IJAS

EFFICIENT LOAD BALANCING TECHNIQUES FOR OPTIMIZING RESOURCE UTILIZATION IN CLOUD COMPUTING ENVIRONMENTS

Krunal Chandrashekhar Pawar[1], Rishi Kushwaha[2]

[1]Research Scholar, Department of Computer Science & Engineering ,Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India
[2]Research Guide, Department of Computer Science & Engineering ,Sri Satya Sai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India

Abstract:This abstract talks about how important load balancing is in cloud computing settings, with a focus on how the service can improve resource use without affecting availability. Load balancing techniques are an important part of any system that wants to get rid of bottlenecks and run as fast as possible. With these methods, work is split between several computers or virtual machines. Some of the load balancing methods covered in this study are round-robin, least connections, and weighted round-robin. These methods are used in places where cloud computing is done. The study also talks about the problems and limits that come up when load-balancing methods are put into place. This group includes things like costs, growth, and security.

Keywords: Load balancing, cloud computing, resource utilization, high availability, virtual machines, servers, algorithms, round-robin, least connections, weighted round-robin, scaling, and security are some of the terms that go with this abstract.

1. INTRODUCTION

Load balancing is one of the most important ideas in cloud computing if you want to get the best speed, dependability, and scalability. Load balancing also makes it possible to add more resources to a system. In a cloud setting, all of the servers must handle requests from a wide range of customers, apps, and services at the same time. This means that the work must be spread out evenly across all of the servers. Load balancing technology gives a way to divide requests among the available resources so that the most resources can be used, reaction times are shortened, and high availability and fault tolerance are guaranteed. To put it another way, load balance technology is a way to get there. Cloud service companies need this technology in order to make sure that the services they offer to their clients are reliable and scalable. In this beginning piece, we'll talk about load balancing in cloud computing systems. We'll talk about how important it is, the problems it faces, and the different ways it can be set up.
Load balancing is used to make sure that tasks are spread out evenly across multiple servers. This improves speed and makes it less likely that one server will get too busy. This makes sure that the system is always up and running, so that people can get help whenever they need

it. Load balancing also makes scaling easy because it makes it possible to add more resources to a system without hurting its speed.

One of the main problems with load balancing in cloud computing is that the cloud is always changing. Because the number of users and the amount of data being handled can change quickly, it can be hard to keep track of resources by hand. Automatic load-balancing solutions are needed to make sure that resources are shared in an effective and efficient way.

Load-balancing methods in the cloud can be Round Robin, Weighted Round Robin, Least Connections, IP Hash, or Content-Based, to name a few. The Least Connections method and the IP Hash method are two more. When deciding how to split up work between servers, these systems look at things like traffic volume, server capacity, and user location.

Overall, load balancing is an important part of cloud computing because it makes sure that all available resources are used well and that users can always use services that work quickly and reliably. Load balancing also makes sure that users can always get to the tools they need. Effective load balancing can make cloud services much more scalable and available, which improves the experience for end users.

Related work

Load balancing is the main topic of research in cloud computing right now, and several studies have been done to look at the problems that come with spreading resources and services well. The following works that are relevant to this topic talk about recent changes in this field:

M. A. Shahid et al. (2020) did a lot of study on the different ways that load balancing is used in cloud computing environments and came up with a new way to deal with errors to make the system more reliable. The authors compared different ways to spread out the work and talked about their pros and cons.

Bayan A. et al. (2021) The best technique for allocating resources to cloud computing applications was shown. It uses a heuristic method for balancing the load. The authors looked at the suggested algorithm and showed how it improves the performance of the system.

Tasneem et al. (2022) gave an overview of load balancing in cloud computing settings and showed the many ways that cloud infrastructure can be optimized for resource use and service delivery. The authors looked at and analyzed the relevant previous research to figure out what was going on in this area and what problems were being faced.

Sansanwal et al. (2021) looked at the pros and cons of the different load-balancing methods used in cloud-based systems right now. The authors looked into the different load distribution methods that are used now and rated how well each method helped system performance.

M. Mehra et al. (2019) did a review of the research on software-defined network load balancing and looked into the different ways to do it well. The writers did some research on the subject and thought about what would be good and bad about working in this field.

