CLOUD COMPUTING AND LOAD BALANCING IN CLOUD COMPUTING

Daisy Sharmah1 Dr. Kanak Chandra Bora2

Department of Computer Science & Electronics, Department of Computer Science & Electronics,

University of Science & Technology Meghalaya, University of Science & Technology Meghalaya, India India

email- sharmah.daisy@gmail.com email- boradrkanak@gmail.com

**ABSTRACT**

The Cloud Computing is a major area of research. A vast use of applications are emerging day by day in cloud computing. As the use of the cloud computing is increasing, the load of the cloud servers are also increasing gradually. As a result of this, the distribution of load in cloud servers is not in a balanced way. Therefore the concept and technique of load balancing is introduced to uniformly distribute the load amongst the cloud servers. In this chapter, we are discussing about the goals of load balancing, its aim towards the cloud computing. The challenges of load balancing are also discussed in this chapter. Along with that we have also discussed the various algorithms used in load balancing.

**Keywords**- Cloud Computing, Load Balancing, Cloud server, Virtual Machine

1. **INTRODUCTION**

The Cloud computing is an emerging technology. It is an on-demand access, via the internet, to computing resources like applications and servers, data storage, development tools, networking capabilities hosted at a remote data center managed by the cloud service providers.

The Cloud Computing is invented with four deployment models that are stated below [3] (Balaji K., 2021)-

* 1. **Public Model**: This model is suitable for the general public.
  2. **Private Model**: This model is available for multiple users.
  3. **Community Model**: This model is best suited for a specific level of community.
  4. **Hybrid Cloud**: This model is the combination of 3 or more cloud models.

Based on the delivery model, the cloud computing is divided into three models-

* 1. **Infrastructure as a Cloud (IaaS)**: This service supplies infrastructure oriented resources as a service of cloud computing. For example, Amazon Web Services (AWS), Cisco Meta-cloud etc.
  2. **Platform as a Service (PaaS)**: This model provides the inbuilt configuration as a service to the users. For example, Google App Engine, Windows Azure etc.
  3. **Software as a Service (SaaS)**: It provides a set of programs and data as a service to the users. For example, Cisco Webex, Facebook etc.

Load Balancing is a methodology which offers methods to maximise throughput, optimized utilisation of resources and better execution of the system. The Load Balancing technique creates a method to store the data for the users based upon its availability. The main objective of Load Balancing is to distribute the load in the entire cloud computing system. Load balancing in the cloud is the process of equivalently administering the functions on virtual machines for proper usefulness of all the hardware and software systems used. There are different types of algorithms used based on different metrics. The aim of load balancer is to help in resource assignment for resource uniformities and also to fulfil the need of the users in optimum price that stimulate the users to find issues in load balancing and to work on resolving them. As the demand of cloud computing is increasing, therefore the workload of this system is also affecting and as a result, the load balancing plays a very major role in cloud computing system. Therefore, on the basis of available research, the existing work has been identified and summarized in a systematic way that depicts issues and challenges for future research work [14] (Kumar P., 2019).

There are two versions of load balancing algorithms: static and dynamic. The static-based balancing algorithms are mostly fit for stable environments with homogeneous system. Dynamic-based balancing algorithms are more adaptable and effective in both homogeneous and heterogeneous environment [15] (Mishra S.R., 2020). The Distributed Load Imbalance System occurs when a number of users make a request to access to the same server while other servers are sitting idle. To overcome this situation, Distributed Load Balance System is introduced for better performance of resource utilisation and also can reduce the time limit of a task execution.

1. **EASE OF USE**
   * 1. **Literature Review/ Related Works**

Afzal et. al [1] (Afzal S., 2019) presented in this paper a detailed review about the load balancing techniques. The advantages and limitations of existing methods are highlighted by the authors with crucial challenges being addressed so as to develop efficient load balancing algorithms in future. The authors also suggested new insights towards load balancing in cloud computing. The problem of load unbalancing in cloud computing was discussed by the authors along with driving factors. However, the challenges of the load balancing algorithms are explored in this paper in order to suggest more efficient load balancing methods in future. Some of the essential QoS metrics are not discussed in this paper.

