WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Using a Hierarchical Taxonomy, Load Balancing in Cloud Computing

SaiKrishna.D[1], Venkataramana.Ch[2], Sandeep.T[3], Venkat Varma[4], Suryakantham[5], Nageswra Rao Moparthi[6]

[1,2,3,4,5,6]Department of CSE, KoneruLakshmaiah Education FoundationVaddeswaram, AP, India

DOI: https://doi.org/10.55248/gengpi.2023.4.4.12261230

ABSTRACT—

The performance and effectiveness of computer resources are decreased by the load unbalancing problem is a multi-variant, multi- constraint issue. Overloading and underloading have two undesirable effects, which are addressed by load balancing procedures as a remedy to load unbalancing situations. disregarding how crucial load balancing methods are for

To our knowledge, there isn't a comprehensive, extensive, systematic, or hierarchical classification of the many load balancing techniques. systems now in use. Moreover, neither the literature nor any studies have been done on the elements that lead to load unbalancing issues. An exhaustive overview of loadbalancing strategies is provided in this piece. Future load balancing algorithms must overcome significant obstacles while pointing out the advantages and disadvantages of existing techniques. New insights on load balancing arealso proposed in the research.

Key words: *Load balances, Cloud computing, Cloud Traffic, Taxonomy, Service Providers, Service Consumers.*

I. INTRODUCTION

By serving clients with a range of needs using The rapid development of communication technology was facilitated by computer resources, cloud computing, and an internet-based network technology.

Whenever a cloud computing model utilizes its resources, most effective way possible, it is considered efficient. This optimum use of the cloud's resources may be attained by implementing as well as preserving effective cloud resource management. Effective resource scheduling, allocation, and scalability techniques are used to manage resources. Customers are given access to these resources in the form ofVirtual Machines (VM) via the virtualization process,which employs a thing (Hardware, Software, or Both), is referred to as a hypervisor. The transformation of a single-user physical system into a multi-user virtual computer is the main advantage of cloud computing. One example of a cloud service provider is challenging task given the available virtual resource and plays a significant role in service delivery to consumers Because to the unbalanced machines' huge user gradient, the cloud service provider is left with them.

The issue of load unbalancing is an unfavourable occurrence on the CSP end, which reduces the computing resources' efficiency and performance as well as the guaranteed quality of service under the established Service Level Agreement between the client and the provider. these situations, load balancing (LB), a particular study issue of interest to academics, becomes necessary. In cloud computing, load balancing can take place at the VM or physical machine level.

When several As jobs are added to a VM, the resources get saturated, indicating no further tasks can be added.

In a distributed system like cloud computing, load balancing is the process of shifting workload to make sure that no computer device is overloaded, under loaded, or idle.

It aims to accelerate many constrained criteria, To improve cloud performance, consider factors like reaction time, execution time, system stability, etc. An NP-hard problem in this optimization method is task scheduling. The majority of the focus has been on resource scheduling, resource allocation, task scheduling, and job

In the many load balancing systems that researchers have created, allocation and resource management are included. As far as we are aware, there isn't thorough and in-depth literature on the causes that lead to load unbalancing situations. Papers from the survey depending on load.

This study makes the following contributions, in brief:

1.Examine the variables that lead to load unbalancingCloud computing has an issue.

II. Provide a thorough rundown of the load balancing techniques now in use and how they have been applied to cloud computing.

3. Describe each load balancing methodology, method, strategy, and algorithm in detail.

4. Examine the difficulties researchers experienced when creating a successful load balancing algorithm.

II. LITERATURE WORK

In the realm of cloud computing generally, a lot of work has been done, especially in scheduling (tasks, VMs).

energy management, load balancing, resource provisioning, resource management, and computing). Nonetheless, load balancing has garnered attention due to its significance in cloud computing between the stakeholders, i.e., Cloud Service Provider and Cloud Service Consumer. the attention of academics. Based on a study of the review literature that has already been published, The inaccuracy of the categorization of the various procedures is one of the explanations cited. This section offers a thorough summary of the literature that has already been published.

A survey Ghomi et al. published a paper on load balancing strategies in cloud computing.. Task scheduling and load balancing methods were categorized by authors.