Shalini Joshi et al. (2016) People talked about how hard it is to balance the load in cloud computer environments. To get the most out of their resources, the authors looked at what was already out there and suggested a load-balancing algorithm that uses both static and dynamic methods.

Overall, these studies give useful information about load balancing in cloud computing settings and show how important it is to use resources and services effectively. These results show how important it is to keep doing research and development in this area to deal with new problems and make load-balancing methods better.

2. METHODOLOGY

Next, we'll look at the suggested framework for ad hoc load balancing in cloud computing and explain how it works. The main goal of the suggested framework is to improve the security of cloud computing applications. It aims to create a highly available cloud environment that can prevent system failures and recover user workloads. As far as the writers know, this framework is the first to use two load balancers and a migration method to solve the problem of fault tolerance in the cloud. The following picture shows the two levels of the planned building:
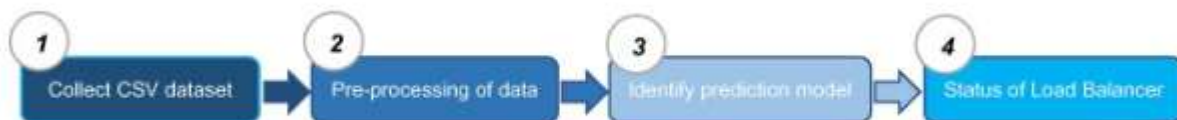

Fig. 1. In other words, this is a predictive study.

- **First of all,** answers to questions from a wide range of clients (application users), including those who use desktop computers and mobile devices. Users can send questions to the cloud from any gadget that is connected to the internet. In cloud computing, the word "Data Center" (DC) is used to describe a large building where cloud servers and data are kept. DC takes requests and sends them to the load manager that is currently in use. In the display layer of the framework, you can find both active and passive load balancers. The very top tier is where you'll find these two load-balancing parts. If something goes wrong with the active load balancer, the passive load balancer can be used as a backup. Failure can be caused by a number of things, like making too many requests to the VMs at once or moving requests around without need.
-
- **In the deepest part of it:** handle calls from users to connect them to VMs. From the picture, you can see that VM3 is in a "overloaded" state and is part of the batch of main VMs. A migration method should be set up so that failed requests can be sent to a different available VM, which may be in the group of secondary VMs. If this happens, the current Load Balancer will be turned off and marked as unavailable, which could cause the cloud-based system to be down for a long time. Passive load balancers, on the other hand, can take over in this situation and continue to re-assign requests to open VMs. The allocation table is then changed whenever a virtual machine (VM) becomes available, whether it is overloaded, idle, or has already been given a certain number of requests.

Principles of replication were used to make the system that has been made. In this literature, problems with fault tolerance in cloud computing have been mentioned as a study need. A standby load balancer can be a good way to solve these problems. There could be a problem with how the main or active load balancer is set up. This could cause problems with the cloud services and, in the end, cause the system to go down. The system explained above will be able to find these kinds of problems and fix them. Because the passive load balancer needs to be ready to take over if the active one breaks, it needs to be set up in a way that is very similar to the active one.

For VMs to work well, they should have the same amount of work, and if they get too busy, a migration method should be used to move requests, as shown in the picture to the right. In its

work, the Load Balancer will use an Optimal Migration Algorithm. The goal of this method is to spread out the load across VMs while moving as little data as possible. It figures out the Migration Threshold and makes it happen. The way the host is used can be used to figure out the barrier.

$$Utilization^{VM} = \frac{\text{Total Requested MIPS}}{\text{Total MIPS for that VM}}$$

The current Load Balancer might sometimes run migrations that aren't required. This behavior can be spotted by training a model. The load balancer keeps track of its health state in 2 different clusters, which can be checked to see if it is working correctly.

