International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 01, January 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A STUDY OF LOAD BALANCING APPROACHES IN CLOUD COMPUTING

M. Arun[1], Dr. M. Marikkannan[2], R.Siva Sankar[3],

[1]Assistant Professor, [2]Senior Assistant Professor, [3]Assistant Professor

Department of CSE, Builders Engineering College, Kangayam, Tiruppur Dt.,Tamilnadu, India.

ABSTRACT

Cloud computing is a kind of Internet-based computing which provides different services over the internet, which eliminate the need for personalized hardware and other resources. It can be implemented in various fields in recent years, such as business, research, industry, and computing to provide computation and resources over the internet via dynamic provisioning of services. Several challenges and issues associated while implementing Cloud Computing. This research paper deliberates on one of the main problems in cloud computing i.e., load balancing (LB). The aim of load balancing is to equilibrate the computation process on the cloud servers such that no host is under or overloaded. Load balancing helps us to acquire optimal resource utilization and maximize throughput and also to provide effective administration and satisfying customer requests for appropriate cloud nodes. In Cloud Computing, Load Balancing algorithms can be applied in Static, Dynamic and Centralized environments. In this paper, analysis of load balancing algorithms along with its advantages and disadvantages have been discussed.

Keywords: Cloud Computing, Datacentre, Load Balancer, Virtual Machine, Cloud Service provider, scheduling, Static, Dynamic, Round Robin, Throttled.

I. INTRODUCTION

Cloud computing services play a major role in today's computing which allows sharing hardware and software resources like networks, servers, storage, software, applications, etc. all over the Internet. Top IT companies like Amazon's AWS, HP, Microsoft, and Google deploy large data centers with extensive hardware networks for effective service delivery to cloud clients. Cloud computing is computing-as-a-service which means that rather than owning its own computing infrastructure or data centers and also offers a pay-as-you-go method in which there is no wastage of resources, users only need to pay for services used. Cloud service providers (CSPs) needed proper resource management and provisioning to allow clients to access cloud services over the Internet. The cloud provides an internet-based scalable, elastic, flexible, utility, and metered computing platform based on a significant financial pay-per-use model.

The main objectives of cloud are to reduce the cost, enhance response time, provide better performance; hence cloud is also called as a pool of Service which allows the users to access the information at anywhere at any time needed by the users. The user need not be present at the same location as the hardware on which the data is stored. The user just needs the internet connectivity to access the services of the cloud-by-cloud service

provider. It delivers all the services needed by the user dynamically through the internet based on service-level agreement (SLA) established between the cloud vendor and service consumers through negotiation [1].

Cloud computing service can be provided in different ways such as, the applications as a service over the Internet and the hardware and systems software in the data centers [2]. The applications are referred to as Software-as-a-Service (SaaS), and the hardware and system software delivered are referred to as Infrastructure-as-a- service (IaaS) and Platform-as-a-Service (PaaS), respectively [3]. some other services offer by CSPs are listed as follows, Database as a Service (DaaS), Storage as a Service (SaaS), Network as a Service (NaaS), Expert as a Service (EaaS), Communication as a Service (CaaS), Security as a Service (SECaaS), Monitoring as a Service (Maas), Testing as a Service (TaaS).

II. LOAD BALANCING

In Cloud Computing, Load Balancing is an important and challenging issue to handle networked or distributed system in such a way that task the allocated to all processor for the efficient utilization of the resources less response time, equal distribution of load and less power consumption, provide services to achieve complete resource utilization.

The term "load" means not only website traffic but also CPU load, network load, and server storage capacity. Load balancing can be referred to as that every machine in the network gets a similar measure of burden at some random time which means that none of them are overburdened or underutilized.

To solve the load balancing issue globally a proper scheduling algorithm is needed to pick the next task in such a fashion to attenuate the execution time, utilization of resources at the data centers. For balancing load across the network many load balancing algorithms have been proposed time-to-time.

