International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 1 Volume No.8 January - February – 2024
DOI:10.46647/ijetms.2024.v08i01.023 ISSN: 2581-4621

EFFICIENT LOAD BALANCING ALGORITHM: MINIMIZING TIME WITH THRESHOLD-BASED APPROACH

V. Mahalakshmi, Jose Reena K
Department of Computer Science, School of Computing Science, *Vels Institute of Science, Technology and Advanced Studies,* Chennai, India
Corresponding Author Orcid ID: 0009-0009-7586-0850

ABSTRACT
The trending technology "Cloud Computing" provides computation and various resources over the internet through providing diverse level of dynamic services at a cost effective rate. The effectiveness of the infrastructure and the usage of cloud services will determine the direction of cloud computing in the future. There are always issues and difficulties in the implementation process of Cloud Computing. In order to better the cloud environment for users, load balancing seeks to meet these needs. Additionally, load balancing makes sure that when workloads are distributed, a particular node is never overloaded or underloaded. It also ensures that tasks are distributed equally. The goal is to investigate on several static and dynamic load balancing techniques in the field of cloud computing. The Threshold based load balancing approach consumesleast amount of processing time at cloud data centers through experimental results.
INDEX TERMS – Cloud computing, load balancing, load balancing algorithms, static load balancing algorithms, dynamic load balancing algorithms, performance metrics

1. INTRODUCTION

In order to provide computing as a service, cloud computing combineshardware, software, networks, storage, services, and interfaces. Virtualization,an abstraction layer of hardware and software, is necessary for cloud computing. The following are theservices that cloud computing provide:
Infrastructure as a Service (IaaS): Servers, networks, data storage, andvirtualization are all managed by a cloudservice provider. The user controls the middleware, programs, and operatingsystem.
Platform as a Service (PaaS): An external cloud service provider provides and manages the application-software platform and hardware. The user controls the required applications and the programs that run on top of the cloud platform.
Software as a Service (SaaS): The cloudservice provider manages and enables the users to access software.
According to several comparison studies, load balancing is a crucial area for research. [1]. The performance and consequently the services can be greatly enhanced with better load management. Load balancing directs the traffic between available servers. Different kinds of load balancers are addressed in Section 1. The performance measures that must be taken into account for load balancing are covered in Section 2. Sections 3 and 4 discuss several static and dynamic load balancing strategies. In Section 5, multiple algorithms' analytical components are displayed. Section 6 presents the outcomes of the experimental analysis using Cloud Analyst. The final conclusion is stated inSection 7.

TYPES OF LOAD BALANCERS

The main objective of the load balancer is to allocate resources to activities in an equitable manner, which

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 1 Volume No.8 January - February – 2024
DOI:10.46647/ijetms.2024.v08i01.023 ISSN: 2581-4621

leads to efficient resource use and user satisfaction at the lowest feasible cost. In a two-level cloud computing context,load balancing can be done. [2].

(a) Virtual Machine Level: At the virtual machine level, a mapping is done between an application that is uploadedto the cloud and a virtual machine. Once the request is approved, the load balancer distributes the load of many applications to the actual computers byallocating the required virtual machine.(VM).
(b) Host Level: An application that maps host resources to the virtual machine can manage a large volume ofincoming requests.
Below is a discussion of various load balancers in further detail:
(a) Hardware Load Balancer: A physicaldevice that controls each server in a network and distributes web traffic among several networkservers [3].
(b) Network Load Balancer: The fastest load balancer tends to falter after it distributes incoming network traffic among web application servers [4].
(c) Internal Load Balancer: Mostly utilized in onsite infrastructure for network area storage,virtual machines, and physical servers are used for regulation, balancing and maintenance [4].
(d) Software Load Balancer: Servers or virtual machines running x86/64 bits that may be swiftly installed andset up with a software defined balancer [4].
(e) Virtual Load Balancer: It serves the same purpose as software loadbalancers despite being different from them. Web traffic is distributed by using the hardware load balancer's software, which is installed on virtual computers [4].

2. LOAD BALANCING PERFORMANCEMETRICS

This section's primary goal is to assess the parameters and estimate how well the load balancing algorithm will work. The following are the various performance metrics:

- ***Throughput:*** It is determined by how many requests or transactionsoccur in a second. This option mimics the capabilities of the server based on its weight. When it comes to the performance of web-based applications, thismetric is crucial. The highest throughput is always required.
- ***Average Response Time:*** The total amount of time spent beginning to fulfill the user's request after it has been processed.
- ***Fault Tolerance****:* The structure can operate under certain system conditions where it normallycould not due to the ability of the load balancing mechanism.
- ***Scalability:*** The algorithm has the ability to scale itself based on prerequisites.
- ***Performance:*** The accuracy and speed of an algorithm's various operations are thoroughlyexamined.
- ***Resource utilization:*** Monitors the overall quantity of resourcesbeing utilized.