Ansar et. al. [2] (Ansar K., 2019) mentioned that load balancing is a challenge in cloud computing and therefore mentioned about implementing meta-heuristic techniques for load balancing in cloud computing. Job scheduling is implemented using K-Mean clustering algorithms based on unsupervised machine learning. The authors proposed a hybrid algorithm as a hybrid scheme based on K-Mean and harmony search algorithm (HSA) algorithm in this paper and finally, the results are shown for the efficiency of the proposed algorithm. However, the proposed algorithm is not implemented in real time environment as stated by the authors. They also suggested for other optimisations for future work of the algorithm.

Balaji et.al.[3] (Balaji K., 2021), illustrated in this paper a report of problems faced by load balancing. In accordance with this study, massive research has been done on several load balancing approaches taking into account the different measures. The authors differentiated the mentioned load balancing technologies to various categories like Natural Phenomenal, Hybrid, Agent, Task and general based load balancing. For individual section, the idea, pros, cons and issues with respect to each and every technology are exhibited.

In this paper, the authors [4] (Belkhouraf, 2015) described some characteristics of cloud computing, exploring some of its critical issues and challenges, namely availability, performance and security. An approach is proposed to take more advantage from load balancing by means of a semi centralized and multi cluster scheme. This promising method is very likely to result in a good overall performance and provide better fault tolerance; it can be deployed in different cloud environments, especially the most demanding in terms of resources. In our future work, the authors suggested for various simulations in order to have a deeper understanding of its possibilities and potential enhancements (Belkhouraf, 2015).

The authors Bhandari et. al. [5] (Bhandari A., 2019) designed a modified throttled load balancing algorithm to consider the virtual machine availability. The availability index is considered in every virtual machines for a given time period by the authors. They also compared their proposed algorithm with Round-Robin Algorithm, Throttled Load Balancing Algorithm and Active Monitoring Load Balancing Algorithm and the performances are evaluated using CloudAnalyst. However, the authors have challenges to execute their proposed algorithm in the real time situation which may vary the results.

The authors Chawla et. al. [6] (Chawla A., 2018) proposed an algorithm in this paper on Packet based Load Balancing Algorithm with the objective to replicate the packets in Virtual Machines during the unavailability of the requested packets by grouping of packages together. As suggested by the authors, the main aim of this algorithm is to balance the load in Cloud Computing Environment with minimum cost, execution time and waiting time. However, the authors have not tested the algorithm in real time situation; it was tested using cloud sim toolkit. They have also mentioned that their proposed algorithm works smoothly if the Virtual Machines work without any fault. The authors have suggested working on automatic creation of the migration tools for the future enhancement.

In this paper, the Authors Ebadifard et. al., [7] (Ebadifard F., 2020) proposed a task scheduling method by using the honeybee algorithm with the goal of load balancing. They suggested this algorithm reduces make span by balancing the existing tasks on virtual machines, but it also proposes a fault tolerance capability and thus, increases the system`s reliability. The proposed method was compared with the Round Robin approach and the method of load balancing of honeybee behaviour. Simulation results indicate that in comparison to the honey bee method, the proposed algorithm improves the average make span by 21.25% when increasing the number of tasks. Moreover, the proposed method also reduces the waiting time. In the future, the authors aim to apply this work to the workflows and consider other factors like cost reduction for the service provider [7] (Ebadifard F., 2020)

Ghom et al.[8] (Ghom E. J., Journal of Network and Computer Applications, 2017 Elsevier Ltd. All rights reserved., 2017) surveyed several metrics for load balancing techniques that considered in future load balancing mechanisms. Based on the observations, the authors presented a new classification of load balancing techniques: (1) Hadoop MapReduce load balancing category, (2) natural phenomenon based load balancing category, (3) agent-based load balancing category, and (4) general load balancing category. In each category, they studied some techniques and analyzed them in terms of some metrics and summarized the results in tables. Key ideas, main objectives, advantages, disadvantages, evaluation techniques, publication year were metrics that considered for load balancing techniques. The authors also focused on two critical metrics, That is, energy saving and reducing carbon dioxide emission. However, they suggested the followings: (1) Study and analyze more recent techniques in each of our proposed categories, (2) Evaluate each technique in a simulation toolkit and compare them based on new metrics.