Main Aim of Load Balancing Algorithms are 1. Maximize the throughput of the network, 2. Minimize the system overhead, 3. Ensure the reliability in the service, 4. Scalability of the system, 5. Resources are efficiently utilized, 6. Fault Tolerance. Load Balancing can be broadly classified into two strategies: one is Static load balancing, another one is Dynamic Load Balancing which can be disguised below.

A. Static Load Balancing:

Static load balancing uses prior information about all the characteristics of the task, the computing resources, and communication network memory are known and provided. Static load balancing algorithm is a type of non-preemptive.

B. Dynamic Load Balancing:

Dynamic load balancing is the way workloads and computing resources are distributed over one or more servers. Here, no prior information about jobs is known. This type of distribution guarantees greatest performance in minimal reaction time. Two or more servers, hard discs, network interfaces or other computer resources are separated to allow better use of resources and time spent responding to system operations.

The main Objective of Load Balancing is Control of traffic surges on one server, Reducing response time for user requests, ratio of resource usage to be improved, performance improvement, Maintaining the system's stability, increase the system's flexibility, workflow and waiting time in a line are minimized.

III. ADVANTAGES OF LOAD BALANCING

- **High-Performance Applications:** Cloud load balancing approaches are less costly and easier to adopt. Companies are able to make their customer apps operate quicker and perform better than at possibly cheaper prices.
- **High scalability:** Cloud balancing helps keep the website traffic scalable and agile. It simply increases user traffic with effective load balancers and distributes it across several servers or network devices. Frontline sites that are dealing with thousands of internet visitors per second, it is very vital. During sales or other promotional services such effective load balancers are needed to spread workloads.
- **Capacity to deal with traffic surges:** During any result announcement, an usually functioning University site might entirely go down because, the possibility of receiving requests at a high scale. They won't have to worry about traffic surges if they use cloud load balancers. Whatever the size of the request, it may be carefully spread among several servers to yield the best results in the shortest amount of time.
- **Continuity with full flexibility:** The main goal is to save or defend a site from abrupt failures by deploying a load balancer. Even if one node fails, the workload may be moved to another active node when distributed among several servers or network units.

IV. LITERATURE SURVEY

Jodayree, Mahdee, Mahmoud Abaza, and Qing Tan [11], proposed a predictive load balancing algorithm in the cloud; it helps to maintain service quality and minimize energy consumption of cloud networks. Here they have compared two algorithms round robin and random algorithm with the proposed C-Rule algorithm, it is a rule-based algorithm. It allows the cloud service providers to prepare their own resources to allocate, the prediction is based on Cicada predictions. It helps in achieving faster and more reliable load balancing algorithms for cloud environments by reducing the CPU waiting time. Li, Dapu, et al [12], author proposed Hybrid genetic algorithm and particle swarm optimization algorithm (HGAPSO) for the load balancing of molecular dynamics simulations. They have combined two algorithms Genetic Algorithm and Particle Swarm Optimization on heterogeneous supercomputers. By combining these two algorithms CUDA is used for the simulation and gives less waiting time for more VMs. Adhikari, Mainak, and Tarachand Amgoth [13], proposed The Heuristic-based load-balancing algorithm (HBLBA) is a load balancing and task scheduling algorithm for the IaaS cloud environment. The algorithm is divided into two steps. In the first step an efficient strategy is used to find the best feasible server configurations to decide the requirement and type of VMs. In the second step to minimize the waiting time mapping, a queuing model is used to schedule the tasks on VMs. The proposed algorithm is compared with the existing algorithms, round robin and min-max in terms of makespan, waiting time, and resource utilizations.

Yelchuri Venkata Sai Harsha, Nagaraj G Cholli [14], author clearly explains about the key hurdle in cloud computing and also discusses many static and dynamic algorithms. As it is a known fact that cloud is heterogeneous in nature. Static algorithms provide easy modeling and environment monitoring but dynamic algorithms are hard to model, but are well adapted for the diverse nature of cloud environments. Zhu, Yongfei, et al [15], suggested an Improved Particle Swarm Optimization for load balancing in cloud computing; it is

