Vinod Patidar et al. 2019, 7:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology

An Open Access Journal

A Survey on Different Techniques of Static and Dynamic Load Balancing

Ph.D Scholar Vinod Patidar
Dept. of Computer Science &
Engineering
AISECT University Bhopal, M.P., India

Dr. Shiv Shakti Shrivastava
Dept. of Computer Science &
Engineering
AISECT University Bhopal, M.P., India
shivshakti18@gmail.com

Abstract- As cloud computing have number of issues related to security, bandwidth efficiency, large information handling, load balancing, etc. Load balancing implies distribution of the workload to node or servers or assets with the goal that one can accomplish maximum utilization of resources, reduce execution time, increase system throughput and so on. This paper gives a concise study of cloud based service balancing methods proposed by different researchers. Here various features of service are detailed for the load balancing and planning. Different sorts of requirement of load balancing was additionally talked with their significance and limitations. So according to the module as well as steps used in techniques classification of load balancing algorithms are done with a comprehensive and comparative understanding of existing literature.

Keywords- Cloud Computing, Load balancing, Machine Learning, Soft Computing, Virtual machines.

I. INTRODUCTION

Cloud computing is a current innovation that worry with online distribution of computing assets and services. In cloud computing, end-client do not worry about the availability of service because end user just pay for the particular model while requirement of hardware and software need to be fulfilled by service provider [1]. One imperative issue related with this field is dynamic load balancing or task planning. Load balancing calculations were examined vigorously in different conditions.

So one load balancing algorithm or system is required as shown in fig. 1. The primary concerns of cloud computing include effectively assigning tasks to the Cloud node with the end goal that the energy and demand handling is done as productively as could reasonably be expected, while having the capacity to tolerate the different influencing limitations, for example, heterogeneity, high correspondence delays, etc. [2].Cloud computing has risen as a popular solution in the business and research world, for its awesome potential to satisfy the clients which can appreciate this service by just submitting the required amount [3].

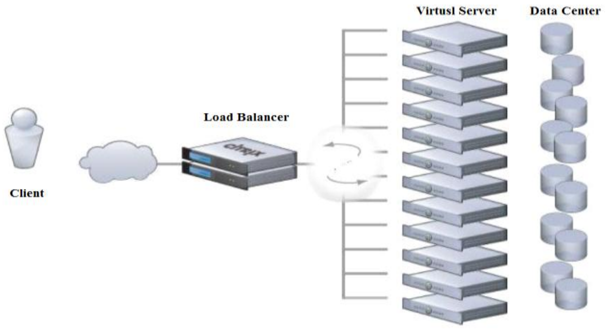

Figure 1: Load Balancing In Cloud Computing [3].

For the most part cloud give clients three levels of access: Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) and Infrastructure as-a-Service (IaaS) [4]. Cloud computing is proficient and expandable, yet keeping up the stability of handling such a significant number of jobs in the distributed computing condition is an exceptionally complex situation as load balancing receiving much consideration for researcher [5]. These days, computing frameworks intensely depend on Virtualization innovation and accordingly makes the servers practical for free applications. Further, virtualization process enhances the power productivity of the data enters (solidification of servers) and subsequently empowering the task of numerous virtual machines (VMs) to a solitary physical server [6]. In cloud framework, virtualization assumes a vital part by permitting web based sharing of computing assets [1]. Cloud utilizes virtualization innovation in conveyed server farms to allot assets for clients as the need them [4].

The rest of this paper is organized as follows: in the second section, the types of load balancing were discussed which was broadly classified as static and dynamic load balancing. Third section list various techniques proposed to handle this problem. While fourth section provide related work of the current approaches achieving load balancing of cloud data centre's will be introduced briefly. Research problem is pointed out, and then the proposed problem is formalized in detail. The conclusion of the whole paper is made in the sixth section.