In this paper, the authors Grover et. al [9] (Grover J., 2013) used Agent Based Dynamic Load Balancing (ABDLB) approach defined as an independent software program that runs on behalf of a network administrator. It has ability to learn and after comparing the proposed method with traditional load balancing scheme as suggested by the authors. However, the Agent Base load balancing scheme reduces the communication cost of servers, accelerates the rate of load balancing which indirectly improves the Throughput and Response Time of the cloud.

Gupta et.al. [10] (A., Load Balancing in Cloud Computing., 2017), in this paper discussed MSLB (Managed Server Load Balancing) techniques for load balancing. The author considered two data centres and 4 Virtual Machines for the research and for presenting result paradigm. The overall response time and the response time by various regions are illustrated by the author in this paper. The cost efficiency of the proposed system is also mentioned in this paper in a graphical representation.

The authors Kapoor et. al [12] (Kapoor C., 2015) proposed algorithm that adds clustering approach so as to divide VMs with similar capacities into groups. K-means clustering approach has been used to divide VMs into cluster. They discussed the issues have been addressed by proposing an algorithm Cluster based load balancing which works well in heterogeneous nodes environment, considers resource specific demands of the tasks and reduces scanning overhead by dividing the machines into clusters. However the experimental results have shown that the proposed approach in this paper by the authors gives better results than throttled and modified throttled algorithms when compared on the basis of waiting time, execution time, turnaround time and throughput.

The Authors Kaur et. al. [13] (Kaur A., 2019) discussed about the paired tree algorithm to improve the performance of the virtual machines based on their overall task computation time. They also mentioned the use of scheduling of workflow algorithms. As suggested by the authors, the paired tree mechanism is used to adjust the load of virtual machines dividing the conditions of the machines. The proposed model suggested by the authors, acts in active state whenever the server gets exhausted and divides the load using subordinate splits recording in the paired tree. However, the authors also mentioned that the proposed system can be enhanced with optimum consumer experience and also suggested to restructure the algorithm including QoS parameters.

The author Kumar P. Et. al. [14] (Kumar P., 2019) suggested many objectives of Load Balancing in Cloud Computing. However, the authors also suggested that there is a need for proper workload mapping and load balancing techniques that consider different metrics.

Mishra et.al. [15] (Mishra S.R., 2020), have described various load balancing techniques in homogeneous and heterogeneous cloud computing environments. Various performances and parameters are listed and evaluated the system performance. The calculation of make-span and energy consumption of the system is explained in details in this paper. A taxonomy is proposed by the authors for the load balancing algorithm in the cloud environment. To analyze the performance of heuristic-based algorithms, the simulation is carried out in CloudSim simulator and the results are presented in detail. However, the authors suggested to evaluate the proposed algorithms in a real-world cloud deployment.

The authors, Mukundha et. al. [16] (Mukundha C., 2017) , surveyed multiple algorithms for load balancing in Cloud Computing and Genetic algorithm for Load Balancing in Cloud Computing is proposed in this paper. The authors described in this paper the genetic algorithm not only balance the load but also help in efficient utilization of resources, increase overall throughput and decrease the response time. All these will reduce the operational cost and will attract more users towards cloud computing as suggested by them. However, the challenges of these algorithms are addressed so that more efficient load balancing techniques can be developed in future.

In the present paper, the authors Panwar et. al. [17] (Panwar R., 2015) proposed a dynamic load management algorithm for distribution of the entire incoming request among the virtual machines effectively. The performance is simulated by using CloudAnalyst simulator by the authors based on various parameters like data processing time and response time followed by comparing the result with previous designed algorithm VM Assign.

