Sarcouncil Journal of Engineering and Computer Sciences

ISSN(Online): 2945-3585

Volume- 01| Issue- 01| 2022

Review Article **Received:** 05-03-2022 | **Accepted:** 15-03-2022 | **Published:** 21-03-2022

A Survey on Load Balancing in Cloud Computing

Nikhil Shrestha and Gajendra Sharma
Department of Computer Science and Engineering, Kathmandu University.

Abstract: Cloud computing is the most rapidly growing invention due to its ground- breaking and significant power. It is a sort of Internet-based computing that allows for the provision of numerous services to users on a cost-effective basis. Virtualization, grid computing, and utility computing are the most common emerging technologies utilized in cloud computing to make it more powerful. However, cloud computing still has a number of key challenges, such as security, load balancing, and adaptability to non-critical failure, among others. The massive growth of cloud computing will result in server overburdening. As a result, network performance will suffer as a result. A good load balancing adjustment may make cloud computing more productive and increase client fulfillment execution. This paper provides a comprehensive overview of cloud computing and load balancing approaches.

Keywords: Cloud computing, Load balancing, Virtual machine.

INTRODUCTION

As seen in Figure 1, cloud computing offers web-based services with a user's information, processing, and software. It also allows for instant access to shared resources such as servers, storage, networks, applications, and services [Samreen, S.N. *et al*., 2018]. Cloud computing gives a good platform and infrastructure to its users. Every one of the administrations provided by servers to consumers is provided by a cloud service provider (CSP), which is fundamentally analogous to acting as an Internet service provider (ISP) in electronic registration. Figure 2 depicts the virtualization idea, in which the user may access all resources over a high-speed network at a minimal cost. This innovation is intended with the new idea of organizations providing services to clients without acquiring these organizations and storing them on their local memory [Choudhary, R. *et al*., 2018]. Virtualization technology in cloud computing enables corporations or organizations to lease registering power to clients in the form of virtual computers. Clients may employ an unlimited number of virtual computers [Dhari, A. *et al*., 2017].

Figure 1: Cloud Computing [1]

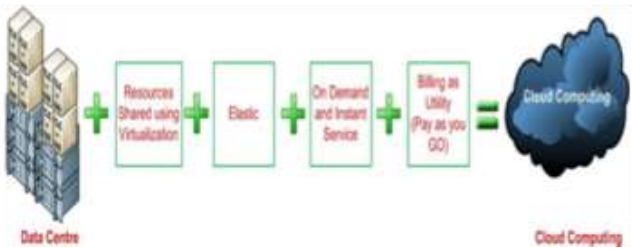

Figure 2: Schematic diagram of Cloud Computing [Diaby, T. *et al*., 2017]

End users and organizations utilize the cloud network to exchange, process, and store data in third-party data centers [Diaby, T. *et al*., 2017]. It is based on the sharing of resources using the

***Corresponding Author:** Nikhil Shrestha

virtualization technology. The memory interface for a visiting OS is virtualized and provides substantial storage capacity via a virtualized memory stockpiling system [Carlos, B.W. *et al*., 2017]. As shown in Figure 3 [Chary, C.N. *et al*., 2017], cloud computing consists of three levels: cloud customer or client or user, cloud service provider (CSP), and cloud network (transmission media or channel of the cloud).

Figure 3: Three level of Cloud Computing [Chary, C.N. *et al*., 2017]

This paper gives an overview of Load Balancing in Cloud Computing. The rest of the paper is structured as follows: Section 2 provides features and characteristics load balancing strategies. Section 3 discusses some of the deployment model. Section 4, list various research areas and Section 5 presents overview of load balancing architecture. Section 6, presents an overview of different load balancing classification and Section 7, with its advantages. Finally, Section 7 concludes paper with remarks.

Features and Characteristics

The following are the features [Fatima, N. *et al*., 2018] and characteristics [Diaby, T. *et al*., 2018] of cloud computing:

Mobility and wide network access: All necessary services are available anywhere on the globe.