•**OK:** The tools given to each VM are enough, and the method is working as planned.
•**CRITICAL:** Too many people moved to keep track of.
By being trained, the model can figure out the best total number of moves. In case a load balancer fails, a passive load balancer must be turned on to keep handling requests and keep high availability. For this, you need a dataset, like the one you get when you run a test of the Optimal Migration Algorithm. The output of the algorithm can be saved as a CSV file with numbers that are separated by commas. Then, use the normal machine learning method below (see Figure 2) to predict the few migrations:
Once the data has been collected, it will be pre-processed to get rid of any mistakes and oddities. When a Machine Learning model is chosen for migration prediction and the number of migrations predicted is more than the maximum, the status of the current Load Balancer will be changed to CRITICAL. In the future, more research will be done to improve this framework and find the best Machine Learning tools for building a complete fault-tolerant architecture in Cloud Computing to improve the performance and availability of services. This framework will be changed so that this can happen.

3. CONCLUSION

In conclusion, load balancing is an important part of cloud computing systems because it makes sure that services work well, are reliable, and can be expanded as needed. Load balancing is becoming more and more important because cloud computing is inherently changeable and the demand for cloud-based services keeps going up. Load balancing is a technology that lets tasks be split between various servers to make better use of resources and improve response times.
Weighted Round Robin, IP Hash, Content-Based, Least Connections, and Round Robin are all examples of load-balancing methods. Two more options are Least Connections and IP Hash. Each method has its own benefits and drawbacks, and the best method to use in a certain cloud environment rests on factors that are unique to that environment.
Automated methods for handling the load are essential for managing resources well. This is especially important in the cloud, where the amount of work can change quickly. Load balancing technology is used to improve the user experience by making online services consistent, easy to use, and able to grow. Load-balancing technology is a must in this case.
Load balancing is important for cloud service providers to make sure that their users can use scalable and reliable services. It makes sure that resources are always available and can handle problems. It also speeds up response times so that resources can be used more effectively. Load balancing technology will be an important part of cloud computing for the

foreseeable future, so its development is important to the success of the business. The future of cloud computing will depend on how well this new idea works.

4. REFERENCE

[1] M. A. Shahid, N. Islam, M. M. Alam, M. M. Su'ud and S. Musa, "A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach," in IEEE Access, vol. 8, pp. 130500-130526, 2020, doi: 10.1109/ACCESS.2020.3009184.

[2] Bayan A. Al Amal Murayki Alruwaili, Mamoona Humayun, NZ Jhanjhi. Proposing a Load Balancing Algorithm For Cloud Computing Applications. International Conference on Recent Trends in Computing (ICRTCE-2021) Journal of Physics: Conference Series 1979 (2021) 012034 IOP Publishing doi:10.1088/1742-6596/1979/1/012034.

[3] Tasneem, R., Jabbar, M.A. (2022). An Insight into Load Balancing in Cloud Computing. In: Qian, Z., Jabbar, M., Li, X. (eds) Proceeding of 2021 International Conference on Wireless Communications, Networking and Applications. WCNA 2021. Lecture Notes in Electrical Engineering. Springer, Singapore. https://doi.org/10.1007/978-981-19-2456-9_113.

[4] Sansanwal, Suman and Jain, Nitin, Survey on Existing Load Balancing Algorithms in Cloud Environment (July 12, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: https://ssrn.com/abstract=3884722 or http://dx.doi.org/10.2139/ssrn.3884722

[5] M. Mehra, S. Maurya, and N. K. Tiwari, "Network Load balancing in Software Defined Network: A Survey," Int. J. Appl. Eng. Res., vol. 14, no. 2, pp. 245–253, 2019.

[6] 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.

[7] Sukhpreet Kaur, Dr. Jyotsna Sengupta, "Load Balancing using Improved Genetic Algorithm(IGA) in Cloud Computing", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 6, Issue 8, August 2017, pp 2278-1323.

[8] Wang Bei, LI Jun, "Load Balancing Task Scheduling based on Multi-Population Genetic Algorithm in Cloud Computing", 2016, Proceedings of the 35th Chinese Control Conference.

[9] Pradhan P, Behera PK, Ray BNB (2016) Modified round Robin algorithm forresource allocation in cloud computing. Proced Comp Sci 85:878–890

[10] Mishra SK, Sahoo B, Parida PP (2018) Load balancing in cloud computing: abig picture. J King Saud Univ Comp Infor Sci:1–32.

