IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10788

A Study of Load Balancing in the View of Task Scheduling in Cloud Computing

Maaz Afnan[1], Dr.Nagaraj G Cholli[2]

Student, Information Science and Engineering, R.V. College of Engineering Bengaluru,India[1]

Associate Professor, Information Science and Engineering, R.V. College of Engineering Bengaluru, India[2]

Abstract: Cloud computing field is a rising technology and despite much research has been carried out but still challenges exist in work related to load balancing in cloud related applications. Load balancing algorithms are viewed from cloud environments such as Static and Dynamic Environment. An efficient task scheduling is an important action in cloud computing and understanding Load balancing from this perspective is essential for designing an algorithm which should be efficient from already existing algorithms . Task Scheduling is an important factor in load balancing and scheduling tasks. The main goal of this paper is to study and understand load Balancing and the parameters that affect the performance of an algorithm, analyze it on the basis of environment and performance in the view of task scheduling by considering task parameters such as execution time, Makespan, resource allocation.This paper also highlights the need of performance metrics from the view of Task scheduling and how it can increase the overall performance of the load balancing algorithm and also be economical at the same time.

Keywords: Cloud computing,resource utilization, load balancing,task scheduling,Virtual machines(VMs),makespan.

I. INTRODUCTION

Cloud Computing technology is becoming an essential part of the business in day to day life due to its applicability in many services which will be offered through various platforms such as applications related to clouds, web browsers and many more. Cloud has become a network of huge infrastructure. Cloud computing can be considered as an infrastructure by which applications can deliver the end users with its services over the Internet and also provide hardware and software to use these services. The resources should be efficiently utilized and should be scalable. However Computing is done based on some certain criteria given by the Service Level Agreement(SLA) document. Infrastructure that is provided through Cloud is made available to the users on the basis of, how much processing power or cloud resources they need and pay them as per acquired.

Smart load balancing makes cloud computing more economical and enhances user satisfaction with smooth experience. Load balancing aims to distribute workload or requests across multiple computers. In load balancing techniques various algorithms are used to transfer or migrations of tasks over loaded machine to below loaded machine and must not affect current user's running tasks. There is an increase in competition between Cloud providers, each of these providers such as AWS from amazon, Google cloud, offer load balancing algorithms where the focus is on elasticity, distribution of workload and traffic management. These enterprises deal with large-scale applications where it has to process millions of incoming requests hence the system should be scalable and available to everybody with good performance.

Cloud computing is the future of many applications but has to answer many questions such as heating and cooling, Cloud providers have to handle an enormous amount of user requests and also provide the service requested, the main goal however is the services but one main problem is the load balancing, the user requests, task allocation which could be answered by designing an algorithm which not only balances load but also helps to utilize resources efficiently, migrate tasks if required and many more.The operations and maintenance is centralised in cloud computing and require highly skilled professionals and experts who can operate at the highest standards.Cloud computing can be either private(restricted access) or public or both(hybrid), obviously while dealing with public cloud the chances of data breach are higher comparatively. One of the goals of cloud computing is to store data and use intelligent models to provide some valuable insights and with the help of many machine learning algorithms and AI find many applications, where an agent can discover more insights so that it can perform more informed decisions.

II. LOAD BALANCING

Load balancing can be considered as a redistributing of tasks also known as load in a distributed system into single nodes and ensuring that no nodes are overloaded, under-loaded, idle or some nodes are under loaded which leaks to inefficient utilization of resources. Effective cloud load balancing may therefore assure continuation in operation with a heavy traffic

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10788

website and ensure system stability. Usually optimization technique in which task scheduling is done for load balancing is an NP hard optimization problem .It can also be considered as a technique where the load is divided among servers, so loads can be sent and received without minimum delay . In load balancing a proper load distribution is done via a load balancer which is required for the incoming tasks or requests from different locations such that congestion does not occur at data centres. If load balancing is done by following the SLA document, resources utilized optimally then minimal resource consumption can be achieved.

