Load Balancing in Cloud Computing

Shivangi Khare, Uday Chourasia, and Anjna Jayant Deen

Abstract Nowadays, many organizations are using cloud-based applications and platforms because of its on-demand service and quick response. The biggest concern of cloud computing is overloading of the system for any individual, group or organization. So, nowadays, load balancing is gaining popularity, its algorithms and solutions are also getting better day by day. This review article gives a brief overview that how the traditional load balancing was done, and the updates made in it with time. This paper also provides comparison table of the major research papers published on load balancing, various algorithms of load balancing, context on which load balancing is used like round robin, throttled, honey bee, max min, min-min, OLB, etc. Also, cloud service providers in CC who deal with powerful computational structures which are entirely determined on usage and offer their solutions in a stable manner. In addition, this research article provides extensive list of areas for research under load balancing and its future scopes.

Keywords Cloud computing · Load balancing · Cloud platforms · Cloud services · Cloud applications

1 Introduction

1.1 Cloud Computing

Cloud computing is very new and effective technology that is used widely in today's world. Cloud computing is the network technology that provides on-demand services to its users. It provides hardware as well as software applications with platforms and testing tools. Cloud computing is the distribution of computer services that includes

S. Khare (✉) · U. Chourasia · A. J. Deen
University Institute of Technology, RGPV, Bhopal (M.P.), India
e-mail: Shivangikhare27@gmail.com

U. Chourasia
e-mail: udaychourasia@rgtu.net

intelligence, databases, servers, networking, storage, etc. It provides data at very low cost, and users can access resources any time whenever they need.

1.2 Cloud Service Model

Software as a service—It is an on-demand software service in which applications are delivered by service providers. It is hosted on remote server that is managed from central location.

Platform as a service—It is a kind of platform on which applications are being developed, tested, run and managed by the programmers. It provides an ability to "auto scale".

Infrastructure as a service—It is also known as hardware as a service. It is cost-effective that means it provides user to avoid cost and complexity of managing and purchasing the physical servers.

1.3 Load Balancing

Load balancing is the contribution of work among different machines or nodes ensuring that no computing device will be overloaded. It enables organization to control or manage applications demand by giving out resources among servers, networks or computers. It is used to circulate the traffic that is on cloud server in order to maintain the system stability. Also, it maintains different parameters of cloud server, i.e. it increases response time, execution time and therefore, it increases the overall performance of the cloud (Fig. 1).

1.4 Parameters of Load Balancing

Overhead—The overhead supports an extra cost of merging the algorithms.

Scalability—Scalability is the kind of parameter by which load of the system is managed or handled. It used to increase or decrease the resource as per need.

Fault tolerance—Whenever the failure occurs in some of cloud components, then fault tolerance helps to continue the operation.

Response time—The time which is taken by the server to process any task.
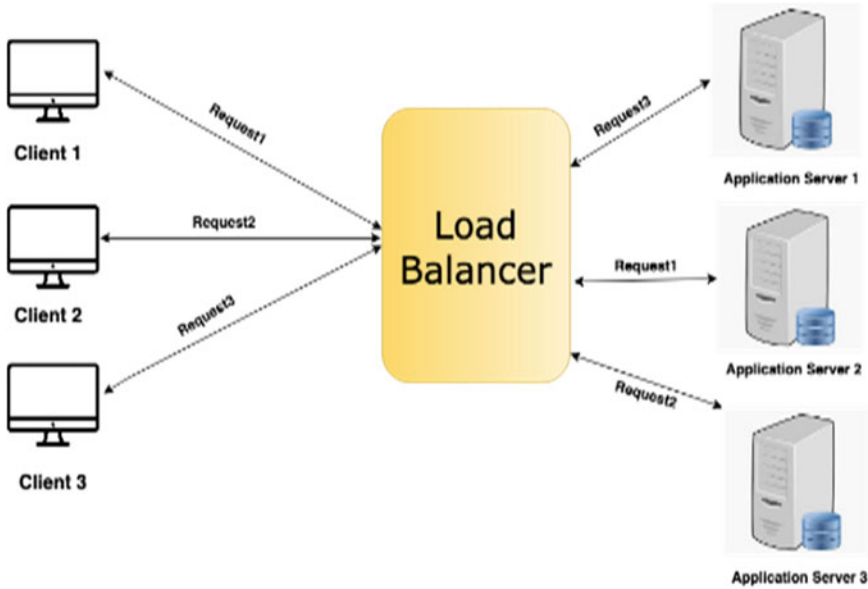
Fig. 1 Load balancing phenomena

2 Literature Survey

In throttled algorithm, Violetta N. Volkova suggested that the load balancer controls a table of virtual machine indexes, as well as their location whether it is reachable or occupied. In the given algorithm, the load balancer controls an index table of a virtual machines and their specific positions. Data centre needs to ask for virtual machine allocation for desired load balancer. Whenever implementing order of consumer, the load balancer gives-1 to the data centre if efficient virtual machine does not get recognized. So the consumer first requests the balancer to pick an optimal virtual machine for implementing essential activities [1].