praised highly nowadays. The proposed algorithm is compared with the red-black tree algorithm; it improves the efficiency and reduces the drawback algorithm. From the result, it proved that the speed of PSO and improved PSO is the same. It gives better results in terms of task solving and load balancing compared to the other algorithms. Sharma, Shabnam, Ashish Kr Luhach, and S. A. Sinha [16] paper, proposed Bat Algorithm based on the natural behavior of bats. When bats find its prey, it flies by changing the velocity, and positions based on the distance between the prey and unpredictably itself. Here, the tasks are considered as artificial bats and virtual machines are considered as prey. They have improved the response time and imbalance degree but, it gives low migration time and scalability. Ghumman, Navtej Singh, and Rajwinder Kaur [17], combine both improved max-min algorithm and ant colony algorithms is a hybrid-based approach for load balancing algorithm for clouds. Improved Max-Min is used to reduce the execution time and combine with Ant colony to find optimum load balancing which minimizes the processing time and cost.

Dam, Scintami, et al [18] proposed a new algorithm by combining Genetic Algorithm and Gravitational Emulation Based Hybrid Load Balancing algorithm with the gravitational emulation local search (GEL). Patni, Jagdish Chandra, et al [19], authors have proposed a load balancing algorithm based on Genetic algorithm (GA) for the cloud environment. The proposed algorithm is a dynamic algorithm. It tries to minimize the completion time of the task. GA is a soft computing approach and comes under heuristic search processes. They have tried to overcome the problem of the inappropriate distribution of the execution time, which is used to create the traffic on the server. Muhammad Asim Shahid, Noman Islam, Muhammad Mansoor Alam, Mazliham Mohd Su'ud, And Shahrulniza Musa [20], described the numerous algorithms for LB & their static load balancing algorithm, dynamic load balancing algorithm. Dynamic nature inspired load balancing algorithm types. The new dynamic LB algorithms which allow better use of resources, minimum make-span, and an improved degree of mismatch, effective task migrations, and minimum time span. Iehab AL Rassan, and Noof Alarifi [21], suggested the new approach which combines the advantages of different task allocation algorithms like Round Robin algorithm, and Random allocation. Kaushik Mishra and Santosh Kumar Majhi [22], proposed binary BSO-LB algorithm initially evaluates the position of particles and keeps updating the position at each iteration. It is mainly used for minimizing the makespan and maximizing resource utilization by using a proper fitness function. L.Shen, J.Li, Y.Wu et al [23], proposed the ABC algorithm based on clusters. This algorithm reduced the make span, response time and resources utility. Hence achieve good load balancing and also adaptivity. K.D. Patel [24] worked on cloud systems and explored better load balancing by adding a pair of algorithms. For priority-based tasks they used modified honey bee behavior encouraged algorithm and for non-priority tasks used enhanced weighted round-robin algorithm. The algorithm was efficient because it enhanced performance of the system, increased the utilization of resources and minimized the completion time.

N.Chitgar, H.Jazayeriy, M.Rabiei [25] proposed an algorithm called DSCTS: Dynamic Stochastic Cloud Task Scheduling. This algorithm enhanced the balancing and rescue utilization by decreasing the cost and making the system more efficient. To execute the tasks depending upon the inductive probabilities for the in-line tasks in every repetition the method selected the finest virtual machine. The performance measures used to check the efficiency of the algorithm are makespan, average utilization (AU), degree of imbalance, time complexity.

A.Alzaqebah, R. Al-Sayyed, R. Masadeh [26] explored the work in which the modification was made by using multi-objectives fitness function rather than a single one. In order to compute the solution two objectives used in the fitness function were cost and make span. They proposed MGWO using multi-Objective function and GWO algorithm and hence named it MGWO. G.Natesan, A.Chokkalingam [27] proposed technique has been evaluated with two datasets (normal and uniform distribution). Using the proposed mean-GWO technique result obtained from the simulation, it showed an improvement in the performance of the task scheduling process when compared to original PSO and GWO techniques. R.Agarwal, C.Science, N.Baghel [28] proposed the Mutation Based PSO (MPSO). MPSO was made to reduce time reduction and fitness function improvement in hope of optimized load balancing in cloud computing. As compared to PSO the MPSO algorithm gave better results. Make Span parameters used as performance metrics. For performing load balancing the proposed algorithm supported preemptive VM scheduling. H.Chen, F.Wang et al. [29] proposed an enhanced LB algorithm that worked on the idea of user-priority of task scheduling relying on Min-Min algorithm (LBIMM) and also gave PA-LBIMM algorithm. These two scheduling algorithms are used to maximize throughput in a cloud environment. Implementation of these proposed new algorithms occurred with a simulation program using Matlab toolbox. Both the proposed LBIMM and PA-LBIMM algorithm reduced the completion time of tasks, enhanced load balancing of resources in all possible cases as shown in experimental results. Also, they acquired better overall performance as compared to the Min-Min algorithm. PA-LBIMM performed well as compared to both Min-Min and LBIMM by reducing average completion time of VIP tasks.