In this paper, the authors Rahman et. al.[18] (Rahman M., 2014) mentioned about the importance of load balancing and desired characteristics in cloud. They also provided complete review on the existing load balancing strategies, their strength, shortcomings and a comparative study. The authors also discussed algorithms, approaches and strategies based on research conducted and analyzed. and presented load balancer as a service model adopted by the major market players. However, the technologies involved are still in their infancy and need to mature for the companies to believe and trust their business with cloud computing as per authors suggestions.

The authors Ren et. al. [19] (Ren H., 2012) proposed in this paper a dynamic migration algorithm in cloud computing environment under considering the heterogeneity of environmental resources on cloud computing applications to trigger strategy based on the fractal methods. As mentioned by them, the strategy determines the timing of the virtual machine migration through forecasting the timing to determine the timing of the virtual machine migration, which can avoid the problem of the peak instantaneous load trigger. However, the authors also mentioned a problem that the instantaneous peak load will be triggered once the virtual machine migration. The trigger strategy proposed fractal achieved through the load forecast, when the load indicator exceeds the specified thresholds are not immediately trigger migration.

In this paper, the authors Sarma et. al. [20] (Sarma P., 2019) compared different load balancing algorithms including Round Robin Algorithm, Modified Throttle Algorithm, Shortest Job First Based Load Balancing Algorithm, Ant Colony Optimization (ACO), ACCLB, Honey Bee Behaviour inspired Load Balancing (HBB-LB), Particle Swarm Optimization (PSO). The simulation results of all the discussed algorithms are also shown in this paper. The authors come to a conclusion that on the basis of the results that amongst Round robin, Shortest Job First, Ant Colony Optimization and Honey Bee Load balancing algorithms for cloud environment, Honey Bee Load balancing algorithm shows significant improvement over the other algorithms. Future work is associated to designing a new dynamic load balancing algorithm for better resource utilization, fast throughput and minimum response time of the cloud computing environment.

The authors Sharma et. al. [21] (Sharma S.C.M., 2022) proposed two multi center two-phase adjustment techniques with the objective to make a better make-span by distributing the load in many Virtual Machine (VM)s. The authors discussed about the various features of the cloud milieu and focused on the bandwidth and delay of the resources rather than resource utilisation feature. The simulation of the proposed algorithm is performed using MATLAB and their results are stated by the authors. They however suggested extending their work by considering criteria on multiple task allocation features.

The authors Singh et. al. [22] (Singh H., 2018) discussed the various challenges of cloud computing include the reliability, some of the legal and compliant factors, security, ownership, performance, interoperability, multi platform support, data management issues and most importantly the load balancing. However, the shortage of tools for geographical distribution of servers and users is also mentioned in this paper.

Singh et. al. [23] (Singh H. G. R., 2014), made a comparison of average response time of active, throttled and proposed Virtual Machine load balancing algorithm with graphical representation in this paper. After the comparative study, the authors came to a conclusion that the whole system can reduce the response time, and not to increase the total cost if we select an efficient virtual machine, one more condition we have find that if we increase the data center it will effect the overall performance. The results are evaluated are based on existing round robin, throttled load balance and active monitoring load balancing algorithm and the experimental result concluded by the authors that if the number of data center is increased this leads to decrease in overall average response time. However, the authors suggested that the TLBA algorithm can be further improved so that cloudlets are allocated pertinent execution environments and the customer charged accordingly and the cloud provider can manage resources optimally based on the environment software cost incurred (Singh H. K. H., 2018).

Sui et. al. [25] (Sui X., 2019) proposed in this paper a virtual machine intelligent scheduling strategy based on machine learning algorithm to achieve load balancing of cloud data center. The experimental results are shown by the authors that compared with other classical algorithms, the proposed virtual machine scheduling strategy reduces the number of virtual machine migration by 94.5% and the energy consumption of cloud data center by 49.13%.