II. TYPES OF LOAD BALANCING

Load balancing is one of the basic problems present in the cloud computation. This load may be of any type like requirement of memory, CPU limit, system or waiting state programs load. It is constantly important to share work load among the diverse node of the dispersed framework to enhance the asset usage and for enhanced execution of the framework. This can evade the condition where node are either intensely loaded or under loaded in the system. Load balancing is the way toward guaranteeing the similarly sharing of work load on the gathering of framework node or processor so that without aggravating, the running task is done. Load balancing algorithms are broadly classified into two categories shown in fig. 2 first was static and other was dynamic:

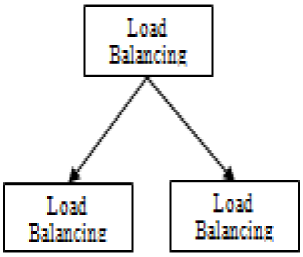

Figure 2: Types of load balancing.

1. Static Algorithm

Here as per the prior knowledge of the whole framework each task is assigned to respected node where changes in the current queue after assignment is not allow. The choice of shifting load does not rely upon current condition of framework. So number of situation arise when static algorithm have dump whole system like sudden failure of required asset, additionally the task is appointed to processors or machines simply after it is made and that assignment can't be moved to other machine in the course of its execution for adjusting the load.

These calculations are appropriate for homogeneous situations. These calculations are non pre-emptive so each machine has only one task for execution and that task is looking for any resource than whole waiting state is start [7]. Round Robin Load Balancing Algorithm In this calculation, settled quantum time is given to the activity. It allots job or task to all node in a round manner. Processors are appointed in a roundabout request and henceforth there is no starvation [7]. This calculation gives quicker reaction on account of equivalent workload conveyance among forms. However, a few node might be over loaded while others stay sit idle and under-used [8].

MIN-MIN Load Balancing Algorithm A series of assignment is kept up and least fulfillment time is computed for all the accessible node. An task with least execution time is allocated to the machine. So the name of the calculation is min-min load balancing algorithm [4]. Refresh the series and running time of the machine. It gives great outcomes when number of low execution task is high [7].

MIN-MAX Load Balancing Algorithm A series of assignment is kept up and least fulfillment time is figured for all the accessible node. An task with most extreme execution time is moved out to the machine. Thus the name of the calculation is min-max [4].

Update the series of job and running time of the machine help in executing other task [7].

1.1 Merit: - It is simple to be implemented by using a circular queuing or sequential array process. It does not require an inter process communication. It provides the best performance only for a special purpose application.

1.2 Demerit: It cannot provide good results in the general case, and when jobs are unequal. It assumes that all servers are similar in their storages, and in their response times. It does not give the priority for more important required tasks.

2. Dynamic Algorithm

It defeats the disadvantages of static approach as choice of adjusting the load relies upon current condition of framework. Any earlier learning of framework isn't required. It enables a procedure to move from an overused machine to underutilized for quicker execution. This permits acquisition which isn't bolstered by static approach [2]. Their calculations are intricate yet they give better execution and adaptation to internal failure. These calculations are more adaptable than static calculations and can without much of a stretch embrace the change and give better outcomes so more reasonable for heterogeneous and dynamic condition moreover. Dynamic load balancing calculations can likewise be of conveyed and non dispersed nature.

2.1 Merit: It is useful in environments where servers are distributed across different logical networks. It always uses the recent load information.

2.2 Demerit: It has more complexity. It includes software on each server to feedback the information about the ongoing performance.

III. TECHNIQUES OF LOAD BALANCING

Honey Bee Behavior: The primary thought behind the Honeybee Foraging calculation is gotten from the conduct of bees characteristics. There are two sorts of bee's activities: discoverers and collectors. The discoverer bees initially go outside of the honeycomb and locate the nectar sources. Subsequent to finding the source, they come back to the honeycomb and complete a waggle dance showing the quality and amount of nectar accessible. At that point, collectors go outside and procure the nectar from those sources. Subsequent to gathering, they come back to apiary and complete a waggle dance. This move demonstrates how much nourishment is cleared out. This calculation requires that every node to keep up a different queue. This calculation requires on every node causes extra overhead, however need balancing centres with respect to reduces the measure of time an undertaking needs to look out for a queue of the VM. Hence, it lessens the reaction of time of VMs. In [10] this strategy was proposed as a calculation named honey bee behavior inspired load balancing (HBB-LB), which intends to accomplish well adjusted load over virtual machines for amplifying the throughput. The proposed calculation likewise balances the needs of assignments on the machines such that the measure of holding up time of the services in the queue is negligible.

