IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105146

Load Balancing in Cloud Computing

Yelchuri Venkata Sai Harsha[1], Nagaraj G Cholli[2]

Student, Information Science, R.V College of Engineering, Bangalore, India[1]

Professor, Information Science, R.V College of Engineering, Bangalore, India[2]

Abstract: Cloud computing is a properly-defined paradigm for computing offerings wherein information and assets are amassed from cloud carrier companies through the internet by the use of well-designed internet-primarily based equipment and packages.. Cloud Computing is simply a combination of computer resources and services and is supplied to people at a cost-effective rate. The sharing of resources might lead to an issue with the availability of such resources, which causes a stalemate scenario. The technique of dispersing network traffic over many servers is known as load balancing. This guarantees that no single server is overburdened. Load balancing increases application responsiveness by distributing tasks equally. It also makes programmes and websites more accessible to users. The purpose of this paper is to understand load balancing.

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

I. INTRODUCTION

Cloud computing is the distribution of diverse offerings along with storage, servers, networking, software programs, intelligence, and analytics, through the internet so as to offer faster innovation, more flexible sources, and economies of scale.

Take an illustration of a site open to everyone. A high number of clients can visit a site or online application whenever. A web application's capacity to deal with these client demands without a moment's delay gets intense. It might even cause system failures. The terrible sense of a website being down or not accessible also delivers lost prospective clients for a website owner whose entire career is based on his portal. Load balancing is crucial in this situation.

Cloud systems have two load balancing tiers.
Fundamental level: In the endeavour to balance the computational load of a number of applications between physical computers, the load balancer distributes the appropriate instances to physical machines when an application is uploaded.
Second level: If many incoming requests are received by an application, each such request should be allocated to the individual instance of the programme to balance the computer load over a group of instances.

II. LOAD BALANCING

Cloud load balancing is the way workloads and computing resources are distributed over one or more servers. This type of distribution guarantees greatest performance in minimal reaction time. Two or more servers, hard discs, network interfaces or other computer resources are separated to allow better use of resources and time spent responding to system operations. Effective cloud load balancing may therefore assure continuation in operation with a heavy traffic website.

The term "load" encompasses not just website traffic but also CPU load, network load, and server storage capacity. Load balancing ensures that every machine in the network gets a similar measure of burden at some random time. This means that none of them are overburdened or underutilised in any way.

Depending on how busy each server or node is, the load balancer distributes data. In the absence of a load balancer, the client will have to wait for his procedure to be completed, which may be too exhausting and demotivating for him.

III. LOAD BALANCING OBJECTIVES

The following are the main objectives of load balancing:

1. Management and control of traffic surges on one server
2. Reducing response time for user requests

DOI 10.17148/IJARCCE.2021.105146

3. To increase the ratio of resource usage.
4. A significant performance improvement.
5. Maintaining the system's stability.
6. Increase the system's flexibility to respond to the changes..
7. The workflow time and waiting time in a line are minimised.
8. To promote pleasure of the user.

IV. ADVANTAGES OF LOAD BALANCING

High-Performance Applications: Unlike their conventional local counterparts, cloud load balancing approaches are less costly and easier to adopt. Companies are able to make their customer apps operate quicker and perform better than at possibly cheaper prices.

Enhanced scalability: Cloud balancing helps keep the website traffic scalable and agile. You can simply match increasing user traffic with effective load balancers and distribute it across several servers or network devices. For online sites that are dealing with thousands of internet visitors per second, it is very vital. Such effective load balancers are needed to spread workloads during sales or other promotional services.

Capacity to deal with traffic surges: During any result announcement, an usually functioning University site might entirely go down. This is due to the possibility of receiving requests at a high scale. They won't have to worry about traffic surges if they use cloud load balancers. Whatever the size of the request, it may be carefully spread among several servers to yield the best results in the shortest amount of time.

Continuity with full flexibility: The main goal is to save or defend a site from abrupt failures by deploying a load balancer. Even if one node fails, the workload may be moved to another active node when distributed among several servers or network units.