Tong et. al. [26] (Tong Z., 2021) proposed a novel DRL-based dynamic load balancing task scheduling algorithm under service-level agreement (SLA) constraints to reduce the load imbalance of virtual machines (VMs) and task rejection rate. The authors proposed the algorithm to get the best performance improving the overall level of cloud computing services. DRL-based dynamic multi objective task scheduling algorithm (DDMTS), and the deep Q-network (DQN) method is used by the authors to select a Virtual Machine to perform the task. They experimented and the results depicted by the authors that the DDMTS algorithm exhibits the best performance in covering the Degree of Imbalance (DI) and task rejection rate. As a result, the authors stated that using this algorithm, the overall quality of cloud computing services had been improved. However, they compared their proposed algorithm with the OLB algorithm and it was discovered that the DDMTS algorithm can be improved in the field of task rejection rate for further modifications.

Tripathi et. al [27] (Tripathi A., 2018) proposed a hybrid algorithm of ant colony optimization and bee colony algorithm. In this paper, the authors combined the parameters of the upgraded bee colony algorithm to form a new ACO method. The Cloud Analyst tool is used to simulate the results for the proposed system by the authors. The authors obtained a better performance of the hybrid algorithm after combining both bee colony algorithm with ant colony optimization algorithm.

Weinhardt et. al. [28] (Weinhardt C., 2009) discussed in this paper the challenges that have to be mastered in order to make the Cloud vision come true, such as the development of a Cloud API, the pricing of complex services, as well as security and reliability. The authors compared Cloud computing with Grid computing and focused on giving a clear distinction between Cloud Computing and Grid Computing by identifying a catalogue of criteria and comparing both paradigms. The authors also mentioned that Clouds require new business models, especially with respect to the licensing of software. They suggested in the paper that a commercial success can only be achieved by developing adequate pricing models that foster an efficient way to allocate and valuate composite services. However, the authors mentioned that the selling of few unique services requires a thorough understanding of portfolio management. Another crucial point for the eventual acceptance of Cloud technology in business industries will be the safety of critical data, both in transfer as in storage.

Wickremasinghe et. al. [29] (Wickremasinghe B., 2010) , mentioned the simulation settings and experimental results in this paper with the variation of average response time. As the results show that bringing the service closer to users improves the quality of service (response time in this case) by the authors. However, as per authors suggestion, the service quality can be increased with the application of load balancing across the data centers, which are supposed to be managed by different service brokerage policies, and also at virtual machine level within data centers.

* + 1. **Major Goals of Load Balancing**

1. To build up the efficiency and performance of the system.
2. To set up a fault tolerant system.
3. Provide support for the balancing of the system.
4. Making a better system using limited resources.
5. Provide better optimisation of user satisfactory level.
6. Reducing the time of waiting and job execution.
   * 1. **Aim of Load Balancing in Cloud Computing**
7. The load balancing provides constant service with suitable resource utilisation.
8. The load balancing also increases the performance of the system in optimum price.
9. It also provides the scalability service without interrupting the existing system.
10. The energy utilisation and carbon emission is reduced by using the load balancing technique as it balances all the resources with equal distribution of load.
11. Load balancing prepares the provision to issue the function uniformly on accessible materials.
12. The main aim of load balancing is to provide simultaneous assistance on non-successful condition of any part of facilities. It is done by equipping and dispossessing the request cases for fulfilment of materials.
13. Load balancing also performs at lower cost by reducing the latency of its scheduled jobs and improve the resource fulfilment.
14. Load balancing contribute in extensible and elasticity for the dynamic applications whose size may increase in future and requires more facilities.
15. Another objective of load balancing is to decrease energy utilisation and carbon emission[14] (Kumar P., 2019).
    * 1. **Challenges of Load Balancing**

Though Load balancing is an useful technique to balance the load of cloud servers, it still has challenges to face. Some of the challenges are mentioned in the following [10](A., Load Balancing in Cloud Computing., 2017)

