

Exploring Load Balancing in Cloud Computing

Archika Jain¹, Pankaj Dadheech²

¹*Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan, India-302017*

²*Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan, India-302017*

Abstract— The efficiency and performance of computer resources are hampered by the load unbalancing problem, a complicated one with many variables and limitations. Unwanted circumstances like overloading and under loading are frequently caused by this imbalance and are lessened by load balancing solutions. To the best of our knowledge, there isn't a thorough, organized, and hierarchical taxonomy of the load balancing solutions that are now in use, which diminishes their significance. Furthermore, the root causes of load unbalancing are not sufficiently examined or addressed in the literature currently in publication. This paper addresses important issues to direct the creation of future load balancing algorithms that are more successful while highlighting the benefits and drawbacks of current strategies. It also offers fresh perspectives on load balancing in cloud computing settings.

Index Terms— Load balancing, Cloud computing, Static, Dynamic.

I. INTRODUCTION

One of the technologies with an exponential growth rate is cloud computing. By meeting customer needs, cloud computing offers computers as a utility. Cloud computing offers internet resources and storage [1, 2]. In order for users to utilize their data over the internet at any time and from anyplace. Because of this, it keeps gaining popularity. It made it one of the technologies with the quickest growth. The need for the size of greater computation both increase daily. Due to an increase in the intense and inconsistent strain for cloud resources. Load balancing [3, 4] can manage this. Cloud computing is user-centric, allowing users to access and own knowledge, such as apps, movies, and images, once connected to the cloud [5-7]. It is task-centered, focusing on how applications operate within the cloud environment. The system is powerful, utilizing large computers with substantial processing power and storage capacity [8, 9]. It is also self-healing, with hot backups ensuring data

availability. Multi-tenancy enables information and costs to be shared across many users, while the intelligence of the system helps optimize its performance [10, 11]. Cloud operations can be programmable, such as automated data backups, and are flexible enough to meet the diverse needs of different consumers [12, 13].

II. LITERATURE REVIEW

One of the technologies with an exponential growth rate is cloud computing. By meeting customer needs, cloud computing offers computers as a utility. Cloud computing offers internet resources and storage [1, 2]. In order for users to utilize their data over the internet at any time and from anyplace. Because of this, it keeps gaining popularity. It made it one of the technologies with the quickest growth. The need for the size of greater computation both increase daily. Due to an increase in the intense and inconsistent strain for cloud resources. Load balancing [3, 4] can manage this. Cloud computing is user-centric, allowing users to access and own knowledge, such as apps, movies, and images, once connected to the cloud [5-7]. It is task-centered, focusing on how applications operate within the cloud environment. The system is powerful, utilizing large computers with substantial processing power and storage capacity [8, 9]. It is also self-healing, with hot backups ensuring data availability. Multi-tenancy enables information and costs to be shared across many users, while the intelligence of the system helps optimize its performance [10, 11]. Cloud operations can be programmable, such as automated data backups, and are flexible enough to meet the diverse needs of different consumers [12, 13].

A. Review outcome in "Performance in Cloud Computing"

A novel heterogeneous load balancing method efficiently distributes workloads across multiple servers by considering host specifications such as CPU, memory, and storage [14]. Virtual machines (VMs) in various datacenters are assigned weights based on their RAM, and a data structure tracks their weight and allocation counts. The algorithm identifies the least loaded VM with available resources, updates its allocation count, and assigns tasks accordingly [15-16]. This method, supported by the Best VM Assign Load Balancing Algorithm, enhances resource utilization by dynamically managing VM workloads. The load balancer indexes VMs, gathers status data, and assigns tasks to the least loaded VMs, ensuring optimal allocation of resources [17].

B. Common Findings of the Issue

- Researchers used dynamic quantum-based algorithms to enhance throughput and response time [18].
- Performance was measured using parameters like waiting time, turnaround time, context switches, efficiency, and response time [19].
- Weighted mean algorithms were explored to resolve starvation issues in SJF and priority scheduling algorithms [20].
- Dynamic time quantum was updated using techniques like mean, median, harmonic mean, integer programming, and fuzzy logic [21].
- Practical implementation of dynamic time quantum in fuzzy logic remains limited [22].
- Deadlock detection and resolution algorithms reduced probe message persistence time [23].
- Intension Locks in Distributed Database Systems introduced additional locks for field-level locking [24].
- Data-centric synchronization approaches improved performance with low annotation overhead [25]

C. Strengths

- Round Robin method faced challenges with unknown process execution durations, leading to uneven node loads [26].
- Weighted Round Robin was avoided due to the inability to predict execution times accurately [27].
- Central Load Balancing Decision Model (CLBDM) compared threshold values and connection times between users and nodes [28].
- Load Balancing Min-Min (LBMM) prioritized the smallest jobs, selecting tasks with minimum completion times first [29].