This work has contrasted the proposed calculation and existing burden balancing and booking calculations. The trial comes about demonstrate that the calculation is compelling when contrasted and existing calculations. Our approach delineates that there is a huge change in normal execution time and diminishment in holding up time of services on queue. The weakness of this calculation is that, it doesn't demonstrate any critical change in throughput, because of the extra queue and the calculation overhead. On the off chance that this work proceeds with bee activities foraging calculation it ends up by over-burden that's why this work will utilize dynamic clustering calculation.

1. Active Clustering: Ruchika Aggarwal et. al. in [7] proposed dynamic Clustering which is a bunching based calculation which presents the idea of grouping in distributed computing. The execution of a calculation can be upgraded by making a group of nodes. Each bunch can be expected as a gathering. The rule behind dynamic bunching is to aggregate comparative nodes together and after that work on these clustering.

The way toward making a bunch spins around the idea of relational relative node. In this procedure, first node chooses a neighbor node called the intermediary node which is of an alternate kind. This relational relative node makes association with its neighbor which is of same sort as the underlying node. At last the matchmaker node gets separated. This procedure is taken after iteratively. The execution of the framework is improved with high accessibility of assets, accordingly expanding the Throughput. This expansion in throughput is because of the productive use of assets. A fast adaptive load balancing method: D. Zhang et al. [11] propose a versatile load balancing technique for different ways that makes it conceivable to accomplish high TCP execution in every way. In regular load balancing strategies, interface usage is the principle parameter to be adjusted among various ways that are set up between an entrance and departure node match. Be that as it may, when this work considers TCP-level execution, balancing activity as far as just connection use may not generally bring about balancing TCP execution on every way.

Our technique uses task flow insights, for example, the quantity of dynamic task flows in every way, which is anything but difficult to quantify, and can improve TCP execution. By adaptively leveling the normal transfer speed utilized per dynamic task flow in every way, which is figured by isolating the data rate to the path by the mean number of dynamic task flows, our strategy accomplishes reasonable and high TCP execution for every path. Not at all other different strategies, new nodes between entrance departures are not required to perform traffic controls or estimations other than typical bundle sending. This work depict a load balancing technique for adaptively leveling the normal transfer speed utilized per dynamic task flow on every path and demonstrate its adequacy under heterogeneous conditions through simulation examination.

IV. RELATED WORK

Jia Zhao et. al. in [2] paper present the idea of process-based load balancing in cloud computing condition, which is rather than the quick load balancing techniques in the present literary works. The advantages are numerous overlap: to decrease superfluous calculation many-sided quality and to build organization effectiveness while satisfying the service execution for clients. What is critical is that LB-BC has long haul potential advantages of advancing outer service capacities and asset usage from the cloud closes viewpoint. Present the model of Bayes hypothesis on ideal arrangement of tasks to work out the probabilities of ideal physical hosts. Propose the bunch based task arrangement calculation by joining with Bayes probabilities to acquire the ideal arrangement of hopeful physical hosts. Propose an information structure of Matrix to decide the last arrangement vector in every calculation cycle.

Ruchika Aggarwal et. al. in [7] paper proposes a productive load-adjusting cloud asset scheduler approach for versatile clients. The proposed approach expands neighborhood cloud benefit pools with open cloud to build the likelihood of meeting the service level assertions. The proposed issue is isolated by open cloud benefit planning and neighborhood cloud benefit scheduling. The framework status data is utilized as a part of the half and half portable cloud framework, for example, the inclinations of versatile applications, vitality, server load in cloud server farm to enhance asset use and nature of experience of versatile client. In this way, the framework status of half and half versatile cloud is observed constantly. The scientific model of the framework and advancement issue is given.