* Throughput: Calculating the execution time of process by CPU.
* Overhead: Improved overhead at execution time.
* Fault tolerance: Less number of faults for better performance.
* Migration Time: Time taken by the processor to transfer one process.
* Response time: Time is consumed by the organisation to the reaction of process.
* Resource utilisation: Ability to efficiently utilise the resources.
* Scalability: Load Balancing on a Virtual Machine with multiple clients.
* Performance: Utilised to evaluate the carrying out of processor.
* Point of Failure: Load Balancing system must be designed in such a way that the entire system should not fail if one central node fails [16] (Mukundha C., 2017).
* Geographically Distributed Nodes: The data centers are distributed according to the geographical features of an area for computation purposes [3] (Balaji K., 2021).
  + 1. **Algorithms used in Load Balancing**

There are mainly three types of load balancing algorithms used in cloud computing and they are[\*\*](Singh H. G. R., 2014)**:**

1. Round Robin Algorithm
2. Active Monitoring Algorithm
3. Throttle Load Balancing Algorithm
4. **Round Robin Algorithm:**

The Authors, Shukla S. and Suryavanshi R. S. (2019) [24] (Shukla S., 2019) proposed the Round Robin algorithm used in load balancing. They also mentioned that there are two types of Round Robin algorithm used in Load Balancing and they are- i. Classic Round Robin Algorithm and ii. Weighted Round Robin Algorithm.

1. **Classic Round Robin Algorithm:**

The authors mentioned the steps of Classic Round Robin Algorithm which are illustrated below [24] (Shukla S., 2019):

Step 1: Distribute the client information across a group of servers.

Step 2: Multiple servers and multiple clients can be involved.

Step 3: The servers are identical and configured to provide the same services to the users.

Step 4: All are configured to use the same domain name with different IP Addresses.

Step 5: Load balancing has the list of all the IP Addresses with associated Internet Domain Names.

Step 6: When the request for the sessions is linked with the Domain Names, they are allocated in a circular sequential form.

Step 7: The first server is allocated the request as the client initiates the request and passes to the server.

Step 8: The same process continues to the second and the third servers.

Step 9: In case of the fourth request, it is allocated to the first server again creating a ring of allocating resources.

1. **Weighted Round Robin Algorithm:**

This algorithm has the following steps:

Step 1: The network administrator assigns a fixed number of weights to each server in the pool.

Step 2: The most efficient and powerful server can be assigned as Weight= 100.

Step 3: The servers with bigger number of weight can be assigned with more requests.

Step 4: The ring is formed based on the requests.

Step 5: It is considered with the system that the higher weightings being assigned with more requests in each cycle.

1. **Active Monitoring Algorithm:**

This dynamic algorithm follows the steps as given below [11] (V., 2019):

Step 1: An index table is maintained by the controllers.

Step 2: The servers are identified based on their loading in current time.

Step 3: If any server is detected with minimum load or in idle state, then the load is allocated to that server.

Step 4: The index table is updated when the load is allocated to a server.

Step 5: The First Come First Serve (FCFS) technique is used to allocate the load.

Step 6: The unique server id is used t assign the tasks to every server.

Step 7: Every task completion allows the controllers to update the index table once again.

Step 8: Based on the request of the user on the Internet usage, the load balancer allocates the load to particular server by checking the index table.

1. **Throttle Load Balancing Algorithm:**

The Throttled Load Balancing Algorithm is dependent on Virtual Machine status. The Virtual Machine allocation is in binary form as yes or no and it defines the status of it. The status information is stored in an index table at the load balancer. The index table consisting of two parameters- first is the ID of the virtual machine and the second is the status whether the virtual machine is available or busy. Initially, all the virtual machine status are set to ‘Available’.

Client Request

Received by

Receives -1, if all VMs Status=busy

Search/ Update

Data Centre

Load Balancer

Forward

Request

|  |  |
| --- | --- |
| **VM\_ID**  Update Request | **STATUS** |
| 1 | Busy |
| 2  If Successful | Available |
| 3 | Busy |
| 4  Yes  Set Status=Busy | Busy |

No Update

Index Table

No

-ve Feedback Notification within the time period

**Fig-1: Working Principle of Throttled Load Balancing Algorithm**

1. **CONCLUSION**

In this chapter we came to know about the cloud computing and the importance of load balancing in cloud computing. The challenges of load balancing are also mentioned here. The main three types of load balancing algorithms are discussed in this chapter. However, there are many hybrid algorithms combing the mentioned three algorithms, used for balancing the load which are not discussed in this chapter. There are also various pros and cons of the mentioned three algorithms. Those are kept for the future discussion.