[11] Mondal B, Choudhury A (2015) Simulated annealing (SA) based load balancing strategy for cloud computing. Int J Comp Sci Info Technol 6(4): 3307–3312.Adhikari M, Amgoth T (2018) Heuristic-based load-balancing algorithm for IaaS cloud. Futur Gener Comput Syst 81:156–165.

[12] 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

[13] Vasudevan, S.K., et al.: A novel improved honey bee based load balancing technique in cloud computing environment. Asian J. Inf. Technol. 15(9), 1425–1430 (2016)

[14] Sajjan, R.S., Yashwantrao, B.R.: Load balancing and its algorithms in cloud computing: a survey. Int. J. Comput. Sci. Eng. 5(1), 95–100 (2017).

[15] Shah, N., Farik, M.: Static load balancing algorithms in cloud computing: challenges & solutions. Int. J. Sci. Technol. Res. 4(10), 365–367 (2015).

[16] Mukati, L., Upadhyay, A.: A survey on static and dynamic load balancing algorithms in cloud computing. In: Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA) (2019).

[17] S. Srivastava and R. Kumar, "Indirect method to measure software quality using CK-OO suite," 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), 2013, pp. 47-51, doi: 10.1109/ISSP.2013.6526872.

[18] Ram Kumar, Gunja Varshney , Tourism Crisis Evaluation Using Fuzzy Artificial Neural network, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011

[19] Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, "A Survey Paper on Altered Fingerprint Identification & Classification" International Journal of Electronics Communication and Computer Engineering Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278– 4209

[20] Kumar, R., Singh, J.P., Srivastava, G. (2014). Altered Fingerprint Identification and Classification Using SP Detection and Fuzzy Classification. In: , et al. Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1602-5_139

[21] Gite S.N, Dharmadhikari D.D, Ram Kumar," Educational Decision Making Based On GIS" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012.

[22] Ram Kumar, Sarvesh Kumar, Kolte V. S.," A Model for Intrusion Detection Based on Undefined Distance", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011

[23] Vibhor Mahajan, Ashutosh Dwivedi, Sairaj Kulkarni,Md Abdullah Ali, Ram Kumar Solanki," Face Mask Detection Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022

[24] Kumar, Ram and Sonaje, Vaibhav P and Jadhav, Vandana and Kolpyakwar, Anirudha Anil and Ranjan, Mritunjay K and Solunke, Hiralal and Ghonge, Mangesh and Ghonge, Mangesh, Internet Of Things Security For Industrial Applications Using Computational Intelligence (August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187998 or http://dx.doi.org/10.2139/ssrn.4187998

[25] Kumar, Ram and Aher, Pushpalata and Zope, Sharmila and Patil, Nisha and Taskar, Avinash and Kale, Sunil M and Gadekar, Amit R, Intelligent Chat-Bot Using AI for Medical Care (August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187948 or http://dx.doi.org/10.2139/ssrn.4187948

[26] Kumar, Ram and Patil, Manoj, Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies (July 22, 2022). Available at SSRN: https://ssrn.com/abstract=4182372

[27] Ram Kumar, Manoj Eknath Patil ," Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies", Turkish Journal of Computer and Mathematics Education ,Vol.13 No.3(2022), 987-993.

[28] Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, "A Survey Paper on Altered Fingerprint Identification & Classification" International Journal of Electronics Communication and Computer Engineering ,Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209.

[29] Aditya, A., Chatterjee, U., Gupta, S.: A comparative study of different static and dynamic load balancing algorithm in cloud computing with special emphasis on time factor. Int. J. Curr. Eng. Technol. 5(3), 1898–1907 (2015).

[30] Pooja Kathalkar1, A. V. Deorankar2, Challenges & Issues in Load Balancing in Cloud Computing, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887, Volume 6 Issue IV, April 2018- Available at www.ijraset.com.

[31] Aarti Singha, Dimple Junejab, Manisha Malhotra "Autonomous Agent Based Load Balancing Algorithm in Cloud Computing", International Conference on Advanced Computing Technologies and Applications , 2015 PP 832-841

[32] Muhammad Asim Shahid,Noman Islam,Muhammad Alam,M. S. Mazliham. A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach July 2020IEEE Access PP(99):1-1 DOI:10.1109/ACCESS.2020.3009184 LicenseCC BY 4.0