Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/15135
DOI URL: http://dx.doi.org/10.21474/IJAR01/15135

RESEARCH ARTICLE

A REVIEW ON CLOUD LOAD BALANCING TECHNIQUES

Sayantan Das, Madhurima Mondal and Prof. Soumen Swarnakar
Department of Information Technology, Netaji Subhash Engineering College, Kolkata, India.

Manuscript Info

Manuscript History
Received: 31 May 2022
Final Accepted: 30 June 2022
Published: July 2022

Key words:-
Cloud Computing (CC), Load Balancing (LB), Load Balancing Algorithms

Abstract

Cloud computing simply means storing & accessing data and programs over the internet instead of your computer's hard drive. The cloud is just a metaphor for the internet. An example of Cloud Computing's process could also be Google's Gmail. Cloud computing is highly efficient, productive, safe & secure.

Cloud computing consists of two terms One is Cloud which is a useful feature of computer technology to perform appropriate tasks. Computing may include computer hardware and software but must involve some form of a computer system. It is a type of operation of a computer that is fully dependent on exchanging computing resources rather than having local servers or personal devices to handle the appeal. Cloud refers to a public or semi-public space on transmission lines that exists between the endpoints of a transmission. Data that is transmitted across a WAN enters the network from one endpoint using a standard protocol.

Introduction:-

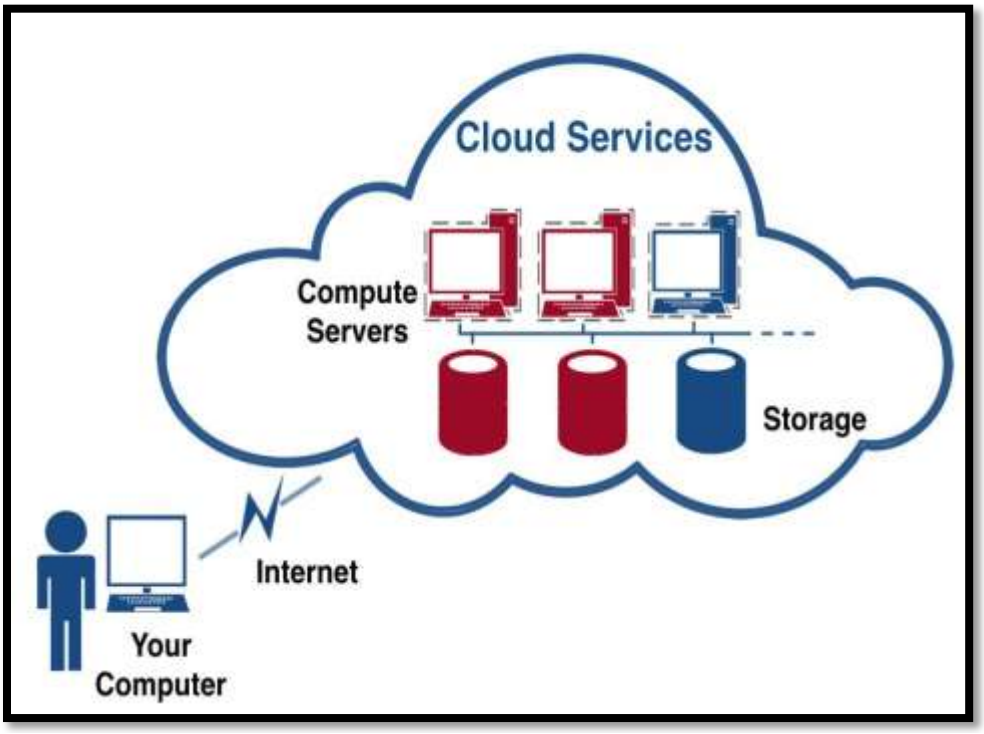
The definition of Cloud Computing provided by Vangie Beal says: "Cloud computing is defined as a type of computing that depends on sharing computing resources rather than having local servers or personal devices to handle applications. Cloud computing is the next stage in the evolution of the internet, it provides the means through which everything from computing power to computing infrastructure, applications, and business processes can be delivered to you as a service wherever and whenever you need them. Cloud computing solutions can simplify how business operates, particularly in terms of hardware needs. Through a cloud solution, users can now connect and access the same information, but now the advantage is any user can connect from anywhere and enjoy a more streamlined technology installation, as compared to before. The use of the word "cloud" refers to the two essential concepts:

• Abstraction: Cloud computing abstracts system implementation details from users and developers. Applications run on physical systems that aren't specified, data is stored in unknown locations, administration of systems is outsourced to others, and access by users is ubiquitous.

• Virtualization: Cloud computing virtualizes systems by pooling and sharing resources. Systems and storage can be provisioned as needed from a centralized infrastructure, costs are assessed on a metered basis, multi-tenancy is enabled, and resources are scalable with agility.

Corresponding Author:- Sayantan Das
Address:- Department of Information Technology, Netaji Subhash Engineering College, Kolkata, India.

So in other words, Cloud computing means that instead of using all the computer hardware and software on our desktop, or anywhere inside our company's network, it's provided for us as a service by another company and accessed by us over the Internet. The physical location of the hardware and software and how it works doesn't matter to us.



Important Components of Cloud Computing Architecture:-

Here are some important components of Cloud computing architecture:

a. Client Infrastructure:

Client Infrastructure is a front-end component that provides a GUI. It helps users to interact with the Cloud.

b. Application:

The application can be any software or platform which a client wants to access.

c. Service:

The service component manages which type of service you can access according to the client's requirements.

Three Cloud computing services are:

1. Software as a Service (SaaS)
2. Platform as a Service (PaaS)
3. Infrastructure as a Service (IaaS)

d. Runtime Cloud:

Runtime cloud offers the execution and runtime environment to the virtual machines.

e. Storage:

Storage is another important Cloud computing architecture component. It provides a large amount of storage capacity in the Cloud to store and manage data.

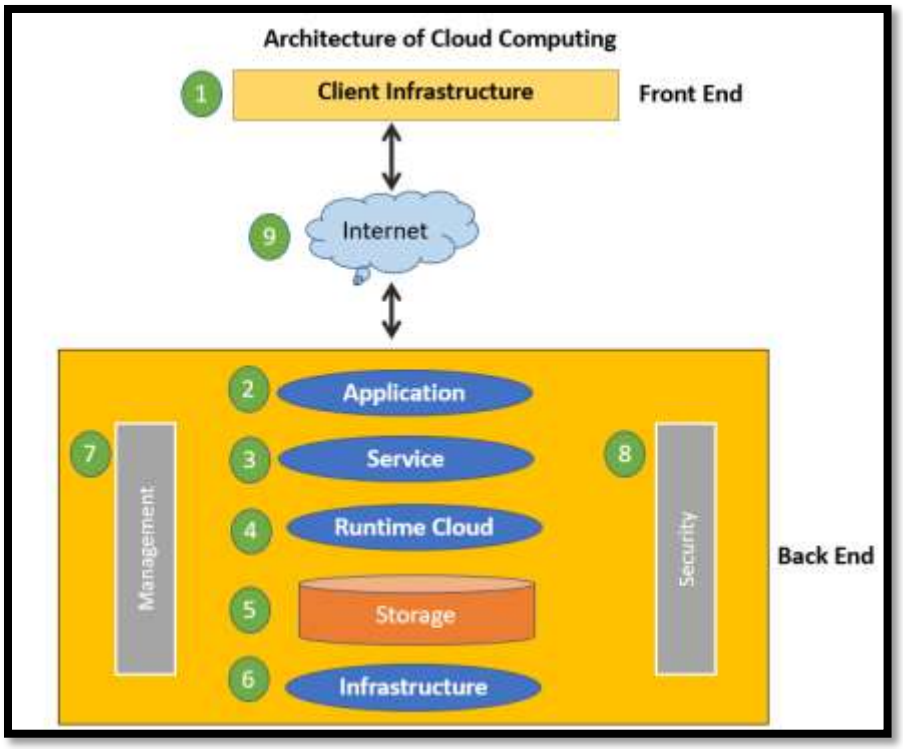
f. Infrastructure:

It offers services on the host level, network level, and application level. Cloud infrastructure includes hardware and software components like servers, storage, network devices, virtualization software, and various other storage resources that are needed to support the cloud computing model.

g. Management:
This component manages components like application, service, runtime cloud, storage, infrastructure, and other security matters in the backend. It also establishes coordination between them.

h. Security:
Security in the backend refers to implementing different security mechanisms for secure Cloud systems, resources, files, and infrastructure to the end-user.

i. Internet:
Internet connection acts as the bridge or medium between frontend and backend. It allows you to establish the interaction and communication between the frontend and backend.



Importance of Cloud Load Balancing:-

Cloud load balancing is specified as the method of splitting workloads and also computer residential properties in a cloud computer. It makes it possible for venture to handle work needs or application needs by distributing resources among many computer systems, networks or servers. Cloud load balancing consists of holding the flow of work website traffic and also demands that exist online.

As the website traffic on the web growing rapidly, which is about 100% yearly of the present traffic. Hence, the workload on the server expanding so quick which causes the overloading of servers primarily for preferred web server. There are two primary services to get over the issue of overloading on the web servers-.

First is a single-server solution in which the server is updated to a greater performance server. However, the brand-new web server may also be strained quickly, demanding an additional upgrade. Additionally, the upgrading process is arduous and also expensive.

Second is a multiple-server solution in which a scalable solution system on a collection of web servers is constructed. That's why it is a lot more cost effective in addition to more scalable to build a server cluster system for network services.

Benefits of cloud load balancing:-

Load balancing provides the same set of benefits, regardless of whether it lives in a local data center or a cloud environment:

a. Better workload scalability and performance:

A single workload or application is fine as long as it handles the incoming traffic and requests in a timely manner, but sometimes, a business must add workload instances to handle greater network traffic volumes and sudden and unexpected spikes in traffic. A load balancer is critical here to queue and distribute that traffic across multiple instances so the overall application runs efficiently and with satisfactory UX.

b. Better workload reliability:

When a single workload handles 100% of incoming requests, the underlying software and underlying hardware pose a single point of failure for the workload. Adding more workload instances and load balancing the traffic between them vitally enhance workload resiliency and availability. If one workload instance (node) fails, others continue to function and direct traffic to remaining instances. This is the heart of high availability workload deployments.

c. Better business continuity (BC) and governance:

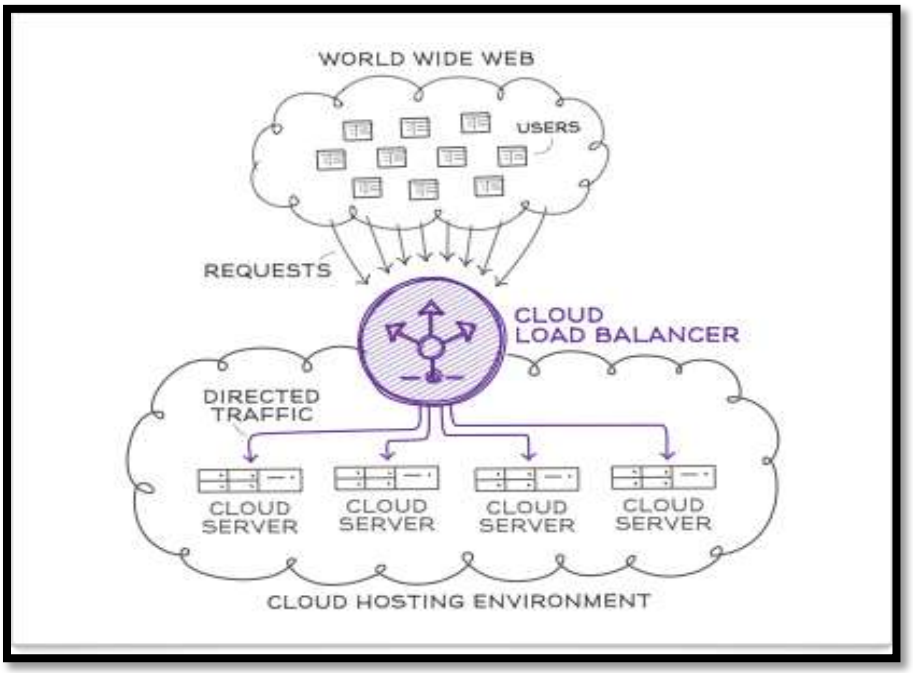
For many businesses, application availability is central to business governance and regulatory compliance. The implementation of multiple workloads into a cluster and sharing traffic with load balancers is a primary means to boost workload reliability, which, in turn