**REFERENCES**

1. Afzal S., Kavitha G. (2019), Load balancing in cloud computing – A hierarchical taxonomical classification, *Journal of Cloud Computing: Advances, Systems and Applications, Springer Open*, 1-24,

doi- [https://doi.org/10.1186/s13677- 019-0146-7](https://doi.org/10.1186/s13677-%09019-0146-7)

1. Ansar K., Javaid N. , Zahid M., Tehreem K., Bano H., Waheed M. (2019),A Hybrid HS-Mean Technique for Efficient Load Balancing in Cloud Computing, *Springer Nature Switzerland AG 2019*, 40–48,

doi- <https://doi.org/10.1007/978-3-030-02613-4_4>

1. Balaji K., Kiran P.S., Kumar M.S. (2021), Load balancing in Cloud Computing: Issues and Challenges, *Turkish Journal of Computer and Mathematics Education,*12(2), 3224 – 3231.
2. Belkhouraf , Kartit A. , Ouahmane H., Idrissi H. K., Kartit Z. , Marraki M. E. (2015), A secured load balancing architecture for cloud computing based on multiple clusters, *2015 IEEE*
3. Bhandari A., Kaur K. (2019), An Enhanced Post-migration Algorithm for Dynamic Load Balancing in Cloud Computing Environment. Springer, Singapore , Advances in Intelligent Systems and Computing, 811, 59-73,

<https://doi.org/10.1007/978-981-13-1544-2_6>

1. Chawla A., Ghumman N.S. (2018), Package-Based Approach for Load Balancing in Cloud Computing. *Big Data Analytics. Advances in Intelligent Systems and Computing*, *Springer, Singapore*, 654, 71-77

Doi-<https://doi.org/10.1007/978-981-10-6620-7_9>

1. Ebadifard F., Babamir S. M., Barani S. (2020), A Dynamic Task Scheduling Algorithm Improved by Load Balancing in Cloud Computing, *2020 6th International Conference on Web Research (ICWR), 2020 IEEE Explore, 177- 183.*
2. Ghom E. J. , Rahman A. M., Qaderb N. N. (2017), Load-balancing algorithms in cloud computing: A survey, *Journal of Network and Computer Applications, 2017 Elsevier Ltd. All rights reserved*., 50-71.

doi- <http://dx.doi.org/10.1016/j.jnca.2017.04.007>

1. Grover J., Katiyar S. (2013), Agent based dynamic load balancing in Cloud Computing, *IEEE 2013 International Conference on Human Computer Interactions,Chennai, India,* 1–6.

doi:10.1109/ICHCI-IEEE.2013.6887799

1. Gupta A. (2017), Load Balancing in Cloud Computing. *International Journal of Distributed & Cloud Computing,* 5(2).
2. Joshi V. (2019), Load Balancing Algorithm in Cloud Computing, *International Journal of Research in Engineering and Innovation*, 3(6), 75-84.
3. Kapoor S. Dabas C. (2015), Cluster based load balancing in cloud computing*, IEEE 2015 Eighth International Conference on Contemporary Computing , Noida, India 2015 Eighth International Conference on Contemporary Computing ,* 76– 81, doi:10.1109/IC3.2015.7346656
4. Kaur A., Kaur B. & Singh D. (2019), Meta-heuristic based framework for workflow load balancing in cloud environment. *International Journal of Information Technology, Springer,* 11**,**119–125.

Doi- <https://doi.org/10.1007/s41870-018-0231-z>

1. Kumar P., Kumar R. (2019), Issues and Challenges of Load Balancing Techniques in Cloud Computing: A Survey, *ACM Computing Surveys*, 51(6), 120:1-120:35.
2. Mishra S.R., Sahoo B., Parida P.P. (2020), Load balancing in cloud computing: A big picture, *Journal of King Saud University – Computer and Information Science, ScienceDirect,* 32, 149-158.