- Honeybee Foraging was applied to large-scale cloud systems [30].
- Biased random sampling selected nodes randomly, comparing walk lengths with a threshold level [31].
- Active Clustering optimized job assignments using a matchmaker node connecting similar nodes [32].

D. Weaknesses

- Map Reduce-based entity resolution generated high processing times [33].
- Central Load Balancing Decision Model using Round Robin faced network congestion due to many ants at the threshold level [34].
- Load Balancing Min-Min algorithm experienced starvation as execution channels were fewer than fetch channels [35].
- Walk-length computation added extra burden, reducing execution performance [36].
- High storage capacity was required for complete data file duplication [37].
- Modified throttled algorithm was not used due to priority factor calculations, which reduced datacenter reliability and latency [38].

E. Gaps

- Focus was primarily on reducing reaction time (RT) in process scheduling, with minimal discussion on processing costs [39].
- Load balancing was achievable by minimizing queue and job procedure times, considering varying hardware configurations [40].
- Efforts aimed at boosting customer satisfaction and optimizing resource utilization [41].
- Central Load Balancer assigned tasks to virtual machines based on priorities and statuses for better efficiency [41].

III. LOAD EQUALISATION TECHNIQUE

Load equalization is a technique that divides a virtual instance's workload among various resources on a distant node in a network without stopping any of the active operations [10].

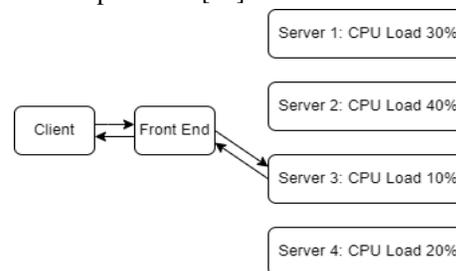


Fig 1: Cloud Load Balancing

A. Type of load balancing algorithms

Static Load equalization: This formula needs previous information concerning system and resources on the market in it. Thus, the choice of load distribution doesn't rely upon this state of the system [13].

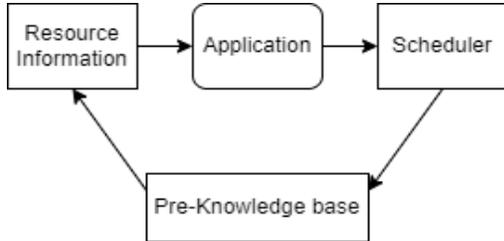


Fig 2: Static Load Balancing

Dynamic Load Balancing: Dynamic load balancing changes the load at run time [13].

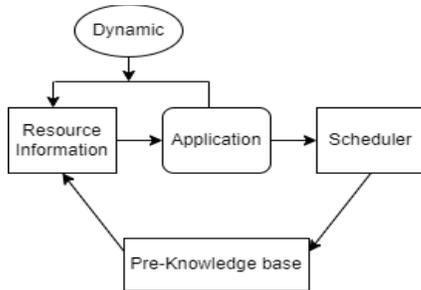


Fig 3: Dynamic Load Balancing

B. Challenges and issues in load balancing

Load balancing faces several challenges and issues that impact its effectiveness. Throughput, for example, reflects the system’s efficiency, indicating how well it performs and the execution time needed for tasks assigned to processors. Overhead refers to the additional computational cost incurred during the execution of the load balancing algorithm, which can reduce overall efficiency. Fault tolerance is another key challenge, as the algorithm must be capable of maintaining consistent performance even when a node fails within the system. Migration time is the maximum time required to transfer tasks between machines, which can affect the responsiveness of the system. Response time is also crucial, as it measures the duration it takes for the algorithm to complete a task. Effective resource utilization is essential, showing how efficiently the system’s resources are being used. Scalability, the ability of the system to handle load balancing with a limited number of processors or machines, is important for maintaining performance under varying loads. Overall performance improves when these parameters are optimized, ensuring the system functions effectively under different conditions.