Lung-Hsuan Hung, Chih-Hung Wu, Chiung-Hui Tsai, And Hsiang-Cheh Huang [30], proposed a two-stage genetic mechanism for the migration-based load balance of virtual machine hosts (VMHs) in cloud computing. In the first stage, performance models of virtual machines (VMs) are extracted from their creating parameters and corresponding performance measured in a cloud computing environment. In the second stage, with the VMH loads estimated by GEP, the genetic algorithm considers the current and the future loads of VMHs and decides an optimal VM-VMH assignment for migrating VMs and performing load-balance. Dalia Abdulkareem Shafiq, Noor Zaman Jhanjh Azween Abdullah, And Mohammed A. Alzain[31], aims to provide an enhanced Load Balancing algorithm which reduces Makespan and provides efficient resource utilization of 78% compared to existing Dynamic LBA. It can function in a dynamic cloud environment where user requests arrive in random order and where there are many changes in the length of the user requests. Mirza Mohd Shahriar Maswood, Rahinur Rahman, Abdullah G. Alharbi and Deep Medhi [32], suggested the fog-cloud environment to minimize resource cost and reduce delay to support real-time applications at a lower operational cost. In this model, a composite objective function is used to minimize the bandwidth cost and provide load balancing.

International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 01, January 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

V. COMPARATIVE ANALYSIS

TABLE 1: Advantage and Disadvantages of Load Balancing Algorithm

S.No	Algorithm	Description	Advantage	Disadvantage
1.	Round Robin	Request is allocated for a fixed period of time	Equal distribution of workload	Process is not known in advance. For a larger task context switching increases.
2.	Weighted Round Robin	According to the processing capacity of VM weight is assigned.	Optimal resource utilization.	Processing time not taken into consideration.
3.	Dynamic Round Robin	It maintains VM retiring state and VM threshold state.	Cost of power consumption get reduced	Does not scale up for large data center
4.	Throttled LB Algorithm	Maintain state of VM busy or idle	Evenly distribution of load	Does not consider current state of the VM
5.	Min-Min Algorithm	Select task with least execution time	Simple to execute	Does not consider existing load
6.	Load Balancing Improved Min- Min LBIMM	Similar to Min- Min algorithm. From the all-available task, the task with the smallest completion time from the most heavily-loaded resource is calculated. If it is less than of Min-Min then the task is reassigned to the resource that produces it. The same process is repeated.	Overall completion time is reduced	Does not consider priority
7.	Max Min Algorithm	Job with higher execution time executed first.	Shorter makespan as compared to Min-Min	Shorter jobs have to wait
8.	User Priority Awarded Improved Min- Min	Divides the task in two groups according to user priority VIP an ordinary	Consider priority and makespan	No deadline

9.	Opportunistic Load Balancing (OLB)	Uses static load balancing algorithm attempt to allocate selected job available VM	Keeps all available VM busy	Does take into account the previous load
10.	Honey Bee Foraging	Distributed load balancing for self-organization	Well suited for heterogeneous environment	Increase in resource not increases efficiency

VI. CONCLUSION AND FUTURE ENHANCEMENTS

The load balancing in cloud computing has great importance which helps to improve the system performance. The paper summarizes different load balancing algorithms and their advantages and disadvantages. Some of the load balancing techniques that have been discussed mainly focus on reducing associated response time, increasing throughput, reducing make-span, and improving performance by some other parameters such as CPU, Memory and disk. The paper summarizes various techniques used for load balancing issues in cloud computing. In future, the based on the advantage and disadvantage of the algorithms disguised here are used for further computation process and analysis can be done for their efficiency can be proved.