V. NEED OF LOAD BALANCING

Load balancing is also a critical element in the scalability of a cloud. Cloud infrastructures should readily be expanded to suit traffic ups and ups. If a cloud "scales up," it usually spins several virtual servers and operates a number of applications. The principal network component that distributes traffic across these new instances is the load balancer.

Without load balancers, virtual servers spun out fresh may not accept incoming traffic in a coordinated way or if at all. Some servers are left without traffic taking care of while others are overwhelmed. Load balancers may also identify unavailable servers and route traffic to those that are still in service. Depending on the algorithms of load balancing, load balancers can even assess if a specific server (or server set) is likely to be overwhelmed faster and to route traffic to other nodes thought to be more healthy. Such proactive skills can greatly lower the likelihood that your cloud services will not be accessible.

Load balancing is additionally important to accomplish green cloud computing. The explanations behind this are:

1. Limited power usage: Load balancing can reduce power usage by limiting excessive workload over core nodes or virtual machinery.
2. Carbon Emission Reduction: Carbon emissions and energy consumption are like heads and tails of a coin. They are both directly proportional. Load balance helps cut energy use, which reduces carbon emissions automatically and so results in Green Computing.

VI. CLASSIFICATION OF LOAD BALANCING ALGORITHMS

They are categorised according to the present status of the system into two types:

1. Static Load Balancing: When it comes to job distribution, a load balancing method is "static" if it ignores the state of the system. The system status comprises actions such as the degree of load of various processors (and occasionally even overflow). Instead, assumptions are established in advance on the whole system, such as arrival periods and incoming resource requirements. The numbers, power and communication speeds of CPUs are also known. Static load balancing is therefore intended to combine a given set of workloads with available processors so that a given performance function is minimised.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105146

Static load balancing systems are usually focused on a router, or Master that distributes and optimises loads. This reduction can take into consideration information on the jobs to be dispersed and a predicted runtime.
The benefit of static algorithms is that in fairly frequent activities they are easily installed and incredibly efficient.

2. Dynamic Load Balancing: Dynamic algorithms take into consideration the current load of each computer unit (also known as nodes) on the system, as opposed to static load distribution techniques. Tasks can therefore dynamically be transferred from an overloaded node to an underloaded node to be processed more quickly. Although these algorithms are significantly harder to build, they can yield superb results, especially if the execution times vary substantially between tasks.

Because a separate node devoted to task distribution is not required, dynamic load balancing design may be more flexible. It is a unique assignment when tasks are allocated to a processor based on its status at a given time. Dynamic assignment, on the other hand, refers to the ability to redistribute duties on a continuous basis based on the condition of the system and its progress. Obviously, a load balancing algorithm that requires excessive communication to obtain its conclusions runs the danger of delaying the overall problem's resolution.

VII. LOAD BALANCING ALGORITHMS

TABLE I LOAD BALANCING ALGORITHMS

Algorithm	State of Algorithm	Job Distribution	Advantages	Disadvantages
Round Robin and Randomized	Static	1. All processors share the same amount of work. 2. Each processor keeps track of the sequence in which processes are assigned. 3. This method is used to handle user requests in a circular manner.	1. It is good to use it when no of processors are significantly less when compared to no of processes. 2. Inter-process communication is not required in Round Robin.	Because various processes take various amounts of time to complete, some nodes may be fully occupied while others are idle and underutilised at any same moment.
Central Manager	Static	1. The central processor has to pick the host for all new processes. 2. The loading processor minimum relies on the total load determined during the process setup.	The load scheduler decides on load balancing based on the system load statistics	A high level of interprocess communication is required which may be expensive..
Min-min	Static	1. For all jobs, the shortest possible completion time is sought. 2. The minimum value is determined from the minimum times.	Good performance with the best number of resources.	Starvation scenario is possible here.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105146

