A Comparative Analysis of Load Balancing Algorithms in Cloud Computing

Sarita Simaiya[1], R Agrawal[2]

Institute of Engineering, CSJM,
Kanpur University, UP

***Abstract*: Cloud computing is a new computational framework for facilitating pay-per-use facilities over the World Wide Web to customers anywhere and at locations worldwide. Load balancing is just an accumulation of the best ways to enhance resource utilization and the efficiency of a virtualized environment. Overloading and under-loading are significant unwanted aspects of load imbalance, but load balancing approaches provide a remedy for both. The importance of load balancing in the cloud computing environment cannot be overstated. Adequate load balancing enables optimized resource utilization by delivering resources to cloud clients on an as-needed premise under a pay-as-you-go model. Load balancing also may assist in prioritizing consumers by utilizing suitable scheduling rules. This article presents a review of load balancing methods for cloud computing.**

***Keywords*: Cloud computing, Load balancing, optimum resource utilization, performance.**

1. Introduction

Cloud Computing is a technology in that application software, platforms, and Infrastructure are delivered via the World Wide Web in response to a customer's request and need. Cloud computing is a "pay-per-use" concept that delivers services whenever they are needed with only a few clicks. The most basic requirement of cloud computing is the share and provisioning of computational capabilities, such as VMs, on a demand basis. The cloud offers a variety of infrastructures, platforms, and applications that reduce money and effort and are also secure, adaptable, and adaptable [1]. Cloud computing is becoming a widely used application in academia and industry, offering a comprehensive and scalable storing and retrieving of files. The main issues is to schedule customer orders to have a short response time, efficient resource utilization, and are not underutilized. Virtualization enhances data centre strength, speed, and quality VMs to be hosted on a standalone computer in the cloud computing environment. Cloud computing encounters several issues, i.e., energy consumption, load balancing, optimum utilization of computing resources, and effective Governance [2].

Load balancing is an important research area in cloud computing that needs more attention. With proper load balancing over the cloud, we can achieve optimum utilization of computing resources [3]. This article concentrates on various research factors- a) review of load balancing methods and general taxonomy, and b) how load balancing can be accomplished efficiently. Even as the field of cloud computing has become more popular, there are many more rigorous activities expected to arrive, the model of assigning cloud work activities equitably so that the terminals inside the cloud computing configuration get an equitable load has become more critical; this resource scheduling method is known as "load balancing." Load balancing significantly impacts cloud computing efficiency because it improves system throughput and productivity, decreases response time, and avoids overloading each resource available. This research article is divided into various sections. Section1 covers the introduction, section 2 covers the cloud computing and basic terminologies, section 3 covers the load balancing algorithms, section 4 covers the related work, and section 5 covers the conclusion and future work.

2. Cloud Computing-

Cloud computing uses virtualization technology to provide elastic applications and services by dynamically assigning hardware, applications, and evaluation metrics. In data centre networks, load balancing methods have been used. IAAS (Infrastructure as a Service), SAAS (Software as a Service), and PAAS (Platform as a Service) are all supported deployment models in cloud computing [4]. The on-demand accessibility of computer system resources, particularly data storage and processing power, without significant effective management by the consumer is referred to as cloud computing. Operations in large clouds are frequently shared among multiple regions in a data centre. As per NIST, a cloud computing can be defined as [5]–

"A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction"

a) Cloud computing model types

The cloud computing models can be divided into the following categories [6]-

a) **Private cloud-** These ecosystems are fully committed to a standard final consumer or team, and the effect is generally run next to that customer's or company's network connection. When the foundational IT infrastructure is committed to an individual customer with relatively remote connectivity, only those clouds are often known as private clouds.

b) **Public cloud-** These are cloud areas that are numerous similarities with IT infrastructure that does not belong to end customers. Microsoft Azure, Alibaba Cloud, Google Cloud, Amazon Web Services (AWS), and IBM Cloud are among the world's biggest cloud service providers.