REFERENCES

[1] Mishra S. K., Sahoo B., Parida P. P., Load balancing in cloud computing:a big picture, Journal of King Saud University-Computerand Information Sciences. 2020, 32(2), 149-58.

[2] Milan S. T., Rajabion L., Ranjbar H., Navimipoir N. J., Natureinspired meta-heuristic algorithms for solving the loadbalancingproblem in cloud environments, Computers & OperationsResearch, 2019, 110, 159-187. https://doi.org/10.1016/j.

[3]J.Prassanna, P. Ajit Jadhav, V. Neelanarayanan, “Towards an Analysis of Load Balancing Algorithm to Enhance Efficient Management of Cloud Data Centres”, 3rd International Symposium on Big Data and Cloud Computing Challenges ISBCC – 2016, VIT University, Chhenai, India, pp 143-155, Springer International Publication, DOI : 10.1007/978-3-319-30438-2-13.

[4] M. Mishra, S. Kumar, B. Sahoo and P. P. Parida, "Load balancing in cloud computing: A big picture," Journal of King Saud University-Computer and Information Sciences, 2018.

[5] M. H. Shirvani, A. M. Rahmani and A. Sahafi, "A survey study on virtual machine migration and server consolidation techniques in DVFS-enabled cloud datacenter: taxonomy and challenges," Journal of King Saud University-Computer and Information Sciences, 2018.

[6] K. Ramana and M. Ponnavaikko, "AWSQ: an approximated web server queuing algorithm for heterogeneous web server cluster," International Journal of Electrical and Computer Engineering, vol. 9, no. 3, p. 2083, 2019.

[7] Z. M. Elngomi and K. Khanfar, "A Comparative Study of Load Balancing Algorithms: A Review Paper," International Journal of Computer Science and Mobile Computing, vol. 5, no. 6, pp. 448-458, 2016.

[8] S. Sharma, A. K. Luhach and S. S. Abdhullah, "An optimal load balancing technique for cloud computing environment using bat algorithm," lndian Journal of Science and Technology, vol. 9, no. 28, 2016.

[9] K. Maheshwari and V. K. Gupta, "Load Balancing in VM in Cloud Computing Using CloudSim.," Cloud Computing, 2019.

[10] M.-A. Vasile, F. Pop, R.-I. Tutueanu, V. Cristea and J. Kołodziej, "Resource-aware hybrid scheduling algorithm in heterogeneous distributed computing" Future Generation Computer Systems, p. 51, 2014.

[11] Jodayree, Mahdee, Mahmoud Abaza, and Qing Tan. "A Predictive Workload Balancing Algorithm in Cloud Services." Procedia Computer Science 159 (2019): 902-912.

[12] Li, Dapu, et al. "A hybrid particle swarm optimization algorithm for load balancing of MDS on heterogeneous computing systems." Neurocomputing 330 (2019): 380-393.

[13] Adhikari, Mainak, and Tarachand Amgoth. "Heuristic-based load-balancing algorithm for IaaS cloud." Future Generation Computer Systems 81 (2018): 156-165.

[14] Yelchuri Venkata Sai Harsha, Nagaraj G Cholli. “Load Balancing in Cloud Computing” International Journal of Advanced Research in Computer and Communication Engineering Vol. 10, Issue 5, May 2021 DOI 10.17148/IJARCCE.2021.105146

[15] Zhu, Yongfei, et al. "A novel load balancing algorithm based on improved particle swarm optimization in cloud computing environment." International Conference on Human Centered Computing. Springer, Cham, 2016.

[16] Sharma, Shabnam, Ashish Kr Luhach, and S. A. Sinha. "An optimal load balancing technique for cloud computing environment using bat algorithm." Indian J Sci Technol 9.28 (2016): 1-4.