3. STATIC LOAD BALANCINGALGORITHMS

Static load balancing technique makes use of prior information about resources and nodes. This method is useful in a homogenous environment and is not scalable. The various static load balancing algorithms arediscussed below:
a. Randomized Approach Algorithm: After obtaining a disk occupancy report, every dispatcher will pre compute a probability distribution on a range of feasible blocks-to-cells assignments that meet thediversity criterion [5]. Thisalgorithm focuses on a distributed, randomized solution to the problemof disk storage load balancing in a large cloud system, with the requirements that blocks of a file are placed on different rows and columns of the cloud matrix and that storage load dispatchers do not communicate or coordinate with each other.

DOI:10.46647/ijetms.2024.v08i01.023 ISSN: 2581-4621

b. Active Monitoring LoadBalancingAlgorithm: AMLB keeps track of every virtual machine (VM) and the different requests that are now associated with it in a table. It locates the virtual machine with the least amount of load when a request comes in. The least anxiousVM is chosen when you are invitedto designate a replacement. The known principal page is selected in that scenario. The VMID is returned to the Info Center Controller by the load balancer. It communicates the new quota across VM quotas to the active VM load balancer and directs requests to the VM identified by that ID [6]. The current load state of each VM is considered inthis algorithm.

c. Weighted Round Robin Algorithm: The Weighted RoundRobin algorithm works by assigning the maximum weight to the server with, the best specifications. For e.g., the server with the best specification is given the most weight if there are twoservers: one with quad coreprocessors and faster processing speed, and the other with dual core processors and slower processing speed. To put it briefly, the Weighted Round Robin algorithm is an instance of the Round Robin algorithm that performs well with servers that have varying specs. [7]. It can deliver more requests to servers that are busier and more capable. Waiting time and Process utilization are the parameters considered in this algorithm.

d. Threshold Based Load Balancing Algorithm: Two types of thresholds are used: Thresholds for overload and thresholds for underload.. We begin allocating tasks to a virtual machine (VM) when its task count falls below the threshold value [8]. The goal ofthis method is to reduce the amount of time that data centers spend processing requests.

4. DYNAMIC LOAD BALANCING ALGORITHMS

The dynamic load balancing technique makes the choice after gathering up-to-date data on all ofthe system's resources. This is suitable for varied environment. Below is a discussion of the several dynamic load-balancing algorithms:

a. Central Queue Algorithm: A central node that makes decisions will be present in this methodology. It maintains a queue for tracking all the requests. When a node has resources available, the local load manager sends a request to the central node for processing the requests [9].

b. Least Connections Algorithm: The least connection algorithm, as its name suggests, automatically directs users to the server with the fewest active connections that can support them over extended periods of time while they are heavily visiting. It makes preservation and equitable distribution among all servers that are available easier [4]. By confirming the quantity of server connections, it prevents a server from being overloaded andguarantees effective load balancing.

c. Stochastic Hill Climbing Algorithm: This technique is based on the imperfect approach to solving optimization problems. One local optimization approach that constantly climbs in order to increase the value is called stochastic hill climbing. It will immediately stop if none of your neighbors have a better value. This operation's primary principle is to keep trying until a solution is found or until no neighbor has a high value. [10]. Response Time is considered in thisalgorithm. Better than RR and FCFS algorithms.

d. Dynamic and Integrated Resource Scheduling Algorithm for Cloud Data centers: Compared to standard cloud load balancing methods, it offers additional advantages as it takes into account only one factor — the CPU utilization of the physical servers. However, their approach takes into account how much CPU, memory, I/O, and network bandwidth areused. They recommended that it is integrated inorder to use both virtual and physical machines. They utilized virtualization to work on actual machines that included servers, clusters, datacenters, etc., they worked on four different queues, 1) Waiting queues, 2) Requesting queue, 3) Optimizing queue, and 4) Deleting queue [11]. This algorithm focuses on the management of CPU, memory, input/output, and network bandwidth.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 1 Volume No.8 January - February – 2024
DOI:10.46647/ijetms.2024.v08i01.023 ISSN: 2581-4621

5. ANALYSIS OF VARIOUS LOAD BALANCING ALGORITHMS

Load balancing is crucial in cloud computing. The performance of thesystem is improved via load balancing. Numerous analysis and descriptions of the various cloud computing techniques are done in this study paper.
A number of observations concerning the various static load balancing methods are made. The Randomized Approach algorithm offers a distributed randomized solution to the problem of balancing the storage demands on disks in a huge cloud system. The present load level of each virtual machine is given weight via the AMLB approach. Reducing wait times and collecting information onprocess utilization are the maingoals of the Weighted Round Robin Algorithm. The threshold-based load balancing system minimizes the processing time at data centers.
These are a few remarks regarding dynamic load-balancing methods. There are numerous complex processes for which the central queue algorithm is helpful. Bycalculating the total number of server connections, the least connections algorithm efficiently divides up the burden. By using the stochastic hill climbing technique, response time is decreased. For cloud data centers, the powerful Dynamic and Integrated Resource Scheduling algorithm efficiently controls the load of CPU, Memory, I/O, and Network bandwidth.