DOI- <https://doi.org/10.1016/j.jksuci.2018.01.003>

1. Mukundha C., Ventakesh N., Akshay K. (2017), A Comprehensive study Report on Load Balancing Techniques in Cloud Computing, *International Jornal of Engineering Research & Development,* 13 (9), 35-42.
2. Panwar R., Mallick B. (2015), Load balancing in cloud computing using dynamic load management algorithm *, IEEE 2015 International Conference on Green Computing and Internet of Things, Greater Noida, Delhi, India, 2015 International Conference on Green Computing and Internet of Things,* 773-778*.*
3. Rahman M., Iqbal S., Gao J. (2014), Load Balancer as a Service in Cloud Computing*, IEEE 8th International Symposium on Service Oriented System Engineering- Oxford, United Kingdom, 2014 IEEE 8th International Symposium on Service Oriented System Engineering ,* 204–211*.*doi:10.1109/SOSE.2014.31.
4. Ren H., Lan Y., Yin C. (2012),  The load balancing algorithm in cloud computing environment*, IEEE 2012 2nd International Conference on Computer Science and Network Technology , Changchun, China , Proceedings of 2012 2nd International Conference on Computer Science and Network Technology,* 925–928*.*
5. Sarma P., Kalita C., & Deka V. (2019). A Survey on Load Balancing Algorithms in Cloud Computing, *International Journal of Computer Sciences and Engineering,* 169-176.
6. Sharma S.C.M., Rath A.K. & Parida B.R. (2022), Efficient load balancing techniques for multi-datacenter cloud milieu. *International Journal of Information Technology, Springer,* 14, 979–989.

Doi-<https://doi.org/10.1007/s41870-020-00529-2>

1. Singh H., Kaur H. (2018), Optimised Environment Allocation for static and Dynamic Tasks based on Throttle Algorithm in Cloud, *International Journal of Engineering Sciences & Research Technology,* 7(7), 1-11.

DOI: 10.5281/zenodo.1305803

1. Singh H., Gangwar R. C. (2014) , Comparative Study of Load Balancing Algorithms in Cloud Environment, *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(10), 3195 – 3199.

doi-<https://doi.org/10.17762/ijritcc.v2i10.3371>.

1. Shukla S., & Suryavanshi R.S. (2019), Survey on Load Balancing Techniques, *International Conferenence of Emerging Trends In Technology and Application*.
2. Sui X., Liu D. , Li L., Wang H. , Yang H. (2019),Virtual machine scheduling strategy based on machine learning algorithms for load balancing, *EURASIP Journal on Wireless Communications and Networking,* *Springer Open,* 1-16.

doi- <https://doi.org/10.1186/s13638-019-1454-9>

1. Tong Z., Deng X., Chen H., Mei J. (2021),DDMTS: A novel dynamic load balancing scheduling scheme under SLA constraints in cloud computing, *Journal of Parallel and Distributed Computing, Contents lists available at ScienceDirect*, 149, 138–148, doi- <https://doi.org/10.1016/j.jpdc.2020.11.007>
2. Tripathi A., Shukla S., Arora D. (2018),A Hybrid Optimization Approach for Load Balancing in Cloud Computing, *Springer Nature Singapore Pte Ltd. 2018*, 197-206,

Doi-<https://doi.org/10.1007/978-981-10-3773-3_19>

1. Weinhardt C. , Anandasivam W. A. , Blau B., Borissov N., Meinl T., Michalk W. W., Stößer J. (2009), Cloud Computing – A Classification, Business Models, and Research Directions, *Business & Information Systems Engineering*, *BISE – State of The Art-Springer* , 391-399, doi- 10.1007/s12599-009-0071-2
2. Wickremasinghe B., Calheiros R.N., Buyya R. (2010), Project web - <http://www.cloudbus.org/cloudsim/>