S. M. Lau et. al. in [12], Sau-Ming Lau et al. have coordinated the two techniques of over burden need and light load need. They have proposed a versatile load appropriation calculation to successfully decrease correspondence overhead of the load adjusting process. Using the covetous calculations can take care of the issue of load circulation. Be that as it may, a few calculations above can't meet voracious decision execution and the idea of ideal sub-structure in the meantime. So these load dispersion approaches frequently acquire the nearby ideal arrangements. What's more, the impact of taking care of the issue of load dissemination under certain unique conditions isn't perfect. Cloud server farm can't achieve load balancing of the whole system.

V. Shrivastava et. al. in [13], VivekShrivastava et al. have exhibited a way to deal with put the virtual machines with solid connection of uses seriously. Be that as it may, the load adjusting issue and cost overheads of cloud server farms weren't considered. It just centered on virtual machines service to reinforce the service of cloud server farms and enhance the performing proficiency of cloud server farms.

ByungChulTak et. al. in [14] First, the general asset utilization at the mutual service which was the total of the asset utilization for various remote elements whose characters are not noticeable to the common service was calculate. Second, regardless of whether such data ends up accessible, basic checking instruments can't convey precise breakthrough of asset utilization since sharing happens at sub instance level. This work contemplate inborn difficulties of performing asset management of shared asset.

Nguyen KhacChien et al. [15] has proposed a load adjusting calculation which is utilized to upgrade the execution of the cloud condition in light of the strategy for assessing the finish of service time. They have prevailing with regards to upgrading the service time and reaction time of the client.

S. Yakhchi et al. [6] talks about that the vitality utilization has turned into a noteworthy test in cloud computing foundations. They proposed a novel power mindful load balancing strategy, named ICAMMT to oversee control utilization in cloud computing server farms. this work have abused the Imperialism Competitive Algorithm (ICA) for identifying over used has and after that this work relocate one or a few virtual machines of these hosts to alternate hosts to diminish their use. At last, this work consider different has as underutilized have and on the off chance that it is conceivable, move the greater part of their VMs to alternate has and change them to the rest mode.

Algorithm	**Type of Algorithm**	**Application**	**Limitation**
Federation Load Balancing Algorithm [15]	Dynamic	VM can be grouped on the basis of requests they handle. Immediately after getting request from client, the group that can handle the client request is checked by the load balancer and the request is allocated to lightly loaded	Some resource get waste due to wrong clustering. Process need to wait as cluster reduced available resource.
Rate Adjustment Policy [16]	Centralized	Job arrives than scheduler makes a decision of where the job should be sent, whether it is to be sent to the queue of node i or to other nodes.	System will undergo in starvation when high rated jobs continue.
Central Queue Algorithm [17]	Centralized	Redistribute the workload equally among the set of processing servers.	System spends more time in managing the jobs rather executing it.
Biased Random Sampling [18]	Dynamic	Develop a virtual graph. The in-degree of a vertex represents the resources that are free at that node. A task is assigned only to a node having at least one free resource available.	Allotting resource by graph lead long execution time as due to checking of each vertex.

Load balancing is one of essential issue faced while utilizing cloud computing. This work managed the issue of over-burden node. The load must be appropriated equitably all through the node. Load balancing helps in ideal usage of assets, expanded execution with vitality utilization. In this survey discussion about different load balancing calculations and the issues related with them was detailed, which must be considered while planning new algorithm. Existing load balancing systems works for reducing the execution time of the overheads present for assigning task to appropriate nodes. It was obtained dynamic load balancing technique which uses some genetic approach are more appropriate as compared to static approaches. In future a flawless calculation is with great component blend is wanted by investigating new load balancing calculations which adjusts the load much better and furthermore helps in green processing.

REFERENCES

[1]. N. K. Chien, N. H. Son and H. D. Loc, "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing," ICACT, pp. 228-233, 2016.