Saves time: Users are able to obtain and read the information they require.

Popular: The majority of users obtain their services via the cloud.

Cost: It is incredibly cost-effective because to the pay-as-you-go model.

Maintenance: maintenance is easy in cloud computing.

Throughput: Because several people may work on the same data at the same time, cloud computing produces good results.

Reliability: The aspect of cloud computing is reliable.

Elasticity: On demand, resources and data are pooled.

Security: The information is secure.

Scalable: The size of cloud computing can be easily growing increased.

On-demand self-service: In cloud computing, clients may keep track of the server's time, capacity, and designated organizing capacity on a continuous basis..

Resources pooling: resources are being used by user as on demand.

Rapid elasticity: Its end client has access to a wide range of services as well as assets.

Measure services: Every asset that is used may be evaluated, managed, and declared for both the provider and the purchaser. IT administrations are billed on a pay-per-use basis.

Cloud Deployment Model

There are four sorts of cloud deployment models [Trilochan. *et al*., 2017]:

Private Cloud

This cloud computing is managed by a single organization or a specialized organization [Aathishvar, M. *et al*., 2018], and its infrastructure is also used by this organization.

Community Cloud

This cloud is managed by a few organizations, supports a certain network, and contains common applications or services.

Public Cloud

This concept is owned and managed by a large cloud service provider (CSP). All clients that require resources on a participation or membership basis [Aathishvar, M. *et al*., 2018].

Hybrid Cloud

It is a hybrid of at least two of the preceding models, namely the public, private, and communal models [Aathishvar, M. *et al*., 2018].

Research Areas in Cloud Computing

There is a lot of room for study in cloud administration since it encompasses so many different locations, difficulties, organizational

Publisher: SARC Publisher

strategies, and particular computing techniques. The subjects that follow provide a wealth of knowledge for cloud framework researchers. Table 1 displays a variety of research disciplines.

Table 1: Various research field in cloud computing

Load Balancing	Cloud access control
Security	Energy improvement
Virtualization	Information isolation security
Data isolation and recuperation	Verifiable calculation
Scheduling for asset improvement	Failure discovery and forecast
Cloud cryptography	Task scheduling

Load Balancing Architecture

Load balancing, as seen in Fig. 5, is composed of four principles: the client, the data center controller, the load balancer, and the computation to be used.

The following steps are used by the client to carry out a solicitation.

1. Every client request is routed to the data center controller.
2. The data center controller queues up all incoming solicitations and queries the central load balancer about solicitation assignment.
3. The central load balancer comprises a database that stores tables that are parsed after the calculation to be utilized determines the most appropriate virtual machine and returns the ID of the selected VM to the data center controller.
4. Finally, the data center controller distributes the request to the VM whose ID is provided by the central load balancer.

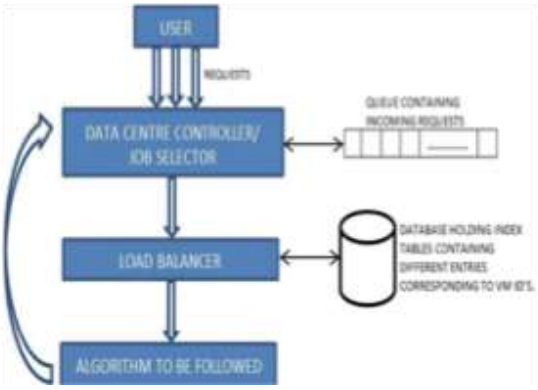

Figure 5: Load balancing architecture [Bhagyalakshmi. *et al*., 2018]

Load Balancing Classification

There are two types of load balancing: **static load balancing** and **dynamic load balancing**.

