IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10720

Load-balancing algorithms in cloud computing

[1.]Barot Hinal, [2]Prof.Riddhi Patel

Dept. of Computer Engineering, LDRP Institute of Technology, & Research, Gandhinagar, India[1]

Dept. of Computer Engineering, LDRP Institute of Technology, & Research, Gandhinagar, India[2]

ABSTRACT- In contempt of the importance of load balancing techniques to the best of our knowledge, there is no comprehensive, extensive, systematic and hierarchical classification about the existing load balancing techniques. Further, the factors that cause load unbalancing problem are neither studied nor considered in the literature. This paper presents a detailed encyclopedic review about the load balancing techniques. The advantages and limitations of existing methods are highlighted with crucial challenges being addressed so as to develop efficient load balancing algorithms in future. The paper also suggests new insights towards load balancing in cloud computing.

Keywords- Cloud computing, Load Balancing Static Algorithms; Dynamic Algorithms, Hierarchical Load Balancing.

I.INTRODUCTION

Cloud computing is the modern internet based service provider which provide on-demand services to clients to access the shared group of resources. The resources can be hardware or software. The cloud computing gets attention from both academic and industrial communities due to its paradigm that "Everything as a Service".Due to the advancement of cloud computing, many enterprises and individuals are using itfor the storage of large data instead of building and maintaining their own local data centres.

In today's world, the services on cloud are provided same as utility services like water and electricity. You need to pay as much as you used these services. The cloud users may also enjoy various types of computing services offered by public cloud. The users now focus on their main objective without worrying about computing resources requirement. With time there is continues increase for the cloud computing resources and it lead to more efficient utilization of computing resources. To provide the has sele free services to its client it is necessary to improve the efficiency which further lead to increase the through put.Because of the implications for greater flexibility and availability at lower cost, cloud computing is a subject that has been receiving a good deal of attention [1].

Cloud Computing is an internet based network technology that shared a rapid growth in the advances of communication technology by providing service to customers of various requirements with the aid of online computing resources. It has provisions of both hardware and software applications along with software development platforms and testing tools as resources[2].

The basic services that the user get from cloud computing architecture are:

1. Software as a Service(SaaS)–In this service end user don't need to install and run different types of applications on their local machine. They can directly use these applications over a cloud network as a service.example is SalesForce.com.

2. Platform as a Service(PaaS)- In this service user can get platform for development and management of Software. In this the software developer can develop and deploy different types application without worrying about tools, languages and API. All these services are provided by cloud service providers. Example is Google App Engine.

3. Infrastructure as a Service(IaaS)-In this service the users are provided storage space in cloud and resources for computation as per their demand and pay as per their usage. Amazon EC2 is an example for the same[1].

Allocation of cloud resources to users on demand gives rise to the problem of load balancing. If workload is not distributed properly, then some nodes in cloud will be heavily loaded and some nodes will be under loaded. In the same way if the resources provided by the cloud are not allocated efficiently, it leads to delay in providing service to the users. Load imbalance may cause system bottleneck. To achieve resource utilization and no delay in providing service, resource allocation should be done in an efficient way.

The main goals of this paper are as follows:

- Studying the existing load balancing mechanisms
- Providing a new classification of load balancing mechanisms
- Clarifying the advantages and disadvantage of the load-balancing algorithms in each class
- Outlining the key areas where new researches could be done to improve the load-balancing algorithms[1].

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

Vol. 10, Issue 7, July 2021

DOI 10.17148/IJARCCE.2021.10720

II.LOAD BALANCING:

The main purpose of load balancing is to provide optimize usage of computing resources to reduce the deployment and operational cost for users. The importance of load balancing in cloud computing is to provide efficient solution to handle multiple request of multiple client by effective use of computing resources available[1].