[17] Ghumman, Navtej Singh, and Rajwinder Kaur. "Dynamic combination of improved max-min and ant colony algorithm for load balancing in cloud system." 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2015.

[18] Dam, Scintami, et al. "Genetic algorithm and gravitational emulation-based hybrid load balancing strategy in cloud computing." Proceedings of the 2015 third international conference on computer, communication, control and information technology (C3IT). IEEE, 2015.

[19] Patni, Jagdish Chandra, et al. "A dynamic and optimal approach of load balancing in heterogeneous grid computing environment." Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2. Springer, Cham, 2015.

[20] Muhammad Asim Shahid, Noman Islam, Muhammad Mansoor Alam, Mazliham Mohd Su'ud, And Shahrulniza Musa, “A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach” Digital Object Identifier 10.1109/ACCESS.2020.3009184, Volume 8, 2020.

[21] Iehab AL Rassan, and Noof Alarifi, “Load Balancing Approach to Enhance the Performance in Cloud Computing” IJCSNS International Journal of Computer Science and Network Security, VOL.21 No.2, February 2021.

[22] Kaushik Mishra and Santosh Kumar Majhi [22], "A binary Bird Swarm Optimization based load balancing algorithm for cloud computing environment" https://doi.org/10.1515/comp-2020-0215, 2020.

[23] L. Shen, J. Li, Y. Wu, Z. Tang, and Y. Wang, "Optimization of Artificial Bee Colony Algorithm Based Load Balancing in Smart Grid Cloud," 2019 IEEE Innov. Smart Grid Technol. - Asia (ISGT Asia), pp. 1131–1134, 2019.

[24] K. D. Patel, "An Efficient Dynamic Load Balancing Algorithm for Virtual Machine in Cloud Computing," no. Iciccs, pp. 145–150, 2019.

[25] N. Chitgar, H. Jazayeriy, and M. Rabiei, "DSCTS: Dynamic Stochastic Cloud Task Scheduling," in 2019 5TH IRANIAN CONFERENCE ON SIGNAL PROCESSING AND INTELLIGENT SYSTEMS (ICSPIS 2019), 2019.

[26] A. Alzaqebah, R. Al-Sayyed, and R. Masadeh, "Task scheduling based on modified grey Wolf optimizer in cloud computing environment," 2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc., pp. 1–6, 2019, doi: 10.1109/ICTCS.2019.8923071.

[27] G. Natesan and A. Chokkalingam, "Optimal task scheduling in the cloud environment using a mean Grey Wolf Optimization algorithm," Int. J. Technol., vol. 10, no. 1, pp. 126–136, 2019, doi: 10.14716/ijtech. v10i1.1972.

[28] R. Agarwal, C. Science, and N. Baghel, "Load Balancing in Cloud Computing using Mutation Based Particle Swarm Optimization," no. Vmm, pp. 191–195, 2020, doi: 10.1109/IC3A48958.2020.233295.

[29] H. Chen and F. Wang, "User-Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing."

[30] Lung-Hsuan Hung, Chih-Hung Wu, Chiung-Hui Tsai, And Hsiang-Cheh Huang, "Migration-Based Load Balance of Virtual Machine Servers in Cloud Computing by Load Prediction Using Genetic-Based Methods" Received January 28, 2021, accepted February 16, 2021, date of publication March 10, 2021, date of current version April 6, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3065170.

[31] Dalia Abdulkareem Shafiq, Noor Zaman Jhanjhi, Azween Abdullah, And Mohammed A. Alzain "A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications" Received March 7, 2021, accepted March 8, 2021, date of publication March 10, 2021, date of current version March 19, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3065308.

[32] Mirza Mohd Shahriar Maswood, Md. Rahinur Rahman, Abdullah G. Alharbi, And Deep Medhi, "A Novel Strategy to Achieve Bandwidth Cost Reduction and Load Balancing in a Cooperative Three-Layer Fog-Cloud Computing Environment" Received May 28, 2020, accepted June 9, 2020, date of publication June 18, 2020, date of current version June 30, 2020. Digital Object Identifier 10.1109/ACCESS.2020.3003263.