Application-oriented Hadoop-map reduce load balancing, general load balancing, agent-based load balancing, load balancing based on natural phenomena,

According to The literature recognizes two domains: workload-specific load balancing and network- aware load balancing. These are based on the process's current condition as well as the person who initiated it.

The many algorithms are gathered from each category, and both their benefits and drawbacks are presented. Based on the survey, Milani et al. assessed existing load balancing techniques and divided the algorithms into three broad categories: static, dynamic, and hybrid. The writers formalized pertinent queries concerning load balancing and addressed major issues with regard to itssignificance, the degree of metrics expected, its function, and the difficulties encountered in doing so.

In their categorization study, Kanakala et al. suggested categorization dividing up the many static and dynamic load balancing strategies currently in use those mentioned in earlier works. They noted difficulties in resolving the load balance issue as well. Geographical node distribution, migration times, system performance, energy management, and security are only a few of the difficulties that have long been documented in the literature. The authors really contrasted many Quality of service parameters, including throughput, speed, reaction time, migration time, etc., against different load balancing algorithms that are currently in use. The study's conclusion was that there are trade offs between measurements. Just eight load balancing techniques out of a large number of algorithms are compared in this research, which is a drawback.

A discussion of each algorithm's benefits and drawbacks was also conducted. The study doesn't approach the literature in a systematic manner.

The research of load balancing techniques in software defined networks by Neghabi et al. was well-organized, systematic, and encouraging. They separated the methods into deterministic and non-deterministic ones and thoroughly examined the related metrics. The paper outlines a few critical questions and makes an effort to shed light on them in terms of their significance, metric analysis, purpose, and challenges that software defined networks face. load balancing. The authors' analysis details the benefits and restrictions of the body of knowledge already written about communication networks. Notwithstanding the fact that there is no perfect load balancing metric, the study nonetheless maintains a firm basis and solid correlation among them.

According to the aforementioned survey papers, it can be said that there is a lack of quality in the available survey studies.

categorization process. While a criteria is established for categorization purposes, no generalization or specialization characteristics are identified, which ultimately results in poor and insufficient findings. Moreover, the present review articles do not look at certain crucial factors like The literature's use of load balancing metrics and the computational complexity of different load balancing techniques. The Quality of service measure set is not thoroughly described in the present survey articles, thus it's probable that more metrics—like migration costs, service level violations, degrees of balance, and task rejection ratios—should have been added. In this study, a taxonomy-based classification is suggested to show its superiority to the existing literature. The classification is still another.

III. IMPLEMENTATION

A good research methodology was used to delve deeply into the reasons of load unbalancing inorder to understand the core of the load balancing process. The literature review was carried out.

According to a broad research strategy that describes how the load-balancing The approach to the problem includes a list of the theories, techniques, algorithms, approaches, and paradigms that are used. The load unbalancing problem, which is broken down into sub-processes, or the components, variables, and parameters linked to load balancing, was studied using the constructive generic framework approach. Moreover, by following Kitchenham's suggested research standards for Systematic Literature Review and placing special attention on studies relating to load balancing techniques in the cloud, the literature review was enhanced.

Several criteria are used to categories In this section, load balancing algorithms.When classifying data, a top-down strategy is suggested and used. a constraint.

The main criticism load balancing methods are not effectively and substantially classified in a hierarchical taxonomic framework, which makes it difficult to judge the effectiveness of current review papers.the exact position of a certain algorithm within the taxonomy.

The many factors utilized for categorization include the "kind of load balancing," "state of the load balancing," "method utilised in load balancing," and "nature of the algorithm" are some examples of these terms. Now, for the first time in literature, a thorough analysis of the LB algorithms has been carried out. The algorithms for load balancing rely on the kind of approach being used.

Because of the nature of the method, load balancing algorithms were first categorized in this paper. LB algorithms base their decisions on this categorization.

This may be divided into proactive and reactive based techniques. Yet in other technical fields, particularly in the communication and networking for mobile adhoc networks (MANETS), the nature of the communication routing protocols has been carefully investigated.

An approach is a proactive-based LB algorithmic technique. of designing algorithms that considers action by instigating change rather than just responding to it when it occurs. Instead of waiting until there is a problem, it is designed to provide a positive result that avoids problems before they arise.