IV. CONCLUSION

In order to follow through and get a definitive conclusion of the study objectives, the analysts defined a number of procedures and methods. Analyzed the 41 research articles on various load balancing techniques in cloud computing that were published between the years of 2004 and 2023. In this we talked about different type of load balancing techniques, findings of the papers, strengths, weaknesses and gaps. On the basis of gaps decrease the RT (reaction time) of process scheduling, reduce the amount of time spent in the queue and throughout the task operation and improve the quantitative link between resource utilization and customer satisfaction. The Central Load Balancer sends tasks to virtual machines in line with their priorities and statuses for increased effectiveness and network functioning. To construct a system for process fault tolerance and maintain instance stability.

REFERENCES

- [1] P. Mell and T. Grance, “The NIST Definition of Cloud Computing”, National Institute of Standards and Technology, Sept. 2011.
- [2] E. J. Qaisar, “Introduction to Cloud Computing for Developers: Key concepts, the players and their offerings”, in proc. of IEEE Information Technology Professional Conference (TCF Pro IT), pp. 1-6, 2012
- [3] O. A. Rahmeh, P. Johnson and A. T. Bendiab, “A Dynamic Biased Random Sampling Scheme or Scalable and Reliable Grid Networks”, in INFOCOMP - Journal of Computer Science, vol.7, no.4, pp. 1-10. Dec. 2008
- [4] M. Randles, D. Lamb and A. Bendiab, “Experiments with Honeybee Foraging Inspired Load Balancing”, in IEEE proc. of 2nd International Conference on Developments in eSystems Engineering(DESE), pp. 240-247, 2009.
- [5] M. Randles, D. Lamb and A. Bendiab, “A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing”, in IEEE proc. of 24th International Conference on Advanced Information Networking and Applications Workshops, pp. 551-556, 2010. [7] Hu, J., Gu, J., Sun, G., and Zhao, T. A Scheduling Strategy on Load Balancing of

- Virtual Machine Resources in Cloud Computing Environment. in 3rd International Symposium on Parallel Architectures, Algorithms and Programming. IEEE, pp. 89-96, (2010).
- [8] Fang, Y., Wang, F., and Ge, J., A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing. Lecture Notes in Computer Science, Jg. 2010(6318): pp. 271-277,(2010).
- [9] S. Wang, K. Yan, W. Liao and S. Wang, "Towards a load balancing in a three- level cloud computing network," in the proc. of IEEE 3rd International Conference on Computer Science and Information Technology (ICCSIT), pp. 108-113, July 2010.
- [10] R. Lee and B. Jeng, "Load-balancing tactics in cloud," in IEEE proc. of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 447-454, October 2011
- [11] A. Jaroodi, J. Mohamed and N. Mohamed, "DDFTP: Dual-Direction FTP," in the proc. of 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 504-503, May 2011
- [12] Y. Lua, Q. Xiea, G. Kliotb, et al. "Join-Idle-Queue: A Novel Load Balancing Algorithm for Dynamically Scalable Web Services", in 29th International Symposium on Computer Performance, Modeling, Measurements and Evaluation, pp.1056-1071, 2011.
- [13] J. M. Galloway, K. L. Smith and S. S. Vrbsky, "Power Aware Load Balancing for Cloud Computing", in IEEE proc. of the World Congress on Engineering and Computer Science(WCECS), pp. 19-21, October, 2011
- [14] J. Ni, Y. Huang, Z. Luan, J. Zhang and D. Qian, "Virtual machine mapping policy based on load balancing in private cloud environment", in IEEE proc. of International Conference on Cloud and Service Computing (CSC), pp. 292-295, Dec. 2011
- [15] N. Kumar, P. Sharma, et al. "Load Balancing of Nodes in Cloud Using Ant Colony Optimization", in IEEE proc. of 14th International Conference on Modeling and Simulation, pp. 3-8, 2012.
- [16] C. L. Hung, H. H. Wang, and Y. C. Hu, "Efficient Load balancing Algorithm for cloud computing network", in International Conference on Information Science and Technology (IST), pp. 28-30, April 2012
- [17] J. kaur, "Comparison of load balancing algorithms in a Cloud", in International Journal of Engineering Research and Applications (IJERA), vol. 2, issue 3, pp. 169-173, 2012
- [18] T. Y. Whu, W. T. Lee, Y. S. Lin, et al. "Dynamic load balancing mechanism based on cloud storage", in proc. of IEEE Computing, Communications and Applications Conference (ComComAp), pp. 102-106, January, 2012
- [19] M. Sharma and P. Sharma, "Efficient Load Balancing Algorithm in VM Cloud Environment", M.Tech. Dissertation, Information Technology Department, Dharmsinh Desai University, 2012.
- [20] James, J. and Verma, B., Efficient Vm Load Balancing Algorithm for a Cloud Computing Environment. International Journal on Computer Science and Engineering (IJCSSE), 4(09): pp. 1658-1663,(2012).
- [21] Sethi, S., Anupama, S., and Jena, K., S, Efficient load Balancing in Cloud Computing using Fuzzy Logic. IOSR Journal of Engineering (IOSRJEN), 2(7): pp. PP 65-71,(2012).
- [22] Subramanian S, N.K.G., Kiran Kumar M, Sreesh P, and G R Karpagam, An Adaptive Algorithm for Dynamic Priority Based Virtual Machine Scheduling in Cloud. International Journal of Computer Science Issues, 9(6),(2012).
- [23] Ratan, M. and Anant, J., Ant colony Optimization: A Solution of Load Balancing in Cloud. International Journal of Web & Semantic Technology (IJWesT), III,(2012).
- [24] Y. Sahu, R. K. Pateriya and R. K. Gupta, "Cloud Server Optimization with Load Balancing and Green Computing Techniques Using Dynamic Compare and Balance Algorithm", in IEEE proc. of 5th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 527- 531, 2013.
- [25] Sharma, T. and Banga, V.K., Efficient and Enhanced Algorithm in Cloud Computing. International Journal of Soft Computing and Engineering (IJSCE), 3(1),(March 2013).