The Load Balancing is one of the major problems in cloud computing. Now a days load balancing becomes a major challenge and concerns because of the following reasons:

1. Improvement of performance in cloud environment
2. Maximize the throughput
3. Minimize Response Time
4. Minimize the Cost incurred to user
5. Optimize the resource utilization
6. Save Energy

Efficient Load balancing will ensure [3]:

Uniform distribution of load on nodes:

- Improves overall performance of the system
- Higher user satisfaction
- Faster Response
- System stability
- Reducing carbon emission

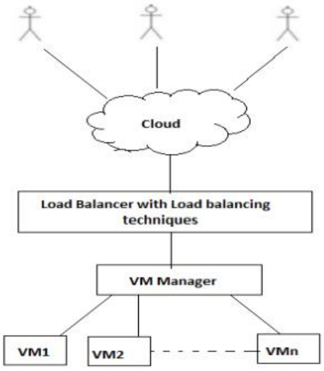

Fig -1: General Structure of Load balancing in Cloud Environment [3].

III.CLASSIFICATION OF LOAD BALANCING

Algorithms There are many load balancing algorithms. Generally Load balancing algorithms are classified into two categories based on the present system state [3].

1).STATIC ALGORITHMS

Static algorithms are best in homogeneous and stable environments. However, static algorithms are not flexible and cannot consider the dynamic changes to the attributes. While assigning tasks to the nodes, static load balancing algorithms will not check the state and functionality of the node in previous tasks[3].

2).DYNAMIC ALGORITHMS :

Dynamic algorithms provide better results in heterogeneous and dynamic environments. These algorithms are more flexible. Dynamic algorithms can consider the dynamic changes to the attributes. However, these algorithms are more complex [3]. Main advantage of this is that selection of task is based on current state and this will help to improve the performance of the system.[3].

- **What is meant by load balancer?**

The main aim of the load balancer helps to assign resources equally to the tasks for resource efficiency and user satisfaction at minimal expense [4], quality output, gripping rapid traffic blast sustain traffic on the website and elasticity which motivates us to identify problems in LB and to work on their resolution . This plays a key role in ensuring the ease of access for customers, business partners, and end-users of the cloud-based applications [4]. The potential of the load balancing and its various applications provide motivation to deliberate on this important challenge, to identify major issues in LB and to resolve these issues [4].

- **LOAD BALANCING** :

DOI 10.17148/IJARCCE.2021.10720

LB's primary objective is to efficiently manage the load across various cloud nodes, so that no nAode is under/ overloaded [4]. LB may be characterized as a process of spreading a burden across network links on multiple devices or system clusters to maximize its use of assets to optimize overall response time. It reduces the device's total waiting period and also avoids excessive replication of assets. Requests spread inside servers in this process so that data can be distributed & processed without waiting. LB is the method of maximizing system performance by moving the device burden. LB provides a systematic mechanism for the equal distribution of the responsibility to the resources available. The goal is to provide reliable service, including adequate use of the resource, in the event of a disaster of the portion of any service by supplying & de-provisioning the device instance. In addition, LB is aimed at reducing response time for tasks & increasing resource efficiency, which increases device efficiency at a lower cost [4].

IV.COMPARISON BETWEEN VARIOUS ALGORITHMS[3]:

Algorithm	Advantages	Disadvantages
Round Robin Load Balancing Algorithm	It is Simple • algorithm and emphasis is on fairness. It works in • circular fashion. Fast response • in the case of equal workload distribution	Each node is • fixed with a time slice. It is not flexible • and scalable. Some node may • possess heavy load and some nodes are idle. Does not save • the state of previous allocation of a VM. Pre-emption is • required
MIN-MIN Load Balancing Algorithm	It is simple • and fast algorithm. It works • better for smaller task.	Selects the task • having minimum completion time. •poor load balancing • Does not consider the existing load on a resource.
MIN-MAX Load Balancing Algorithm	It is simple• algorithm. It runs short• tasks concurrently.	Selects the task • having the maximum completion time •there is a starvation. Larger tasks will execute first, while the smaller tasks need to wait. • Poor load balancing.
Honeybee Foraging Behavior Load Balancing Algorithm	Selforganizing, • nature inspired algorithm. • Performance will be achieved by increasing the system size. •suitable for heterogeneous environment	Increase in • resources will not increase the overall throughput.
Throttled Load Balancing Algorithm	List of VMs• is maintained along with the status of each VM •good performance •better resource utilization	Scans the entire • list of VMs from the beginning •does not consider the current load on VM.
Biased Random Sampling Load Balancing Algorithm	• Fully decentralized • Suitable in large network	Performance is • degraded with an increase in diversity
Modified Throttled Load Balancing Algorithm	• Index table is parsed from the index next to already assigned VM. •faster response than throttled algorithm	• Does not consider the current load on VM.
Hierarchical Load Balancing	•faster response • Suitable for homogeneous and heterogeneous environment	• Less fault tolerant

V.NATURE OF THE ALGORITHM:

The first categorization of load balancing algorithms in this work has been done on the basis of nature of algorithm. On the basis of this classification, LB algorithms are classified as proactive based approaches and reactive based approaches. However in other fields of technology particularly in the communication and networking for mobile networks (MANETS), the nature of the communication routing protocols has been extensively studied under these two variants[2].

A proactive based LB algorithmic technique is an approach to algorithmic design which takes into consideration action by causing change and not only reacting to that change when it happens. It is intended to yield a good outcome to avoid a problem in advance rather than waiting until there is a problem. Proactive behavior aims at identification and exploitation of opportunities and in taking preemptory action against potential problems and threats. The limitation of existing approaches is that a limited number of proactive approaches have been used and that too in a traditional manner with no novel concepts[2].

VI.STATE OF THE ALGORITHM:

On the basis of state information of system that an algorithm relies on, LB algorithms are widely classified as static, dynamic and hybrid. From existing literature survey, it is evident that this is most widely used classification system for LB algorithms. Majority of work on comparative studies on load balancing begin the algorithmic taxonomy by placing this category on top of taxonomy. In static load balancing, traffic load is segregated uniformly across the servers. This is done by algorithm having the prior knowledge about system resources and task requirements. The static LB algorithm schedules tasks to VM for execution at compile time[2].

VII.CHALLENGES IN CLOUD-BASED LOAD BALANCING:

Review of the literature shows that load balancing in cloud computing has faced some challenges. Although the topic of load balancing has been broadly studied, based on the load balancing metrics, the current situation is far from an ideal one. In this section, we review the challenges in load balancing with the aim of designing typical load balancing strategies in the future[5]. Some studies have mentioned challenges for the cloud-based load balancing (Palta and Jeet, 2014; Nuaimi et al., 2012; Kanakala and Reddy, 2015a, 2015b; Khiyaita et al., 2012; Ray and Sarkar, 2012; Sidhu and Kinger, 2013).

1) Virtual machine migration (time and security) :

The service-on-demand nature of cloud computing implies that when there is a service request, the resources should be provided. Sometimes resources (often VMs) should be migrated from one physical server to another, possibly on a far location. Designers of load-balancing algorithms have to consider two issues in such cases: Time of migration that affects the performance and the probability of attacks (security issue)[5].

2) **Spatially distributed nodes in a cloud**:

Nodes in cloud computing are distributed geographically. The challenge in this case is that the load balancing algorithms should be designed so that they consider parameters such as the network bandwidth, communication speeds, the distances among nodes, and the distance between the client and resources[5].

3) Single point of failure

some of the load-balancing algorithms are centralized. In such cases, if the node executing the algorithm (controller) fails, the whole system will crash because of that single point of failure. The challenge here is to design distributed or decentralized algorithms[5].

4) Algorithm complexity

The load-balancing algorithms should be simple in terms of implementation and operation. Complex algorithms have negative effects on the whole performance[5].