Load balancing is the main focus in many cloud computing applications such as cloud gaming, cloud based services which face a lot of challenges while delivering the data, factors such as delay , performance and accuracy are very important to deliver their services. Load balancing should also consider environmental aspects such energy conservation and c02 emission. In a dynamic environment[14], which is mostly preferred over static, VMs are dynamically clustered depending on their sizes, so that the incoming job requests are mapped according to the needs of processing speed; this avoids the overload of tasks to a great extent. In a task scheduling scheme the algorithm has to assign deadlines to the tasks[16], hence the goal here is to minimize the deadline miss as much as possible which mainly takes place while processing large media

III. LOAD BALANCING ON THE BASIS OF CLOUD ENVIRONMENT

Cloud computing can be divided depending upon various factors such as node distribution or environment, if environment is considered then cloud computing can be either static or dynamic depending on the configurations of the cloud as needed for the application in use.

1.Static Environment : An Environment where depending on the needs the cloud provider installs homogeneous resources in the cloud which are not flexible and hence cannot be modified when required which makes the environment static. In this environment a prior knowledge is required such as node capability, the maximum power the nodes can withstand along with various other parameters. Algorithms proposed for static environments cannot be adapted if the load is required to change in runtime, however implementing load balancing here is comparatively easy to simulate but fails in heterogeneous environments and for resources allocated during run time. Few algorithms like Round Robin provide load balancing in this type of environment as decided by the cloud providers. Usually in a static based algorithm resources are allotted as they arrive in a queue and scheduled in a time sharing manner. The nodes with the minimum number of connection requests are addressed first and tasks are allocated accordingly.

2. Dynamic Environment: Dynamic environment is when a cloud provider wants to use heterogeneous resources which are also flexible. In this environment we take into account the run-time statistics which increases its flexibility. Hence run time changes can be easily adopted by the algorithms that work in this environment. Its difficulty in simulation is overshadowed by its high adaptability with the cloud computing environment. Algorithms like Weighted Least Connection(WLC) in a dynamic environment are allocated on the basis of least weighted task by taking into account node capacity which reduces the task starvation and task is assigned to a node. Dynamic Algorithms are highly scalable and hence are a better choice than static.

IV.RELATED WORK

A load balancing algorithm was proposed[12], for the data centres where frameworks were provided for both top-layer and bottom layer such that the proposed load balancing algorithm can be optimized for various cloud computing applications, this paper also discusses the importance of task Scheduling in load balancing and VM priority while task migration. In [1] , the algorithm was proposed on GA with the aim to solve the load balancing situation for virtual machines by integrating GA and gravitational emulation local search (gel).And in [2], the algorithm is mainly used for comparison and balance based on sampling so that it reaches an equilibrium solution and also decreases its migration time of Vms by shared storage. In [3] , the algorithm was aimed for single workflows so that it is divided and for each division called as partition a deadline is assigned the algorithm is also designed to work on multiple workflows which focus on execution time.In [4] , in this paper Task scheduling was addressed which is a significant goal of load balancing, as the clients increase the cloud can lead to improper task/jobs scheduling, Table 1 shows the comparison of different methods.

Table 1 COMPARISON OF RELATED WORKS

Algorithm	Objective	Issue
Genetic Algorithm based on double fitness adaptiveness	Designing a greedy approach based algorithm with a Double fitness function to increase the efficiency of Load	Performance could have been increased with a better Resource Utilization

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10788

	balancing by optimizing Task Scheduling	approach.
Particle swarm Optimization based algorithm which improved existing PSO by utilizing a complex network model.	Task Scheduling was addressed in terms of between VMs from overloaded to underloaded.	Improve Performance and Resource utilization
Ant Colony optimization algorithm	Load balancing based on Task scheduling to minimize makespan	Minimize makespan
Bee Colony Optimization based algorithm	VM Migration when overloading of requests at the data centre.	Reduce parameters such as makespan , migration time.
Compare and balancing algorithm	To reduce migration time while assigning tasks to Vms while overloaded.	Increase the response time and decrease migration time.