- [26] Sharma, T. and Banga, V.K., Proposed Efficient and Enhanced Algorithm in Cloud Computing. International Journal of Engineering Research & Technology (IJERT), 2(2),(2013).
- [27] G. Soni and M. Kalra, "A Novel Approach for Load Balancing in Cloud Data Center", in IEEE International Advance Computing Conference (IACC), pp. 807-812, 2014.
- [28] S. G. Damanal and G. R. Reddy, "Optimal Load Balancing in Cloud Computing by Efficient Utilization of Virtual Machines", in IEEE proc. of 6th International Conference on Communication Systems and Networks (COMSNETS), Jan. 2014
- [29] Singh, A., Bedi, R., and Gupta, S., Design and implementation of an Efficient Scheduling algorithm for load balancing in Cloud Computing. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(1),(2014).
- [30] Dave, S. and Maheta, P., Utilizing Round Robin Concept for Load Balancing Algorithm at Virtual Machine Level in Cloud Environment. International Journal of Computer Applications, 49(4),(2014).
- [31] Singhal, U. and Jain, S., A New Fuzzy Logic and GSO based Load balancing Mechanism for Public Cloud. International Journal of Grid Distribution Computing, 7(5): pp. 97-110,(2014).
- [32] J. M. Galloway, K. L. Smith and S. S. Vrbsky, "Power Aware Load Balancing for Cloud Computing", in IEEE proc. of the World Congress on Engineering and Computer Science(WCECS), pp. 19-21, October, 2011.
- [33] J. kaur, "Comparison of load balancing algorithms in a Cloud", in International Journal of Engineering Research and Applications (IJERA), vol. 2, issue 3, pp. 169-173, 2012.
- [34] T. Y. Whu, W. T. Lee, Y. S. Lin, et al. "Dynamic load balancing mechanism based on cloud storage", in proc. of IEEE Computing, Communications and Applications Conference (ComComAp), pp. 102-106, January, 2012.
- [35] R. Lee and B. Jeng, "Load-balancing tactics in cloud," in IEEE proc. of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 447-454, October 2011.
- [36] X. Ren, R. Lin and H. Zou, "A dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast", in IEEE proc. of International Conference on Cloud Computing and Intelligent Systems (CCIS), pp. 220-224, Sept. 2011. [37] A. Jaroodi, J. Mohamed and N. Mohamed, "DDFTP: Dual-Direction FTP," in the proc. of 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 504-503, May 2011.
- [38] M. Sharma and P. Sharma, "Efficient Load Balancing Algorithm in VM Cloud Environment", M.Tech. Dissertation, Information Technology Department, Dharmsinh Desai University, 2012.
- [40] S. G. Domanal and G. R. Reddy, "Load Balancing in Cloud Computing Using Modified Throttled Algorithm", in the proc. of IEEE Cloud Computing in Emerging Markets (CCEM), pp. 1-5, October, 2013.
- [41] Kumar, Pramod, Mahesh Bunde, and Devendra Somwansi. "An adaptive approach for load balancing in cloud computing using mtb load balancing." 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE). IEEE, 2018.