Working of Cloud Load Balancer:-

Here, load refers to not only the website traffic but also includes CPU load, network load and memory capacity of each server. A load balancing technique makes sure that each system in the network has same amount of work at any instant of time. This means neither any of them is excessively over-loaded, nor under-utilized.

The load balancer distributes data depending upon how busy each server or node is. In the absence of a load balancer, the client must wait while his process gets processed, which might be too tiring and demotivating for him.

In short, a load balancer acts as the “traffic cop” sitting in front of the servers and helps in routing client requests across all servers capable of fulfilling those requests in a manner that maximizes speed and capacity utilization. It ensures that no one server is overworked, which could degrade performance. When a new server is added to the server group, the load balancer automatically starts to send requests to it.



Various information like jobs waiting in queue, CPU processing rate, job arrival rate etc. are exchanged between the processors during the load balancing process. Failure in the right application of load balancers can lead to serious consequences, data getting lost being one of them.
Different companies may use different load balancers and multiple load balancing algorithms like static and dynamic load balancing. One of the most commonly used methods is Round-robin load balancing**.**

It forwards client request to each connected server in turn. On reaching the end, the load balancer loops back and repeats the list again. The major benefit is its ease of implementation.

Cloud load balancing algorithms:-

Static load balancing algorithms

Static load balancing is generally used for homogeneous environment when the load is low. Generally the load or traffic is divided evenly among the servers. Static Algorithms are made for those systems which have very low variations in load. In Static Algorithm, the entire traffic is equally divided amongst the servers. This algorithm needs an in-depth knowledge of server resources for better performance of the processors which is determined at the beginning of the implementation.

However, the decision of load shifting does not depend on the present state of the system. There is one major drawback of Static Load Balancing Algorithm that is load balancing tasks work only after they are created, it could not be implemented to other devices for load balancing.

Dynamic load balancing algorithms

Dynamic load balancing algorithms are those algorithms which search for the lightest server in the network and then designated appropriate load on it. In this, work load is distributed among the processors at runtime. The algorithms in this category are considered complex, but have better fault tolerance and overall performance.

Existing load balancing Algorithms

In this section some existing different kinds of load balancing algorithms and the difference between them. The paper also provides details about various parameters, used to compare the existing techniques.

Round Robin Load balancing algorithm

Round-robin load balancing is one of the simplest and most used load balancing algorithms. Client requests are distributed to application servers in rotation. For example, if you have three application servers: the first client request to the first application server in the list, the second client request to the second application server, the third client request to the third application server, the fourth to the first application server and so on.

This load balancing algorithm does not take into consideration the characteristics of the application servers i.e. it assumes that all application servers are the same with the same availability, computing and load handling characteristics.