[2]. Jia Zhao, Kun Yang, Xiaohui Wei, Yan Ding, Liang Hu, GaochaoXu. "A Heuristic Clustering-based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment". IEEE Transactions On Parallel And Distributed Systems, June 2014.

[3]. Lei Yu, Liuhua Chen, ZhipengCai, HaiyingShen, Yi Liang, Yi Pan. "Stochastic Load Balancing For Virtual Resource Management In Datacenters". IEEE Transactions On Cloud Computing, Vol. , No. , November 2014.

[4]. RenGao and Juebo Wu, "Dynamic Load Balancing Strategy for Cloud Computing with Ant Colony Optimization", Future Internet, 465-483, doi:10.3390/fi7040465, 2015.

[5]. S. Garg, R. Kumar and H. Chauhan, "Efficient Utilization of Virtual Machines in Cloud Computing using Synchronized Throttled Load Balancing," 1st International Conference on Next Generation Computing Technologies (NGCT-2015), pp. 77-80, 2015.

[6]. S. Yakhchi, S. Ghafari, M. Yakhchi, M. Fazeli and A. Patooghy, "ICA-MMT: A Load Balancing Method in Cloud Computing Environment," IEEE, 2015.

[7]. Ruchika Aggarwal, LatikaGupta . "Load Balancing In Cloud Computing". ISSN 2320–088X IMPACT FACTOR: 6.017 IJCSMC, Vol. 6, Issue. 6, pg.180 – 186, June 2017.

[8]. LI CHUNLIN, ZHOU MIN AND LUO YOULONG. "Efficient Load-Balancing Aware Cloud Resource Scheduling for Mobile User". Computer And Communications Networks And Systems The Computer Journal, 2017

[9]. B. Godfrey, K. Lakshminarayanan, S. Surana, R. Karp, I. Stoica, "Load balancing in dynamic structured P2P systems", in: INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, vol. 4, IEEE, pp. 2253–2262, 2004.

[10]. Dhinesh Babu L.D, P. Venkata Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing 13 2292–2303, 2013.

[11]. Ryoichi Kawahara, "An adaptive load balancing method for multiple paths using flow statistics", Teletraffic Science and Engineering, Volume 5, Issue null, Pages 301-310, 2003.

[12]. S. M. Lau, Q. Lu, and K. S. Leung, "Adaptive Load Distribution Algorithms for Heterogeneous Distributed Systems with Multiple Task Classes," Journal of Parallel and Distributed Computing, vol. 66, no. 2, pp. 163-180, 2006.

[13]. V. Shrivastava, P. Zerfos, K. W. Lee, H. Jamjoom, Y. H. Liu, and S. Banerjee, "Application-aware Virtual Machine Migration in Data Centers," Proc. IEEE INFOCOM, pp. 66-70, 2011.

[14]. Byung Chul Tak, Youngjin Kwon, and BhuvanUrgaonkar. "Resource Accounting of Shared IT Resources in Multi-Tenant Clouds". IEEE Transactions on Services Computing, 10.1109/TSC..2453980, 2015.

[15]. Seyedeh Leili Mirtaheri, SeyedArmanFatemi, LucioGrandinetti.. "Adaptive Load Balancing Dashboard in Dynamic Distributed Systems". The Authors 2017. This paper is published with open access at SuperFri.org.

[16]. N. K. Chien, N. H. Son and H. D. Loc, "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing," ICACT, pp. 228-233, 2016.

[17]. S. Kapoor and D. C. Dabas, "Cluster Based Load Balancing in Cloud Computing," IEEE, 2015.

[18]. R. Panwar and D. B. Mallick, "Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm," IEEE, pp. 773-778, 2015.

[19]. M. Belkhouraf, A. Kartit, H. Ouahmane, H. K. Idrissi,, Z. Kartit and M. . E. Marraki, "A secured load balancing architecture for cloud computing based on multiple clusters," IEEE, 2015.