The reactive techniques react to a situation rather than trying to control it. The issue of load unbalancing is resolved by the reactive approach to load balancing.

when it occurs and its effects become apparent.Here is where the majority of load balancing algorithms fit.

The primary issue with load balancing that has been evaluated from the literature of previous research is that load unbalancing.

When the problem has occurred, researchers suggest several methods to address it by improving specific load balancing parameter.

Infrastructure as a service cloud scheduling and load balancing technique to decrease task completion time, make span, waiting time, and increase resource consumption. the reactive methods used in the current LB strategies. While proactive methods try to prevent problems before they arise and reactive methods offer solutions after problems arise, the former are more successfulthan the latter.

Algorithmic state

LB algorithms are frequently divided into static, dynamic, and hybrid categories dependent on the system's status information upon which they rely. The literature analysis makes it evident that this is the most well-liked categorization system for LB algorithms. The bulk of load balancing comparison studies begin by layering this category on top of the algorithmic taxonomy. The traffic load is uniformly distributed across the servers through static load balancing. This is carried out by an algorithm that is already aware of the system requirements and task requirements. The virtual machine's workload is planned at build time via the static LB algorithm. Although static algorithms offer the advantage of being simpler, they are fatally constrained by their inability to Moreover, dynamic algorithms are separated into two categories: batch mode, often known as offline mode, and are rather sophisticated algorithms. Dynamic load balancing can handle unexpected processing loads and takes into account the system's present status.

Functionality

Load balances are categorized according to their capabilities.

as hardware load balancing and elastic load balancing. Hardware load balances are concerned with the task distribution at the hardware level, which includes the CPU, memory, and storage. Elastic Load Balancing automatically divides up incoming application traffic using a variety of destinations, including Amazon EC2 instances, containers, and IP addresses. It may control varying loads of user application traffic in a single availability zone or across many.

Elastic load balancing offers three distinct forms of load balances, all of which meet the requirements for high availability, automatic scaling, and robust security for fault-tolerant user applications.The targets that the application load balance directs traffic to at the request level are EC2 instances, containers, and IP addresses.

IV. RESULTS DISCUSSION

In this part, the findings of a comparative analysis of numerous load balancing techniques utilized in cloud computing are presented. The percentage is.

Several scheduling possibilities are available in proactive load balancing methods. It is clear that proactive methods are used more often. used to schedule tasks and resources than they are to schedule virtual machines, each of which contributes 55.62%.

hybrid algorithm, which makes up 16%, and algorithms, which account for 30%. whereas % of proactive techniques are single-objective, 54% of proactive approaches are multi-objective. 16, 65% of reactive techniques are single goal whereas 59% are multi-objective .

The testing environment used to assess a certain approach's performance indicators . It is evident that CloudSim simulator—which accounts for 46.56% of experimental implementation—is widely utilized for performing simulation experiments. Cloud Analyst simulator came in second with 23.78% of experimental implementation. Matlab's load balancing implementations in C and C++ each account for 14.34% of the total, while others make for 17.68%. There is virtually little real-time application of cloud load balancing techniques.

V. CONCLUSION

The review that is offered in this paper discusses certain significant topics that have not been adequately addressed in either the technical or survey literature but which are severely required by cloud load balancing.

In this part, we thus describe some unfinished studies.

The intricacy of the algorithm has a big impact on how well any load balancing solution performs. It is discovered that only 6 of the 42 prospective technical articles taken into consideration in this study define the appropriate algorithmic difficulty, amounting to 48% of the total, and that do not define the algorithmic complexity, amounting to 68% of the whole. Because the bulk of studies do not address algorithmic complexity, it is recommended for future researchers to use algorithm complexity as a standard when creating a new load balancing strategy with increased practicality.

Always include migration, namely task migration, in your reactive load balancing strategy. The expense associated with moving work is known as the migration cost.In contrast to The analysis reveals that little research has been done on cloud load balancing that concentrates on migration costs, Service Level Violations, Task Rejection Ratio, and Power Usage. It is important for upcoming researchers that reactive strategies that have minimal migration costs continue to improve in this area.