In [5], parameters that affect the resource utilization in an effective way without violating the SLA and additionally considering constraints such as Deadline, priority etc. Randles et al.[6] proposed load balancing by adopting a technique inspired by the movement of ants in search of food(Honeybee foraging) for a distributed load system by using a queue data structure for its implementation. Another technique for distributed systems by using virtual graphs as a knowledge base was proposed known as biased random sampling[7].In [8], the paper was a survey on how cloud computing could face challenges in computing requirements, also considering factors such as optimality to handle issues in a cloud environment. The paper also discusses the sustainability of swarm based optimization with their applications in a cloud environment and in a specific area. In [9], this paper talks about the challenges faced by cloud nodes while balancing its workload specific to the cloud environment.The paper has also classified load balancing techniques into categories. The techniques are analyzed in terms of metrics by considering the critical metrics. Load balancing in terms of Energy emission by considering the environment aspects such as carbon dioxide emission.In [10], this paper has done a comparative study of various algorithms in various environments and also has discussed pros and cons of all schemes.Static and Dynamic load balancing are analyzed and its implementation at level node with its effectiveness.In [11], this paper has discussed the challenges faced due to data redundancy and also the network congestion which causes response delay. The process of routing is discussed that are responsible for the delay. In [15], a paper was proposed for a heterogeneous based cloud which uses an adaptive based approach for its task allocation scheduler and it resulted in performance increase. The paper also discussed the methods to determine factors that affect the system such as the response time and also the backup tasks, and also solve them so that the system can effectively respond to its ability.

V. LOAD BALANCING CHALLENGES

Much research has been done on Load Balancing and many algorithms have been proposed but still many gaps are to be filled in terms of performance and identifying different parameters that may lead to optimize an algorithm. Below are some of the most common factors that need to be dealt with that can help a load balancing algorithm reach its goal.

1. Throughput: It's the time given to the processor to execute a process typically it should be increased for good performance, research is still going on to increase throughput
.
2. Migration time:It's the total time taken by a processor to migrate one process from one VM to another(by priority of VMs). Minimizing this time is a challenging task.

3. Response time:Any application that works on cloud needs to have a good response time. As the load and complexity increases there is a need to handle the response time in an efficient manner.

4. Resource utilization:Any machine should be able to utilize maximum resources for smooth functioning.

5. Scalability: Load balancing provides scalability by which network performance and its growth can be maintained. Due to growth nodes required increases but performance is compromised as VM's perform load balancing for the increased nodes hence this factor should be increased.

6. Performance: An algorithm in both static and dynamic environments is aimed to achieve high performance, however sometimes an algorithm can alone not increase the performance in cloud centres, required frameworks such that the algorithm can provide better results are to be designed .A framework includes a Top-layer(to handle incoming client requests), a Bottom-layer(physical allocation of tasks to the VMs).

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10788

VI. TASK SCHEDULING

A major aspect of Load Balancing is Task Scheduling which has to be addressed while dealing with load balancing. Task Scheduling can be defined as a mode where resources are allocated for end users. In cloud computing there are massive amounts of tasks which cannot be allocated manually .When the number of cloud users increases problems leading to improper scheduling hence the need for Task Scheduling. So the main goal of Task Scheduling is to execute the tasks efficiently such that maximum utilization of resources takes place and enables Multiprogramming where cloud can handle different processes at the same time with minimum delay.

Task Scheduling for a cloud computing environment is done in two ways namely, spaced shared and time shared. The major difference between them is that in space sharing mode the task will not be completed unless the resource allocation is done, whereas in time sharing mode tasks undergo completion while the resources are continuously preempted at the other end. Cloud providers are at a competition to provide high quality services, this occurs majorly because of unbalanced load scheduling there are many reasons one can be due to high Makespan time which can violate the SLA agreement[17], thus a task can be rejected due to starvation in a overloaded system.