6. EXPERIMENTAL ANALYSIS –CLOUD ANALYST TOOL

First and foremost, choosing an appropriate tool is crucial for modeling large-scale applications; thus, it appears that researchers or users go for a tool with an intuitive user interface. Thus, Cloud Analyst tool offers a graphical user interface that is included in a program Kit and allows you to configure different cloud environment characteristics. The data center processing time is considered as the performance evaluation criterion. The Active Monitoring load balancing algorithm, Weighted Round Robinalgorithm and Threshold based load balancing algorithm are takento analyze the data centre processing time. Important components of the Cloud Analyst tool include data center controller, virtual machine load balancer, cloud application service broker, internet, user base, and region. The overall data centre processing time for the three algorithms are as follows:

TABLE 1
DATA CENTRE PROCESSING TIMECALCULATION USING LOAD BALANCING ALGORITHMS

LB Algorithm Used	Avg(ms)	Min(ms)	Max(ms)
AMLB	0.60	0.06	1.12
Weighted Round Robin	0.58	0.04	1.08
Threshold Based	0.40	0.01	1.05

7. FINAL CONCLUSION

Load balancing is a major task and a major difficulty in cloud infrastructures. Its goal is to evenly and efficiently distribute the workload of incoming or dynamic traffic among all cloud data centers, ensuring optimal resource allocation and audience satisfaction.

The static and dynamic load balancing techniques are focused and their importance in relation to performance, as heterogeneity inclouds is a known reality. Static algorithms make modeling and monitoring of the environment simple, but they cannot imitate the variety of cloud nature. Although difficult to model, dynamic load balancing algorithms are ideally suited for the variety of cloud environments.

It is concluded that the threshold- based load balancing algorithm takes the least amount of time of 0.40 ms to process data at cloud data centers after analyzing the data center processing time using the methods listed in Table 1. The data centre processing times of AMLB and Weighted Round Robin algorithms are 0.60 ms and 0.58 ms, respectively.

The different load balancing strategies discussed in this research may be expanded by comparing the algorithms in order to reduce overall response times, boost throughput for a given number of jobs, manage resources efficiently by keeping an eye on their utilization, etc.,

8. REFERENCES

[1] Rajeshwari Nema, Sahana T.Edwin, "A Review: Load balancing algorithm using Cloud Analyst Environment", "*International Journal of Science, Management &Engineering Research(IJSMER)",* Vol.01, issue 5, October 2016.

[2] Jianzhe Tai, Juemin Zhang, Jun Li, Waleed Meleis and Ning Fang Mi, " ARA:Adaptive Resource Allocation for Cloud Computing Environments under Bursty Workloads", "*Institute of Electrical and Electronic Engineers*", 2011.

[3] S. K. Mishra, B. Sahoo, and P. P. Parida, ``Load balancing in cloudcomputing: A big picture," , "*J. King Saud Univ.-Comput. Inf. Sci.*," vol. 32,no. 2, pp. 149_158, Feb. 2020.

[4] Bhavya Alankar , Gaurav Sharma , Harleen Kaur , Raul Valverde and Victor Chang, "Experimental Setup for Investigating the Efficient Load Balancing Algorithms on Virtual Cloud", "*MDPI",* 21, December,2020.

[5] Liang Liu, Lance Fortnow, JinLi, Yating Wang, Jun Xu, "Randomized Algorithms for Dynamic Storage Load- Balancing", "*Proceedings of the Seventh ACM Symposium on Cloud Computing", Santa Clara, CA, USA,October 5-7, 2016*. pages 210-222,ACM, 2016.[doi]

[6] D Singh, V Bhalla, N Garg, "Load Balancing Algorithms with the Application of Machine Learning: A Review", "*MR International Journalof Engineering and Technology",*2023.

[7] Manikandan N, Pravin A, "An Efficient Improved Weighted RoundRobin load balancing Algorithm in Cloud Computing", *"InternationalJournal of Engineering andTechnology",* 2018.

[8] Shusmoy Chowdhury, Ajay Katangur, "Threshold Based Load Balancing algorithm in Cloud Computing" , *"IEEE InternationalConference on JCC",* August,2022.

[9] Vaibhav V Athani, Sandhya S, "Survey on Load Balancing Techniques in Cloud", *"International Research Journal of Engineering and Technology (IRJET)",* Volume:08,Issue:07, July 2021.

[10] Rajeshkannan R, Aramudhan M, "Comparative study of load balancing algorithms in cloud computing environment", *"Indian Journal of Science and Technology",* Vol 9(20), DOI: 10.17485/ijst/2016/v9i20/85866,May 2016.

[11] Tian W, Zhao Y, Zhong Y, Xu M, Jing C. (2011), "A dynamic and integrated load-balancing scheduling algorithm for cloud datacenters", *"IEEE International Conference on Cloud Computing and Intelligence Systems*", IEEE 2011; 311–315.

[12] 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","*Instituteof Electrical and Electronic Engineers",* July 2020.