Static Load Balancing

This kind of load balancing is only employed in identical situations and is not adaptable, which means that it cannot modify its property. All of the entities have a fixed nature. This approach does not examine the incoming requests or the node's status [Sajjan, R.S. *et al*., 2018]. Here are some examples of static algorithms:

- Round robin load balancing algorithm (RR),
- Load balancing min-min algorithm (LB Min-Min),
- Load balancing min-max algorithm (LB Min-Max).

RR Algorithm

The task or assignment is allocated a defined quantum time in this computation. It distributes employment to all hubs in a roundabout way. A roundabout request is used to assign processors. Because of the proportionate remaining job at hand appropriation among methods, this calculation provides a speedier answer. Regardless, a few centers or hubs may be overcrowded, while others remain dormant and underutilized.

LB Min-Min

A list of available jobs is maintained, and the shortest fulfillment time for all available hubs is determined. The machine is assigned the job that requires the least amount of time to complete. It produces excellent results when a small amount of effort is expended.

Publisher: SARC Publisher

LB Min-Max
A list of available jobs is maintained, and the shortest fulfillment time for all available hubs is determined. The job with the longest completion time is assigned to the machine.

Dynamic Load Balancing
To spread the load, this technique makes use of the hub's previous and present states. At runtime, the user needs and resources might be adjusted. They are suitable for both homogeneous and diverse situations. There are two kinds of dynamic load balancing.

Distributed
This method of load balancing modifies the load calculation performed by all data centers and distributes the load among them.

Centralized
In this case, a central hub is in charge of balancing the load over the whole structure. This focal hub communicates with other hubs [Sajjan, R.S. *et al*., 2017].

Some common dynamic algorithms are as follows:

- Throttled load balancing algorithm,
- Equally spread current execution (ESCE) load balancing algorithm,
- Modified throttled load balancing algorithm,
- Throttled modified algorithm (TMA),
- Throttled load balancing algorithm [Sajjan, R.S. *et al*., 2017; Choudhary, R. *et al*., 2018; Phi, N.X. *et al*., 2018–Panchal, B. *et al*., 2018].

This approach is used for load distribution on VMs that are entirely deployed. When the client delivers the request, the load balancer rapidly receives notice and scans for the gathering that can supervise efficiently and distributes that request.

ESCE [Sajjan, R.S. *et al*., 2017]
This approach maintains the shutdown of whole virtual machines and jobs. When this algorithm receives a solicitation, it filters the list of VMs. If a VM that can deal with the customer's request is identified, the request is allocated to that specific VM. This calculation spreads the equal weight over all VMs.

Modified Throttled Load Balancing Algorithm [Phi, N.X. *et al*., 2018; Bhagyalakshmi. *et al*., 2018]
The load balancer maintains a file table of virtual machines including the province of VMs in this algorithm. The computation employs a strategy for selecting a VM for handling consumer solicitation in which the most easily available VM is selected. If the machine is available, it is distributed with the request and the VM's ID is returned to the data center controller; otherwise, (-1) is returned. When the next solicitation appears, the list table is filtered from record beside effectively allocated VM and the next VM is selected based on the province of VM.

Throttled Modified Algorithm (TMA) [Phi, N.X. *et al*., 2018]
The TMA load balancer adjusts the strain by updating and maintaining two record tables. Accessible index: the status of VMs is '0'. Busy index: VM status is not accessible '1'. It outperforms other load balancing algorithms in terms of performance.

Advantage of Load Balancing
Various advantages of load balancing are described as below [Narayanan, S.S. *et al*., 2016; Amandeep, V.Y. *et al*., 2014]:
Throughput: it improves the result or throughput.

Fault tolerance: system must be free from failure.

Migration: resources are moved from hub to hub to improve performance.

Response time: measuring to take the resources for service.

Scalability: improve the system scale and performance.

Cost-effective: system improves their performance with lower cost.

CONCLUSION

Cloud computing gives a good platform and infrastructure to its users. Every one of the administrations provided by servers to clients is provided by CSP, which is fundamentally analogous to acting as the ISP in electronic registration. Load balancing aids in the efficient use of resources and enhances the framework's presentation. It distributes all workload requests from diverse resources over several PCs, frameworks, or servers. As seen in Table 1, there are several study topics accessible in load balancing. The primary aims are to maintain framework solidity and to improve framework execution.