Algorithm

```
Round_Robin_Load_Balancing ()
{
   Initialize all the VM allocation status to AVAILABLE in the VM state list;
   Initialize hash map with no entries;
While(new request are recived by the Data Centre Controller)
  Do {
          Data Center Controller queue the requests;
          Data Centre Controller removes a request from the beginning of thequeue;
If(hash map contain any entry of a VM corresponding to the current requesting user base
          VM allocation status = AVAILABLE)
          The VM is reallocated to the user base request;
       }
   Else
       {
          Allocate a VM to the user base request using Round Robin
          Algorithm;
          Update the entry of the user base and the VM in the hash map and the
          VM state list;
       }
   }
}
```

Opportunistic Load Balancing Algorithm

Opportunistic load balancing algorithm attempts to keep each node busy. Therefore it does not consider the present workload of each computer. OLB dispatches unexecuted tasks to currently available nodes in a random order regardless of the node's current workload. Since OLB does not calculate the execution time of the node, the task to be processed will be processed in a slower manner resulting in bottlenecks despite some of the nodes being free.

Min-Min Load Balancing Algorithm

Min-min load balancing algorithm begins by finding the minimum completion time for all tasks. Then among these minimum times, the minimum value is selected which is the minimum time amongst all tasks on any resource. According to that minimum time, the task is then scheduled on the corresponding machine. The execution time for all other tasks is updated on that machine and that task is removed from the list. This procedure is followed until all the tasks are assigned the resource. In scenarios where the number of small tasks is more than the number of large tasks, this algorithm achieves better performance. However, this approach can lead to starvation. The algorithm is as follows.

Algorithm

```
Procedure Minmin(Task Ti)
{
   Find execution_Completetion_time of each task
   Store the execution_Completetion_Time of task Ti inorderQueue
  Repeat
   {
for each task ti in orderQueue
   }
  Obtain minimum completetiontime from orderQueue;
  Assign task to vm;
  Update the execution_Completetion_Time;
   }
  Until orderQueue empty;
}
```

Max-Min Load Balancing Algorithm

Max-min load balancing algorithm is similar to the min-min algorithm except the following: after finding out the minimum execution times, the maximum value is selected which is the maximum time amongst all tasks on the resources. Then according to the maximum time, the task is scheduled on the corresponding machine. The execution time for all other tasks is updated on that machine and the assigned task is removed from the list of tasks that are to be assigned to the machines. Since the requirements are known beforehand, this algorithm is expected to perform well.

Throttled Algorithm

Throttled load balancing algorithm depends on state of virtual machine. Here two virtual machine states are there based on which the algorithm is performed that are AVAILABLE/BUSY .This state is based on the allocation of virtual machine to the request, is allocated or not. A hash table maintained at loadbalancer will contain this state information. ID and status(AVAILABLE/BUSY) of the virtual machine are the two parameters of a hash table. At the beginning every virtual machine is on available state.

When the data center receives client request it simply forward them to load balancer to find suitable virtual machine for request. Among all virtual machine with AVAILABLE state the virtual machine (VM) which is to be assigned to a request is determined by load balancer.

For this, load balancer will start look through the hash table from top-to-bottom. After identifying an available virtual machine from the hash table, datacentre controller will be notified with the id of that virtual machine for request allocation. If data center successfully allocate the virtual machine id to the specified request, it will notify the success of its operation to load balancer and accordingly, then hash table will be updated by load balancer.During allocation.

If datacenter controller detecting any problem then a negative feedback will send to load balancer which result no updation on hash table. If all virtual machine's are in BUSY state datacenter will receives a notification, based on which datacenter will start queuing user requests at its own pool. The virtual machine itself notify job done to the data center controller after the completion of allocated job to that virtual machine and it will inform the same to the load balancer,and load balancer will make appropriate.

Honey Bee Behaviour Load Balancing Algorithm

This is nature inspired algorithm for self-organization. HB consists of a queen and foragers, where forager is of two types; employed and unemployed. The foragers are informed about the food available nearby by waggle dance by scout bee (unemployed), the dance is to give the information to the other foraging bees about the distance, quality, direction and other information which is useful in getting the food12. This algorithm has similar principle in balance the work of the virtual machine. The HBB algorithm calculates the virtual machine workload, then it decides whether it is overloaded, light weighted or balanced. The high priority of the task is off from the overload virtual machine and tasks are waiting for the lightweight virtual machine. These tasks are known as scout bee in the next step. Honey Bee Behaviour inspired Load Balancing technique reduces the response time of VM and also reduces the waiting time of task.