c) **Community cloud –** This is a cloud environment designed for a particular community or professor and used by only similar types of consumers.
d) **Hybrid cloud-** This is an ostensibly solitary IT dynamic environment composed of numerous ecosystems linked by local area networks (LANs), wide area networks (WANs), virtual private networks (VPNs), and APIs.

b) Cloud computing service models

The cloud computing service models can be divided into the following categories [7], [27-29]-

a) **SaaS –** It is known as software as a service model. It is a service that enables customers to access application software managed by the CSP. SaaS apps are generally mobile or web software that can be accessed through an internet browser. The customer is responsible for automatic updates, fixing bugs, and other standard software servicing, and they communicate to cloud systems via a center console or Application programming interface.
b) **PaaS-** It is known as a platform as a service model. It signifies that the external cloud provider provides and manages the devices and software, but even then the consumer manages the applications that run on the highest point of the framework and the data on which the software depends.
c) **IaaS-** It is known as the Infrastructure as a service model. IaaS refers to a service provider who handles your Infrastructure, web server, communications system, configuration management, and storing systems on the online platform.

3. Load Balancing Algorithms in Cloud Computing-

Load balancing is a methodology in which the jobs on one VMs are transferred to the other VMs in a network without interfering with the running activity. Load Balancing methodologies are classified into three types, which are as follows [8]:

a) **A sender-based method-** When the sender initiates the resource allocation problem and applies the load balancing method.
b) **A Receiver-based method -** When the receiver initiates the resource allocation problem and applies the load-balancing method.
c) **Symmetric method-** This is a mixture of sender and recipient initiated. The massive amounts have been equally distributed, consistently, overloaded, and minimally amongst some of the system's end devices.

A "hardware-based" load balancer is a standard solution to handle software applications. Since the load balancer expects the online software's IP address, all conversation with the website goes through the first one. A load balancer is linked with one or even more exactly equal web applications within the server-side. A load balancer directs packet data to various web servers for computation based on the customer session and the load upon every web server. The hardware-based load balancer is built to handle a large amount of traffic and can increase existing [9].

3.1 Load balancing methods- Load Balancing methodologies focus on minimizing consumption of resources, empowering scalability, preventing bottlenecks, and over-provisioning. There are static and dynamic load balancing methodologies. Load balancing for Web Distributed Services with low connection percentages [10].

a) Static load-balancing methods: Suitable for more minor distributed conditions with high data speeds and negotiable network congestion.
b) Dynamic load-balancing methods: Its primary goal is to reduce transmission delay and processing time in distributed network ecosystems.
c) Hybrid load-balancing methods: All such methodologies are concerned with the symmetrical distribution of designated computational tasks and the reduction of transaction cost among distributed processing end devices.

4. Related Work-

A cloud computing technology is effective only if services are used as effectively as possible, which can be accomplished by utilizing and maintaining appropriate different cloud management. Resource governance is accomplished through optimized energy planning, distribution, and resource manageability methodologies. The review of various existing research is as follows-

4.1 Load balancing based on process initiation: Table 1 represents the review of existing load balancing methods based on process initiation.

Table 1: Load balancing methods based on process initiation

References	Method type	Method Initiation	Key benefits	
[11]	Sender Initiated	On workload arrival	Demand-driven, threshold policy	Random
[12]	Receiver Initiated	On work departure time	Demand-driven, threshold policy	Random
[13]	Symmetric	It takes place at both levels	Demand-driven, threshold policy	Random

4.2 Load balancing based on service type- Table 2 represents the review of existing load balancing methods based on method type.

Table 2: Review of Load balancing methods based on the type

References	Method type	Key points	Challenges
[14]	Static method	It takes less time for task completion	Higher delay and no priority
[15]	Dynamic method	It has minimum response time and makespan time	It cants work with higher priority jobs
[16]	Mixed method	It can manage the workload equally	It takes more time in the exaction of jobs.