The importance of Task Scheduling can be depicted using the below figure, a cloud broker acts as the intermediate between end users and the virtual infrastructure and is the main focus needed while designing an algorithm. When an algorithm schedules tasks to VMs parameters which are essential for the efficient scheduling of jobs are to be considered and made sure it meets the specific requirements. The cloud broker's efficiency depends on the user's request sent via internet to the VMs and it's execution/completion maintained within a specific deadline.

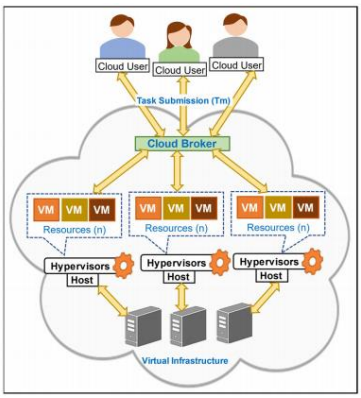

Fig 1

The above illustration in Fig 1 is for Task Scheduling in IaaS cloud computing[12], the hypervisors are responsible for the execution of multiple VMs located on a hardware. An example is VMware for the types of hypervisors that are located in a host machine.

A. PERFORMANCE OF LOAD BALANCING ALGORITHMs

The performance of an algorithm is most affected while the system is subjected to millions of incoming client requests and maintaining a balanced workload in the environment. It is a challenging situation as Vms need to be allocated properly such that it does not get subjected to under-loading , overloading or sitting idle . The main area to focus in an algorithm is priority assignment to the VMs, which directly affects the workload balancing. Hence there is a need to consider Task Scheduling parameters .However many papers have proposed several new approaches for load balancing but these algorithms focus only on a few parameters of Task Scheduling and have not yet addressed the workload migration fully. When an algorithm is designed it may take inputs such as the length of the task it has to schedule, the deadline associated with the task without violating the SLA document, however the goal remains the same that is to manage the incoming loads such that congestion does not occur in other terms workload balancing of VMs in a cloud environment. When an incoming task has a higher completion time than of assigned the algorithm should migrate the load/task to another VM to avoid traffic delay. The proposed architecture should be simple in terms of its implementation and operation at nodes, however the algorithm can be complex when we consider various performance parameters which makes the whole algorithm more complex. Some other factors which might hinder its performance are CPU load, background processes and also race conditions. The scheduling of tasks and its resource utilization is a vital factor in determining the algorithm's performance. However, considering performance metrics and designing an algorithm which brings these metrics into ideal condition is a challenging task. Simulation of cloud related issues is rising due to its importance such as CloudSim which can effectively eliminate the expenses of an actual computing facility[13], the resources and entities required for

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 7, July 2021
DOI 10.17148/IJARCCE.2021.10788

load balancing can be virtually modeled in a cloud environment and makes it easy to compute the performance of an algorithm.

B. PERFORMANCE METRICS

A Load Balancing Algorithm can be analyzed by various parameters below are a few metrics that need to be addressed by the algorithms.

1. Makespan : The sum of total time taken to schedule a given task to all nodes is known as makespan. This is a prominent metric of many scheduling algorithms . This metric should be minimized so as to schedule tasks faster which might also have an impact on resource allocations.
2. Execution Time : After the task is allocated it utilizes the resource for execution the total time it takes is Execution Time. Performance of an algorithm can be improved by reducing this time .
3. Resource Utilization : This metric determines the utilization of the CPU and determines its efficiency. When a task is scheduled, resources allocated to the task should be utilized in an efficient way so that minimal resource wastage can be achieved.

VII. CONCLUSION

As from the above discussion it is evident that task scheduling is a crucial factor while designing load balancing algorithms hence improving the task parameters contribute highly to load balancing algorithms. The things that might affect the performance of the algorithms have been discussed which should be taken care of while designing a load balancing algorithm. A comparative study on the algorithms has been done which need performance improvements mainly with respect to migration time. Improving an algorithm can result in efficient utilization of cloud resources hence it can reduce the costs to great extent if we consider balancing at large scale. The algorithm should also consider SLA violations for the parameters used in the algorithm.