Ant Colony based Algorithm

The ant colony algorithm is based on the behaviour of the real ants. The ant can notice the optimal path where the source food is available. The Ants while seeking a path from their colony in search of food secrete a chemical called pheromone on the ground thus leaving a trail for other ants to follow the path. But this chemical evaporates with time.

Comparative Analysis for all Algorithms:-

Algorithms	Throughput	Overhead	Fault Tolerance	Migration Time	Response Time	Resource Utilization	Scalability	Performance
Round Robin	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Min-Min	Yes	Yes	No	No	Yes	Yes	NO	Yes
Max-Min	Yes	Yes	No	No	Yes	Yes	NO	Yes
OLB	No	No	No	No	No	Yes	No	Yes
Throttled	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Honeybee	No	No	No	No	No	Yes	No	No
Ant Colony	No	No	Yes	No	No	Yes	Yes	No

Conclusion:-

This article has described about different types of existing cloud load balancing algorithms and their merits and demerits. Cloud computing is an art of computer technology which distributesincoming tasks among different virtual machines or cloud servers. Load Balancing is one of the solutions with Cloud Computing to overcome the problem of overloading serverswhich lead to dreadful results in computation time. So effective Load balancing algorithm is required for effective use of resources. The main goal of Load balancing is to meet user needs by distributing the workload across multiple network nodes & maximizing resource usage & growing device efficiency. This research paper discusses the merits demerits of different existing conventional load balancing algorithms associated with cloud computing. Different research works are carried on by different researchers for overcoming the problem of load balancing in cloud computing. As a conclusion it can be said that an effective load balancing is very important for reliability, system efficiency, throughput optimization, resource usage, makespan time, response time minimization.

Reference:-

1. S. Swarnakar, Z. Raza, S. Bhattacharya and C. Banerjee, "A Novel Improved Hybrid Model for Load Balancing in Cloud Environment," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)**,** 2018, pp. 18-22, doi: 10.1109/ICRCICN.2018.8718697.
2. S. Swarnakar, N. Kumar, A. Kumar and C. Banerjee, "Modified Genetic Based Algorithm for Load Balancing in Cloud Computing," 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE)**,** 2020, pp. 255-259, doi: 10.1109/ICCE50343.2020.9290563.

3. S. Swarnakar, R. Kumar, S. Krishn and C. Banerjee, "Improved Dynamic Load Balancing Approach in Cloud Computing," 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), 2020, pp. 195-199, doi: 10.1109/ICCE50343.2020.9290602.
4. Dam, S., Mondal, G., Dasgupta, K., Dutta, P.: An ant colony based load balancing strategy in cloud computing. Adv. Comput. Netw. Inf. 2, 403–413 (2014)
5. GeethuGopinath P P(1), Shriram K Vasudevan(2), "An in-depth analysis and study of Load balancing techniques in the cloud computing environment", published in 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).
6. James, J., Bharma, B.: Efficient VM load balancing algorithm for a cloud computing environment. Int. J. Comput. Sci. Eng. (IJCSE) 4(9), 1658–1663 (2012)
7. N. S. Raghava and D. Singh, "Comparative Study on Load Balancing Techniques in Cloud Computing," vol. 1, no. 1, pp. 18-25, 2014.
8. Suriya Begum, Dr.PrashanthC.S.R,"Review of Load Balancing in Cloud Computing", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, 2013.
9. f. Harvindersingh, Rakesh Chandra Gangwar, "Comparative Study of Load Balancing Algorithms in Cloud Environment", International Journal on Recent and Innovation Trends in Computing and Communication ,Vol 2, Issue 10, Oct 2014.
10. H. Rai, S. K. Ojha and A. Nazarov, "Cloud Load Balancing Algorithm," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 861-865, doi: 10.1109/ICACCCN51052.2020.9362810.