5) Emergence of small data centers in cloud computing

Small data centers are cheaper and consume less energy with respect to large data centers. Therefore, computing resources are distributed all around the world. The challenge here is to design load-balancing algorithms for an adequate response time[5].

VIII.HIERARCHICAL LOAD BALANCING ALGORITHM

Hierarchical Load Balancing involves different levels in load balancing decisions. Every node is managed or balanced by its parent node [3]. Parent node is responsible for load balancing [3]. Hierarchical load balancing can be used in homogeneous as well as heterogeneous environment [3]. Cluster can also be used in hierarchical load balancing. Clustering is the process of organizing similar type of objects into groups. VM's having similar characteristics are logically grouped. VM's are at last level[3].

IX.ACTIVITIES INVOLVED IN LOAD BALANCING

Scheduling and allocating tasks to VMs based on their requirements constitute the cloud computing workload. The load balancing process involves the following activities [2].

1. Identification of user task requirements

This phase identifies the resource requirement of the user tasks to be scheduled for execution on a VM[2].

2. Identification of resource details of a VM

This checks the status of resource details of a VM. It gives the current resource utilization of VM and the unallocated resources. Based on this phase, the status of VM can be determined as balanced, overloaded or under-loaded with respect to a threshold[2].

3. Task scheduling

Once resource details of a VM are identified the tasks are scheduled to appropriate resources on appropriate VMs by a scheduling algorithm[2].

4. Resource allocation

The resources are allocated to scheduled tasks for execution. A resource allocation policy is being employed to accomplish this. While, scheduling is required for speeding up the execution, allocation policy is used for proper resource management and improving resource performance. The strength of the load balancing alalgorithm is determined by the efficacy of the scheduling algorithm and the allocation policy[2].

5. Migration

Migration is an important phase in load balancing process in cloud and latter is incomplete without the former. Migration is of two kinds in cloud based on entity taken into consideration- VM migration and task migration. VM mimgration is the movement of a VM from one physical host to another to get rid of the overloading problem and is categorized into types as live VM migration and non live migration[2].

X.PARAMETERS OF LB

The parameters concerning cloud LB in a much more practical sense will not only enhance output processing by LB the process but also make the theoretical basis for studying efficient algorithms to boost LB efficiency on CC [103]. LB refers to the efficient methods used for cloud workload allocation between VMs. Within a cloud network, the versautility of the VMs depends on the degree of load distributed across existing resources. A decent scheduler allows for a reliable method of load control. Parameters of CC, namely, are important. The performance measurements recognized in the LB methods are divided into two major quantitative & qualitative parameter classifications. In fact, the parameters can also be either receptive or autonomous. The taxonomy of LB metrics is shown in Figure -2[4].

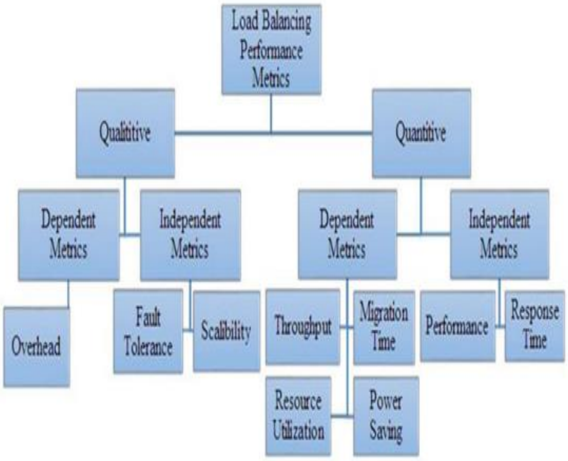

Fig -2.LB parameters

In addition to the current load balance parameters, there are a few performance parameters added in this work. thermore, if novel parameters are identified in the future, they may be put according to their characteristics in the categorization. Taxonomy classifies the cloud LB parameters into four distinct groups [4]:

1) LB performance parameter with qualitative attributes & dependent nature [4].
2) LB performance parameters with qualitative attributes & independent nature [4].
3) LB performance parameters with quantitative attributes & dependent nature [4].
4) LB performance parameters with quantitative attributes & independent nature [4].