Publisher: SARC Publisher

REFERENCES

1. Samreen, S.N., Valmik, N.K., Salve, S.M., Khan, P.N. "Introduction to cloud computing." *IRJET* 5.2 (2018).
2. Diaby, T. and Rad, B.B. "Cloud computing: a review of the concepts and deployment models." *International Journal of Information Technology and Computer Science* 9.6 (2017): 50-58.
3. Chary, C.N., Ashok, D. and Dilip, P. "Advantages and solutions in cloud computing." *Int J Innovative Comput Sci Eng* (IJEACSE) 4.3 (2017):149–152.
4. Fatima, N. and Parveen, Z. "Cloud computing issues and countermeasures." *International Journal of Engineering and Applied Computer Science* 2.02 (2017): 69-74.
5. Trilochan. and Verma, A. "Cloud computing: evolution and challenges." *IJESC* 7.4 (2017).
6. Paul, V., Pandita, S. and Randiva, M. "Cloudcomputingreview." *Int. Res. J. Eng. Technol.* (IRJET) 5.3 (2018).
7. Aathishvar, M., Kumar, N.M. and Ganesh, K. "Study on cloud computing." *Int J Contemp Res Comput Sci Technol (IJCRCST)* 4.1 (2018).
8. Sindhu, S. and Sindhu, D. 'Cloud computing models and security challenges." *IJESC* 7.4(2017).
9. Narayanan, S.S. and Ramakrishnan, M. "A comprehensive study on load balancing algorithms in cloud computing environments." *Research Journal of Applied Sciences, Engineering and Technology* 13.10 (2016): 794-799.
10. Sajjan, R.S. and Yashwantrao, B.R. "Load balancing and its algorithms in cloud computing: A survey." *International Journal of Computer Sciences and Engineering* 5.1 (2017): 95-100.
11. Amandeep, V.Y. and Mohammad, F. "Different strategies for load balancing in cloud computing environment: a critical study." *International Journal of Scientific Research Engineering & Technology (IJSRET)* 3.1 (2014): 85-90.
12. Choudhary, R. and Kothari, D.A. "A novel technique for load balancing in cloud computing environment." *Int. J. Softw. Hardw. Res. Eng* 6.6 (2018): 1-5.
13. Dhari, A. and Arif, K.I. "An efficient load balancing scheme for cloud computing." *Indian Journal of Science and Technology* 10.11 (2017): 1-8.
14. Carlos, B.W., Yong, W.L., Duncan, B., Olmsted, A., Vassilakopoulos, M. and Lambrinoudakis, C. "Cloud computing 2017. The eighth international conference on Cloud Computing, GRIDs, and Virtualization, Athens, Greece." (2017).
15. Patel, S. and Bhatt, M. "Implementation of Load balancing in Cloud computing through Round Robin & Priority using cloudSim." *International Journal for Rapid Research in Engineering Technology & Applied Science* 3.11 (2017).
16. Phi, N.X., Tin, C.T., Thu, L.N.K. and Hung, T.C. "Proposed load balancing algorithm to reduce response time and processing time on cloud computing." *Int. J. Comput. Netw. Commun* 10.3 (2018): 87-98.
17. Bhagyalakshmi. and Malhotra, D. "Review paper on Throttled Load balancing algorithm in cloud computing environment." *IJSRSET* 4.2 (2018): 2395-1990.
18. Panchal, B. and Parida, S. "A review: different improvised Throttled load balancing algorithms in cloud computing environment." 5.7 (2018) www.ijetmas.com.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Shrestha, N. and Sharma, G. "A Survey on Load Balancing in Cloud Computing." *Sarcouncil Journal of Engineering and Computer Sciences* 1.1 (2022): pp 1-5

Publisher: SARC Publisher