		3. The work is allocated based on that minimal time.		
Max-Min	Static	This method is quite similar to the above method. However there is one major difference: After obtaining the shortest execution times, the one which is maximum is picked.	Good performance with the best number of resources.	Starvation scenario is possible here.
Honey Bee Foraging Behavior	Dyna mic	Based on the behaviour and approach of the honeybees to harvest honey. The global load balancing is achieved by local server activities.	The virtual machine has reduced response times and waiting time.	Throughput is indirectly proportional to resource number.
Biased Random sampling	Dyna mic	Each server is considered a node's vertex, and the indegree symbolises the nodes' available free resources. The work is assigned based on the in degree. If each node has more than one degree, work will be assigned to that node. When a work is given to a node, the degree of the node is decreased by one, and it is increased after the work is completed.	Performs better with a large and similar resource population.	Degrades with increasing population variety.
Active Clustering	Dyna mic	This algorithm's fundamental premise is to group similar nodes together and operate with those grouped nodes. The resources can enhance throughput more efficiently by grouping nodes together.	1. When there are a lot of resources being used, performance is good. 2. Using additional system resources to boost performance.	Degrades with increasing population variety.
ACCLB(Ant Colony and Complex	Dyna mic	Small-world and size-free features of a complex network make it possible to	This methodology eliminates heterogeneity, is	1. Used in networks wich are complex only.

network Theory)		balance loads effectively.	dynamic, is great in tolerance of failure and so helps improve system performance.	2.performance and Scalability of this network are poor.

VIII. Challenges of Load Balancing

Migration of virtual machines: The notion consists on creating a machine as a file or a file set. The burden on a loaded computer can be reduced by effective movement of the virtual machine. The aim is to eliminate and minimise load on cloud computers when load is spread dynamically on the machine.

Management of energy: Advantages of using cloud include the scale economies. Energy conservation is a crucial issue for a global economy. Since diminished providers support a scope of worldwide assets, and every one has its own assets. How can data center portion be employed while maintaining its throughput acceptable?

Data storage and management: The storage of information is another crucial necessity. So how can data be disseminated with the most suitable storage and rapid access in a cloud system?

Cloud node spatial distribution: Some methods are only offered for nodes that are close together and have minor communication latency. However, designing an effective load balancing method that can be effectively articulated for spatially scattered nodes remains a challenge.

LB Scalability: Accessible and on-demand scalable cloud services allow guests to access resources for quick scaling at any time. A solid load balancer should take into account quickly changing requirements in computing circumstances, memory, device architecture, and so on.

IX. Conclusion

Simply explained, Cloud computing is a method for several users to access diverse resources through the internet on an as-needed basis. However, there are significant hurdles with respect to cloud computing.

Load balancing is a key hurdle in cloud computing. Many static and dynamic algorithms are discussed in this paper. As it is a known fact that cloud is heterogeneous in nature. Static algorithms provide easy modelling and environment monitoring, but do not simulate diverse cloud nature. Dynamic load balancing algorithms are hard to model, but are well adapted for the diverse nature of cloud environments.

This paper gives an outline of load balancing, its advantages, need and obstacles and discusses several existing methods for load balancing.

References

[1]. Ms. Shalini Joshi , Dr. Uma Kumari “Load Balancing in Cloud Computing:Challenges & Issues” , Conference: 2016 2nd International Conference on Contemporary Computing and Informatics, DOI:10.1109/IC3I.2016.7917945.

[2]. Muhammad Asim Shahid, Noman Islam, Muhammad Mansoor Alam, Mazliham Mohd Su’ud, Shahrulniza Musa,”A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach ” . DOI: 10.1109/ACCESS.2020.3009184

[3]. Foram F Kherani, Prof.Jignesh Vania, “Load Balancing in cloud computing ”, 2014 IJEDR | Volume 2, Issue 1 | ISSN: 2321-9939

[4]. Shahbaz Afzal and G. Kavitha, “Load balancing in cloud computing – A hierarchical taxonomical classification”, Journal of Cloud Computing volume 8, Article number: 22 (2019). DOI: https://doi.org/10.1186/s13677-019-0146-7

[5]. Abhijit Aditya, Uddalak Chatterjee and Snehasis Gupta, “A Comparative Study of Different Static and Dynamic Load Balancing Algorithm in Cloud Computing with Special Emphasis on Time Factor ”,International Journal of Current Engineering and Technology, Vol.5, No.3 (June-2015)

[6]. Bhawesh kumawat , Rekha kumawat,”A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment using Cloud Analyst”, IJESC Volume 7 Issue No.3.