XI.FUTURE WORKS

Balancing of the workload among cloud nodes is one of the most important challenges that cloud environments are facing today. In this paper, we surveyed research literature in the load balancing area, which is the key aspect of cloud computing. We found in the literature, several metrics for load balancing techniques that should be considered in future load balancing mechanisms. Based on our observations, we have presented a new classification of load balancing techniques:
(1) Hadoop Map Reduce load balancing category
(2) natural phenomenon based load balancing category
(3) agent-based load balancing category
(4) general load balancing category. In each category, we 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 we considered for load balancing techniques. Recently, load balancing techniques are focusing on two critical metrics, That is, energy saving and reducing carbon dioxide emission. As future works, we suggest 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.

XII.CONCLUSION:

Cloud computing allows wide range of users to access distributed, scalable, virtualized, hardware and softwareresources over the Internet. Load balancing is one of the most important issue of cloud computing. It is a mechanism which distributes workload evenly across all the nodes in the whole cloud. Through efficient load balancing, we can achieve a high user satisfaction and resource utilization. Hence, this will improve the overall performance and resource utility of the system. With proper load balancing, resource consumption can be kept to a minimum which will further reduce energy consumption and carbon emission rate. Through hierarchical structure of system, performance of the system will be increased.

REFERENCES:

1) Pooja Arora, Anurag Dixit," An optimized Load Balancing Algorithm in Cloud Computing" (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-5, June 2020
2) Shahbaz Afzal* and G. Kavitha," Journal of Cloud Computing: Advances, Systems and Applications"@2019.
3) Sajjan R.S, Biradar Rekha Yashwantrao" Load Balancing and its Algorithms in Cloud Computing" Volume-5, Issue-1 on 16 February 2017.
4) MUHAMMAD ASIM SHAHID, NOMAN ISLAM, MUHAMMAD MANSOOR ALAM , MAZLIHAM MOHD SU'UD , AND SHAHRULNIZA MUSA "A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment" date of publication July 14, 2020.
5) Einollah Jafarnejad Ghomi , Amir Masoud Rahmani,⁎ , Nooruldeen Nasih Qader" Load-balancing algorithms in cloud computing" 08 April 2017
6) J. Rathore, ''Review of various load balancing techniques in cloud computing,'' Comput. Sci. Electron. J., vol. 7, no. 1, p. 5, 2015.
7) Qiiu, Weidong Cai, Jian Shen, Xiaodong Liu, Nigel Linge, "An Adaptive Approach to Better Load Balancing in a Consumer-centric Cloud Environment," IEEE Transactions on Consumer Electronics, Vol: 62, no: 3, pp: 243 - 250, August 2016.
8) Narander Kumar and Diksha Shukla," Load Balancing Mechanism Using Fuzzy Row Penalty Method in Cloud Computing Environment", Information and Communication Technology for Sustainable Development
9) Patel G, Mehta R, Bhoi U (2015) Enhanced load balanced min-min algorithm for static meta task scheduling in cloud computing.
10) A. A. Jaiswal, Dr. Sanjeev Jain," An Approach towards the Dynamic Load Management Techniques in Cloud Computing Environment", 2014 International Conference on Power, Automation and Communication (INPAC)
11) Surbhi Kapoor, Dr. Chetna Dabas," Cluster Based Load Balancing in Cloud Computing", 2015 Eighth International Conference on Contemporary Computing (IC3).
12) G. Wang, S. Deb and L. d. S. Coelho, "Elephant Herding Optimization, "In proceedings of 3rd International Symposium on Computational and Business Intelligence (ISCBI),pp:1-5,2015.
13) A.N. Ivanisenko; T. A. Radivilova , "Survey of major load balancing algorithms in distributed system ", Information Technologies in Innovation Business Conference (ITIB), 2015.