VI. FUTURE SCOPE

In the paper, load balancing strategies used in evaluated publications are compared. The issue of load unbalancing in cloud computing was explored, as well as the contributing causes. An The tasks involved in the load balancing process were also briefly covered along with the abstracted load balancing model. The topic was researched according to rules using the Constructive Generic Framework (CGF), which the Systematic Literature Review method helped to enhance. We came up with a list of pertinent questions and discussed them at work. Data for this study were gathered from IEEE Explore Digital Library, Science Direct, ACM Digital Library, Springier, and Elsevier, five recognized prospective databases.

Many tools and sophisticated filtering choices were used to aid in the data search process. The information was gathered from June 2011 to June 2019.

In this study, a multilevel taxonomy-based categorization that bases its decision on five criteria was presented. The "Nature of Algorithm" criterion was the most significant one employed in this paper. On the basis of these characteristics, we divided 35 articles into two major groups: 28 pieces were reactive in nature, while 10 were proactive. According to the statistics, proactive strategies are always dynamic but reactionary strategies do not always need to be. Moreover, we generalized that while all dynamic techniques may not always be proactive, all proactive approaches are dynamic. The study also showed that work scheduling has received a lot of attention.

There are still a lot of problems, it is concluded from the review done during this study procedure.

There are still gaps in the load balancing process that may be filled in the future, especially in terms of new Quality of Service metrics and algorithm complexity assessment, by using an effective and advanced load balancing algorithm. The study also includes ataxonomy of algorithms, such as machine learning, algorithms derived from mathematics, and algorithms inspired by nature, that might help future researchers cope with the load unbalancing problem successfully.

VII. REFERENCES

1) Dam, S., Mandal, G., Dasgupta, and P. (2017, February) In the cloud, a hybrid load balancing technique is based on genetic algorithms and gravity emulation computing.
2) Dave A, Patel B, Bhatt G (2018, October) Load balancing in cloud computing using optimization techniques: a study. In: International Conference on Communication and Electronics Systems.
3) (2016) December Achar R, Rao SS. Afzal, G. Kavitha (2018, December) Using the IMDLB technique, work transfer costs in cloud computing infrastructure are optimized. International Conference on Circuits and Systems in Digital Enterprise Technologies cloud load balancing based on virtual machine live migration.
4) N. Pasha, A. Agarwal, and R. Rastogi (2015) In a cloud computing context, the VM load balancing technique uses a round-robin approach.
5) Luhach AK, Abdhullah SS (2017) a load balancing method that uses the bat algorithm and is ideal for cloud computing environments. Indian J. Sci. Technology.
6) R. Kapur (2016, July) a method for resource scheduling in cloud computing that balances workload.

7) SS Rajput and VS Kushwah (2017) An enhanced min-min task scheduling technique for load balancing in cloud computing that is genetically based.

8) Karthik K, Reddy VK, and Kanakala VR (2016) Analysis of the effectiveness of load-balancing methods in a cloud computing environment. International Conference on Electrical, Computer, and Communication Technologies.

9) S. Dam, G. Mandal, K. Dasgupta, and P. Dutta (2017, February) hybrid load balancing approach in cloud computing based on genetic algorithms and gravity emulation. Third International Conference on Computer, Communication, Control, and Information Technology Proceedings.

10) Dave, B. Patel, and G. Bhatt (2016) A research on load balancing in cloud computing utilizing optimization methods. Symposium on Communication and Electronics Systems.

11) Modified round Robin method for resource allocation in cloud computing, Pradhan P, Behera PK, Ray BNB (2018).

12) Lumb I, Choi E, and Rimal BP A classification and analysis of cloud computing systems

13) Achar R, Thilagam PS, Soans N, Vikyath PV, Rao S, and Vijeth AM cloud load balancing based on virtual machine live migration.

14) Malhotra, D. Juneja, and A. Singh (2016) cloud computing technique for autonomous agent-based load balancing.

15) Efficient resource management for virtual machines in a cloud context using a dynamic, well- structured load balancing algorithm

16) Using an updated weighted round robin method for non preemptive dependent jobs to balance load in a cloud computing environment.

17) Cloud computing's priority-based modified active monitoring load balancing technique.