Dina and Fatima proposed storing and retrieving data on Internet rather than hard drive of computer. The components taken part in cloud computing are clients, datacentres and distributed servers. It means to divide the workload among different nodes equally so that no single node will be overloaded. In this, we introduced an architecture and algorithm which will upgrade the load balancing problems by reducing the response time. We have introduced the amplified version of subsist modulate balancing proposal for cloud computing greedy algorithm of balancing the load. To check the performance of presented approach, we have used the cloud analyst simulator [2].

Scheduling of task that depends on LB-Y. Fang proposed a two-level scheduling of task technique depends on to meet dynamic demands of customers which acquire high resource deployment. It receives load balancer by first mapping task to

virtual machines and then upgrading the performance, response time and resource deployment of the cloud environment [3].

Priyanka Sagwan and Manmohan Sharma proposed the fundamental part of operating system is multiprogramming. It is a strategy for operating various procedures all in the memory. We have rolled out a progress in round robin algorithm so that the bringing out of CPU can be moved forward and compared it with original round robin scheduling and exhibit the consequences of why our model is systematic and well organized than the original one [4].

R. Achar proposed distributed intra cloud load balancing algorithm to balance and compare depends on sampling to reach balanced solution. Cost of running virtual machine on each of the host was determined and it ensures that virtual machine locates from physical machine with respect to larger cost to those with low and high was difference of cost [5].

Gao proposed a multi-objective virtual machine placement by ant colony algorithm system. This exerts to carry out Pareto solution which at same time reduces resource power consumption and wastage. This will use the value of pheromone concentration of a virtual machine sets on the server for resolution. Requirements of a virtual machine put on the server is inversely proportionate to the amount of resource wastage and utilization associated with the placement plan [6].

Accessibility at one of the most important aspects is environmental systems. Load balancing is used in it to ensure Internet availability by distributing traffic across many data centres of computer hardware, mitigating recourse limitations and software failures. It talks about the load balancing and after reveals a case study on accessibility of system depends upon a typical management of database of hospital solution [7].

Balancing the load under cloud environment has basic effect on the performance. This object defines a best load balancer model for the public cloud, which is founded on cloud division theory and provides a switch mechanism that involves selecting multiple strategies for different circumstances. To improve efficiency in the cloud context, the system combines game theory to the computation offloading technique. [8].

Michael D proposed a graph technique to express and economic networking. It has both complex and simpler connections between platforms. They employed graph method to build some static and random fabrication networks. Here gives detailed information on these aspects of economic communication networks along a broad range of industries [9].

Bhathiya has proposed two virtual machine load balancing strategies have been deployed for load balancing in cloud server farms at the very same time. The very first algorithm is the automatic inspection base station, which divides the platform's load consistently. The throttled is the second algorithm that makes sure that only of a predefined jobs should be located to individual virtual machine at any certain given time. The major issue happens whenever the hardware layout of VMs is different and it fabricates the undergoing load and over load conditions in VMs [10].

Pham proposed a latest community depends on search algorithm known as bees algorithm. This algorithm emulates the food foddering behaviour of flocks of the

honey bees. In this bees locality, spy bees are casting for exploring food origin by moving casually from one origin to other. After when they got success in finding food origin with standard, threshold the spy bees back to swarm and dance on the floor. Bees, those having the most fitness's are taken as picked bees and locations visited by those bees are considered for neighbourhood explore [11].

Li et al. A dynamic min-min and list scheduling approach is being used as a resource provisioning strategy. This technique initiated for infrastructure as a service layer, and works depend on the job response times. To boost virtual machine, migration technique, we are putting forward a better load balancing algorithm that is stimulated from direction of the human's hormone system that is known as endocrine algorithm. This algorithm authorized the overloaded virtual machines to get off their additional jobs to suitable virtual machines through communications that is stimulated from feedbacking technique between hormones. There are benefits of this technique [12].

Sukhpreet Kaur et al. presented an improved genetic algorithm (IGA) for assigning virtual machines to users' tasks. The goal of this proposed strategy is to increase resource consumption while cutting task completion costs and spending as few effort as possible [13].

According to Sheetal Karki et al. data is kept in an unified VMs known as the cloud, and cloud service companies are responsible for sending the offerings to end clients [13].

Cloud technology is one of the areas in computation that is changing fast at the present. Companies and individuals demand vital services, and CC delivers them in the shape of IaaS, SaaS and PaaS. The Internet has made it easier to develop apps as well as provide capabilities via the web [14].

3   Comparison Tables

See Table 1.

4   Conclusion and Future Work

After having a deep exploration of all mentioned research papers evaluated in this article, it is observed that load balancing had abundant limitations. So, in recent times, many other algorithms came like flexible load sharing algorithm, dynamic load balancing, firefly algorithm, etc., which enhance the overload problem to some extent.