REFERENCES

[1]. Dam S, Mandal G, Dasgupta K, Dutta P. Genetic Algorithm And Gravitational Emulation Based Hybrid Load Balancing Strategy In Cloud Computing. In Computer, Communication, Control and Information Technology (C3it), 2015 Third International Conference on 2015 Feb 7 (Pp. 1-7). IEEE
[2]. Zhao Y, Huang W. Adaptive Distributed Load Balancing Algorithm Based On Live Migration Of Virtual Machines In Cloud. In Inc, Ims and Idc, 2009. Ncm'09. Fifth International Joint Conference On 2009 Aug 25 (Pp. 170-175). IEEE
[3]. u, M., Cui, L., Wang, H. & Bi, Y. (2009). A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing. 2009 IEEE International Symposium on Parallel and Distributed Processing with Applications 978-0-7695-3747-4/09 $25.00 © 2009 IEEE DOI 10.1109/ISPA.2009.95.
[4]. A. Aruna Irani, D. Manjula, and V. Sugumaran, ‘‘Task scheduling techniques in cloud computing: A literature survey,’’ Future Gener. Comput. Syst., vol. 91, pp. 407–415, Feb. 2019, doi: 10.1016/j.future.2018.09.014.
[5]. M. Kumar, S. C. Sharma, A. Goel, and S. P. Singh, ‘‘A comprehensive survey for scheduling techniques in cloud computing,’’ J. Netw. Comput. Appl., vol. 143, pp. 1–33, Oct. 2019, doi: 10.1016/j.jnca.2019.06.006
[6]. Randles, M., Bendiab, A. T. & Lamb, D. (2008). Cross layer dynamics in self-organising service oriented architectures. IWSOS, Lecture Notes in Computer Science, 5343, pp. 293-298, Springer
[7]. Randles, M., Lamb, D., Bendiab, A. T. (2010). A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing. 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops 978-0- 7695-4019-1/10 $26.00 © 2010 IEEE. DOI 10.1109/ WAINA.2010.85./
[8]. Akash Dave, Prof Bhargesh Patel, Prof. Gopi Bhatt.”Load Balancing In Cloud Computing Using Optimization Techniques: A Study” journal of network and computer Applications.
[9]. Einollah Jafarnejad Ghomi , Amir Masoud Rahmania, Nooruldeen Nasih Qader “Load-balancing algorithms in cloud computing : A survey” journal of network and computer Applications.
[10]. Mayanka Katyal, Atul Mishra “A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment”
[11]. Yuxin Liu , Zhiwen Zeng , Xiao Liu , Xiaoyu Zhu and Md Zakirul Alam Bhuiyan “A Novel Load Balancing and Low Response Delay Framework for Edge-Cloud Network Based on SDN”
[12]. Dalia Abdulkareem Shafiq , Noor Zaman Jhanjhi , Azween Abdullah, (Senior Member, Ieee), and Mohammed a. Alzain, “A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications”
[13]. A. V. Lakra and D. K. Yadav, ‘‘Multi-objective tasks scheduling algorithm for cloud computing throughput optimization,’’ Procedia Comput. Sci., vol. 48,. 2015
[14]. Komarasamy, D, Muthuswamy V, “A novel approach for dynamic load balancing with effective Bin packing and VM reconfiguration in cloud”. Indian J. Sci. Technology 2016
[15]. Yang, S.J., Chen, Y.R. “Design adaptive task allocation scheduler to improve MapReduce performance in heterogeneous clouds” J. Netw. Comput. Applications 2015 (ICMSAO).
[16]. Cloud and Services Computing, Bok, K., Hwang, J., Jongtae Lim, J., Kim, Y., Yoo, J., “An efficient MapReduce scheduling scheme for processing large multimedia data. Multimed. Tools Application” 2016.
[17]. S. Afzal and K. Ganesh,Jan. 2019, ‘‘A taxonomic classification of load balancing metrics: A systematic review,’’ in Proc. 33rd Indian Eng. Congr.,