4.3 Comparisons of load balancing methods- Table 3 represents the comparative review of existing load balancing methods based on performance measuring parameters.

Table 3: Comparative analysis of various load balancing methods

References	Method type	Throughput	Response Time	Performance	Fault Tolerance	Energy Consumption	Job Migration	Load balancing
[17]	Round Robin method	Yes	Yes	No	No	No	No	No
[18]	Max-Min method	Yes	Yes	No	Yes	Yes	No	No
[19]	Min-Min method	Yes	Yes	No	Yes	Yes	No	No
[20]	Honey bee method	Yes	Yes	Yes	Yes	Yes	Yes	Yes
[21]	Ant colony optimization	Yes	Yes	Yes	Yes	Yes	Yes	Yes
[22]	PSO method	Yes	Yes	Yes	Yes	Yes	Yes	Yes
[23]	Throttled method	Yes	Yes	Yes	No	Yes	Yes	Yes
[24]	First Come, First Serve method	Yes	Yes	No	NO	No	No	No
[25]	Shortest Job first method	Yes	Yes	No	Yes	No	No	No
[26]	The genetic Load Balancing method	Yes	Yes	Yes	Yes	Yes	Yes	Yes

5. Conclusion & Future Work-

In cloud computing, load balancing is a significant task for optimizing resource consumption. The prime objective of load balancing is to significantly boost customer satisfaction, maximize system throughput, and improve the effectiveness of the service system, resulting in lower energy utilization. This paper reviewed the various load balancing methods for cloud computing and presented a comparative analysis based on various performance measuring parameters. Load balancing methodologies focus on minimizing consumption of resources, empowering scalability, preventing bottlenecks, and over-provisioning. There are static and dynamic load balancing methodologies. This review presents a systematic review of various load balancing methods, provides comparative analysis and taxonomy. In the future, we will create a load balancing model for cloud computing that will balance the workload and improve the system's performance.

References

[1] Kansal, Nidhi Jain, and Inderveer Chana. "Cloud load balancing techniques: A step towards green computing." IJCSI International Journal of Computer Science Issues 9, no. 1 (2012): 238-246.
[2] Sidhu, Amandeep Kaur, and Supriya Kinger. "Analysis of load balancing techniques in cloud computing." International Journal of computers & technology 4, no. 2 (2013): 737-741.
[3] Desai, Tushar, and Jignesh Prajapati. "A survey of various load balancing techniques and challenges in cloud computing." International Journal of Scientific & Technology Research 2, no. 11 (2013): 158-161.
[4] Shaw, Subhadra Bose, and A. K. Singh. "A survey on scheduling and load balancing techniques in cloud computing environment." In 2014 international conference on computer and communication technology (ICCCT), pp. 87-95. IEEE, 2014.
[5] Mishra, Nitin Kumar, and Nishchol Mishra. "Load balancing techniques: need, objectives and major challenges in cloud computing-a systematic review." International Journal of Computer Applications 131, no. 18 (2015): 0975-8887.
[6] Ahmed, Tanveer, and Yogendra Singh. "Analytic study of load balancing techniques using tool cloud analyst." International Journal of Engineering Research and Applications 2, no. 2 (2012): 1027-1030.
[7] Chaczko, Zenon, Venkatesh Mahadevan, Shahrzad Aslanzadeh, and Christopher Mcdermid. "Availability and load balancing in cloud computing." In International Conference on Computer and Software Modeling, Singapore, vol. 14, pp. 134-140. IACSIT Press, 2011.