Then also the load balancing problem is at risk to high order because if system become overloaded then the system may be crashed, fault tolerance and other parameters. So, it can be enhanced further with some better modern algorithm or combination of two or more algorithms which can give result for improvement of balancing

Table 1 Comparative study among previous work on the basis of different parameters and performances

Parameter	Security	Cost-effectiveness	Scalability and elasticity	Execution time	Performance	Algorithm
Violetta et al.	Intermediate	Low	High	Intermediate	high	Throttled algorithm
Dina et al.	Low	High	High	Low	Intermediate	Greedy load balancing algorithm
LB-Y. Fang et al.	Intermediate	Intermediate	Low	High	Intermediate	Two-level task scheduling algorithm
Priyanka et al.	High	Intermediate	Intermediate	High	Intermediate	Round robin algorithm
R. Achar et al.	Low	Intermediate	High	High	High	Distributed intra cloud load balancing algorithm
Gao et al.	High	Intermediate	Intermediate	Intermediate	High	Ant colony system algorithm
G. Xu et al.	Intermediate	Low	High	Intermediate	High	Cloud partitioning for public cloud
R. kanniga et al.	Low	High	High	Low	Intermediate	The game theory to the load balancing algorithm
Michael D et al.	Intermediate	Intermediate	Low	High	Intermediate	Graph technique
Bhathiya et al.	High	Intermediate	intermediate	High	Intermediate	Two virtual machine load balancing algorithms
Pham et al.	Low	Intermediate	High	High	High	Bees Algorithm
Li et al.	High	Intermediate	Intermediate	Intermediate	High	Token-ring algorithm

the load. This research concentrated on the problems of cloud computing and their challenges. It is a computer technology that gives client support every time. For load balancing, this paper narrated out the effective algorithms.

Keeping this data on an open network naturally creates concerns about balancing load and its parameters. We simulate the approach on a network with over 50 nodes installed on edge computing or fog edge. The results of the experiments show that the suggested approach is effective, detecting dropping workload, data packets received on the network.

References

1. Shahid MA, Islam N, Alam MM, Su'Ud MM, Musa S (2021) A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach. IEEE Access 8:130500–130526. doi: https://doi.org/10.1109/ACCESS.2020.3009184
2. Altayeb DF, Mustafa FA, History M (2019) Analysis on load balancing algorithms implementation on cloud computing, Vol 6, no. 02, pp 32–36
3. Al Nuaimi K, Mohamed N, Al Nuaimi M, Al-Jaroodi J (2012) A survey of load balancing in cloud computing: challenges and algorithms. In: Proceedings of the IEEE 2nd symposium networking cloud computer applications, NCCA 2012, pp 137–142. doi: https://doi.org/10.1109/NCCA.2012.29
4. Sangwan P, Sharma M, Kumar A (2017) Improved round robin scheduling in cloud computing. Adv Comput Sci Technol 10(4):639–644
5. Mohamed Shameem P, Shaji RS (2013) A methodological survey on load balancing techniques in cloud computing. Int J Eng Technol 5(5):3801–3812. doi: https://doi.org/10.9790/0661-182 15561
6. Thakur A, Goraya MS (2017) A taxonomic survey on load balancing in cloud. J Netw Comput Appl 98(February):43–57. https://doi.org/10.1016/j.jnca.2017.08.020
7. Chaczko Z, Mahadevan V, Aslanzadeh S, Mcdermid C (2011) Availability and load balancing in cloud computing
8. Xu G, Pang J, Fu X (2013) A load balancing model based on cloud partitioning for the public cloud. Tsinghua Sci Technol 18(1):34–39. https://doi.org/10.1109/TST.2013.6449405
9. Kanniga Devi R, Murugaboopathi G, Vijayakumar P (2017) A graph-based mathematical model for an efficient load balancing and fault tolerance in cloud computing. In: Proceedings of the 2017 2nd international conference on recent trends challenges computer model ICRTCCM 2017, pp 136–140. doi: https://doi.org/10.1109/ICRTCCM.2017.25
10. Soni G, Kalra M (2014) A novel approach for load balancing in cloud data center. In: Souvenir 2014 IEEE international advanced computing conference on IACC 2014, no. February, pp 807–812, 2014. doi: https://doi.org/10.1109/IAdCC.2014.6779427
11. T. Nadu (2014) Nature inspired preemptive task scheduling for. No. 978, pp 0–5, 2014
12. Aslanzadeh S, Chaczko Z (2015) Load balancing optimization in cloud computing: applying endocrine-particale swarm optimization. IEEE Int Conf Electro Inf Technol 165–169. doi: https://doi.org/10.1109/EIT.2015.7293424

13. Kalpana, Shanbhog M (2019) Load balancing in cloud computing with enhanced genetic algorithm. Int J Recent Technol Eng 8(2, Special Issue 6):926–930. doi: https://doi.org/10.35940/ijrte.B1176.0782S619
14. Amal Murayki Alruwaili BAA, Humayun M, Jhanjhi NZ (2021) Proposing a load balancing algorithm for cloud computing applications. J Phys Conf Ser 1. doi: https://doi.org/10.1088/17426596/1979