					1763
avg					440.63
max					729

Step 3: Mutation table

String no	Offspring after crossover	Mutation chromosome for flipping	Offspring after mutation	X values	Fitness function
1	01101	10000	11101	29	841
2	11000	00000	11000	24	576
3	11011	00000	11011	27	729
4	10001	00100	10100	20	400
sum					2546
avg					636.5
max					841

Result: Thus, the string no (1) will great chance of selection because it's have highest fitness function.

XI. LITERATURE SURVEY

In [11] Proposed a load balancing mechanism based on improved GA using PR (Population Reduction Method) gives satisfactory results. It's first identified that resource that has good population rather than others and try to gives priority to complete the jobs. It's handled the overload resource and transfer that load to ideal. After selecting resources by PR Method overload resource and transfer that load to ideal are handle. In paper they used cloud-sim simulator for checking the allocation and load on resources. When are entered into cloud simulator then first population is define on population PRM method and after completing all jobs the response time for resource will given an idea for finishing job time.

In [4] Proposed a load balancing mechanism based on comparing different load balancing together and find out that the ant colony optimization based on the action of ants and seeking of an optimal path during collecting their food .Workload are distribute efficient among the nodes and optimal job scheduling is achieved.

In [12] Proposed a load balancing strategies Dynamic Feedback technique provide better results. In this strategy uses a monitor daemon to check various parameters such as availability, load and response time to balance load. In dynamic load balancing algorithm consider the server load information, server capability, response time, active connection etc in every field dynamic feedback technique work in better and efficient manner.

Conclusion

Cloud computing provides everything to user as a service over network. Load balancing is very important issues in cloud computing terms. Overloading of a system can cause poor performance, delay in response or delay in execution of tasks can make cloud computing fail or unsuccessful. To make proper and efficient utilization of resources and services, the load balancing is required.

References

- [1]. B. Patel, S. Patel, "Various Load Balancing Algorithms in cloud computing", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.1, Issue.2 pp.23-29, 2015.
- [2]. R. Nallakumar, N. Sengottaiyan, S. Nithya, "A Survey of Task Scheduling Methods in Cloud Computing", International Journal of Computer Sciences and Engineering, Vol.2, Issue.10, pp.9-13, 2014.
- [3]. Z. Chaczko, V. Mahadevan, S. Aslanzadeh, C. Mcdermid, "Availability and load balancing in cloud computing", International Conference on Computer and Software Modeling, Singapore pp.138-140, 2011.
- [4]. AS. Togadiya , PC. Togadiya, "A Comparative Analysis of Load Balancing Algorithm in Cloud Computing Environment", International Journal of Advanced Research in Computer Engineering & Technology, Vol.1, Issue.3, pp.120-124, 2012.
- [5]. D. Saranya, LS. Maheswar, "Load Balancing Algorithms in Cloud Computing: A Review", International Journal of Advanced Research in Computer Science and software Engineering, Vol.5, Issue.7, pp.1107-1111, 2015.
- [6]. S. Sakshi, NS. Ghumman, "Cloud computing model and its load balancing algortihms," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp.2940-2943, 2016.
- [7]. Reena Panwar, Bhawna Mallick, "Comparative Study of Load Balancing Algorithms in Cloud Computing" International Journal of Computer Applications, Vol.117, No.24, pp.1-17, 2015.
- [8]. ShangLiang Chen, YunYao Chen, SuangHong Kuo, "CLB: A novel load balancing architecture and algorithm for cloud services", Computers and Electrical Engineering, Issue.56, Issue.2, pp.154-160, 2016.
- [9]. R. Gupta, R. Bhatia, "An Enhanced and Secure Approach of Load Balancing in Cloud Computing", International Journal of Computer Sciences and Engineering, Vol.2, Issue.8, pp.112-116, 2014.
- [10]. Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing", International Journal Of Scientific & Technology Research, Vol.3, Issue.11, pp.115-119, 2014.
- [11]. RR Patel, SJ Patel, "Improved GA using population reduction for load balancing in cloud computing", 2016 International Conference on Advances in Computing Communications and Informatics (ICACCI), Jaipur, pp. 2372-2374, 2016.
- [12]. M. Rahman, S. Iqbal and J. Gao, "Load Balancer as a Service in Cloud Computing", 2014 IEEE 8th International Symposium on Service Oriented System Engineering, Oxford, pp. 204-211, 2014.
- [13]. P. Kaur, S. Majithia, "Implementation and Analysis of Replicated Agent Based Load Balancing In Cloud Computing", International Journal of Computer Sciences and Engineering, Vol.2, Issue.8, pp.21-27, 2014.