[8] Singh, Athokpam Bikramjit, Sathyendra Bhat, Ragesh Raju, and Rio D'Souza. "Survey on various load balancing techniques in cloud computing." Adv. Comput 7, no. 2 (2017): 28-34.
[9] Chen, Shang-Liang, Yun-Yao Chen, and Suang-Hong Kuo. "CLB: A novel load balancing architecture and algorithm for cloud services." Computers & Electrical Engineering 58 (2017): 154-160.
[10] Galloway, Jeffrey M., Karl L. Smith, and Susan S. Vrbsky. "Power aware load balancing for cloud computing." In Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp. 19-21. 2011.
[11] Gabi, Danlami, Abdul Samad Ismail, and Anazida Zainal. "Systematic review on existing load balancing techniques in cloud computing." International Journal of Computer Applications 125, no. 9 (2015).
[12] Katyal, M. and Mishra, A., 2014. A comparative study of load balancing algorithms in cloud computing environment. arXiv preprint arXiv:1403.6918.
[13] Mishra, S.K., Sahoo, B. and Parida, P.P., 2020. Load balancing in cloud computing: a big picture. Journal of King Saud University-Computer and Information Sciences, 32(2), pp.149-158.
[14] Kaur, Jaswinder, and Supriya Kinger. "A survey on load balancing techniques in cloud computing." International Journal of Science and Research (IJSR) 3, no. 6 (2014): 2662-2665.
[15] Sreenivas, Velagapudi, M. Prathap, and Mohammed Kemal. "Load balancing techniques: Major challenge in Cloud Computing-a systematic review." In 2014 International Conference on Electronics and Communication Systems (ICECS), pp. 1-6. IEEE, 2014.
[16] Kashyap, Dharmesh, and Jaydeep Viradiya. "A survey of various load balancing algorithms in cloud computing." Int. J. Sci. Technol. Res 3, no. 11 (2014): 115-119.
[17] Dasgupta, Kousik, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mandal, and Santanu Dam. "A genetic algorithm (ga) based load balancing strategy for cloud computing." Procedia Technology 10 (2013): 340-347.
[18] Sajjan, R. S., and Biradar Rekha Yashwantrao. "Load balancing and its algorithms in cloud computing: A survey." International Journal of Computer Sciences and Engineering 5, no. 1 (2017): 95-100.
[19] Mesbahi, Mohammadreza, and Amir Masoud Rahmani. "Load balancing in cloud computing: a state of the art survey." International Journal of Modern Education and Computer Science 8, no. 3 (2016): 64.
[20] Panwar, Reena, and Bhawna Mallick. "A comparative study of load balancing algorithms in cloud computing." International Journal of Computer Applications 117, no. 24 (2015).
[21] Shoja, Hamid, Hossein Nahid, and Reza Azizi. "A comparative survey on load balancing algorithms in cloud computing." In Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2014.
[22] Hans, Abhinav, and Sheetal Kalra. "Comparative study of different cloud computing load balancing techniques." In 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), pp. 395-397. IEEE, 2014.
[23] Rathore, Rahul, V. SHARMA, and KK GOLA. "A new approach for load balancing in cloud computing." International Journal of Engineering and Computer Science 2, no. 5 (2014): 1636-1640.
[24] Ajit, M., and G. Vidya. "VM level load balancing in cloud environment." In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2013.
[25] Soni, Gulshan, and Mala Kalra. "A novel approach for load balancing in cloud data center." In 2014 IEEE international advance computing conference (IACC), pp. 807-812. IEEE, 2014.
[26] Ray, Soumya, and Ajanta De Sarkar. "Execution analysis of load balancing algorithms in cloud computing environment." International Journal on Cloud Computing: Services and Architecture (IJCCSA) 2, no. 5 (2012): 1-13.
[27] Abdelltif, Ahmed Abdelaziz, Ejaz Ahmed, Ang Tang Fong, Abdullah Gani, and Muhammad Imran. "SDN-based load balancing service for cloud servers." IEEE Communications Magazine 56, no. 8 (2018): 106-111.
[28] LD, Dhinesh Babu, and P. Venkata Krishna. "Honey bee behavior inspired load balancing of tasks in cloud computing environments." Applied soft computing 13, no. 5 (2013): 2292-2303.
[29] Begum, Suriya, and C. S. R. Prashanth. "Review of load balancing in cloud computing." International Journal of Computer Science Issues (IJCSI) 10, no. 1 (2013): 343.