RESEARCH ARTICLE | MAY 02 2022

Survey on various load balancing algorithms in cloud computing

K. Vinod Kumar ✉; R. Balakrishna; Santhosh Kumar

AIP Conference Proceedings 2463, 020002 (2022)
https://doi.org/10.1063/5.0080325

Survey on Various Load Balancing Algorithms in Cloud Computing

Vinod Kumar K [a)], Balakrishna R, Santhosh Kumar

Department of Computer Science and Engineering, Vels Institute of Science Technology and Advanced Studies (VISTAS), Tamil Nadu, India.

[a)] Corresponding author: kethineni.vinod@gmail.com

Abstract. In today's world Cloud computing is perhaps the most utilized technology. Distributed computing offers web administrations, information stockpiling and computing assets over the web effortlessly. Cloud computing permits clients to get to the IT assets anyplace and whenever on a compensation for every utilization premise. As the interest for distributed computing develops quickly, the traffic additionally increases. There are two answers for this issue, one is to upgrade a single server to a high-performance server but upgraded server may also overload soon and second is multi server (gathering of workers). Multi server arrangement is adaptable and practical. While making a group of servers, the issue is Load adjusting. Load balancing is one of the critical issues in cloud computing. Load balancing is an interaction of parting the unique responsibility and dispersing the responsibility among every one of the hubs with the end goal that no hub in the cloud climate is overloaded/under loaded or inactive. Load balancing in Cloud computing is utilized to improve the proficiency and utilizing resources effectively. This journal paper objective is on investigation on cloud load balancing, Load balancing methods, Load balancing algorithms.

Keywords: Cloud computing, Load balancing, Load balancing algorithms.

INTRODUCTION

The term Cloud refers Network or Internet. Cloud can offer types of assistance through networks for i.e on public or private organizations. Distributed computing is an attractive technology in the field of software engineering. It offers online data storage, framework and application. Cloud computing is a compensation for every utilization model methods clients need to pay for what they use. Above all, the critical focal point of cloud computing is to make a data center more exceptional so they offer dynamic and versatile kinds of administrations to the cloud clients [6].

Cloud computing became a worldwide trend by providing services Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Identity as a Service (IDaaS) and Network as a Service (NaaS).

1. Infrastructure as a Service (IaaS) Model: Iaas gives admittance to essential assets like actual machines, virtual machines, virtual storage and so on.

2. Platform as a Service (PaaS) Model: PaaS ensures the run time environment for applications.

3. Software as a Service (SaaS) Model: SaaS model licenses us to pass on software applications to the end clients.

4. Network as a Service (NaaS) Model: Naas permits us to access network infrastructure directly and securely. Cloud computing has a lot of benefits, yet there are a few issues to manage for example, load adjusting among the assets, scheduling of tasks, VM relocation, security, insurance of information, recovery of information and accessibility, interoperability, portability, fault tolerance, Cloud computing administration and some more.

International Conference on Recent Innovations in Science and Technology (RIST 2021)
AIP Conf. Proc. 2463, 020002-1–020002-8; https://doi.org/10.1063/5.0080325
Published by AIP Publishing. 978-0-7354-4192-7/$30.00

This paper focuses on how load balancing among the resources takes place. If clients in the cloud expands, the responsibility additionally builds, clients send requests to servers in the cloud. In some cases, a few servers just cycle and complete those requests, while some are quiet. This leads to overloading/under loading of servers. For this issue Load balancing came into the picture.

CLOUD LOAD BALANCING

Load balancing is an interaction of partitioning the workload similarly among every one of the accessible hubs to accomplish high asset use proportion and client fulfillment [1]. Load balancing utilizes a procedure to part responsibility on the virtual machines so every virtual machine does an equivalent measure of work. Load balancing is utilized to accomplish effective asset use which expands throughput and decreases reaction time [2].It is used to maximize the throughput. Reducing cost, energy consumption and workload traffic, get minimal response time, improving the utilization of resources and performance. Load balancing allows firms or organizations to manage application or workload demands by splitting and distributing resources among various desktops, computers, networks or servers (nodes). The main aim of a cloud load balancing scheme is to maintain system consistency, firmness, performance and protect against system failures [3].

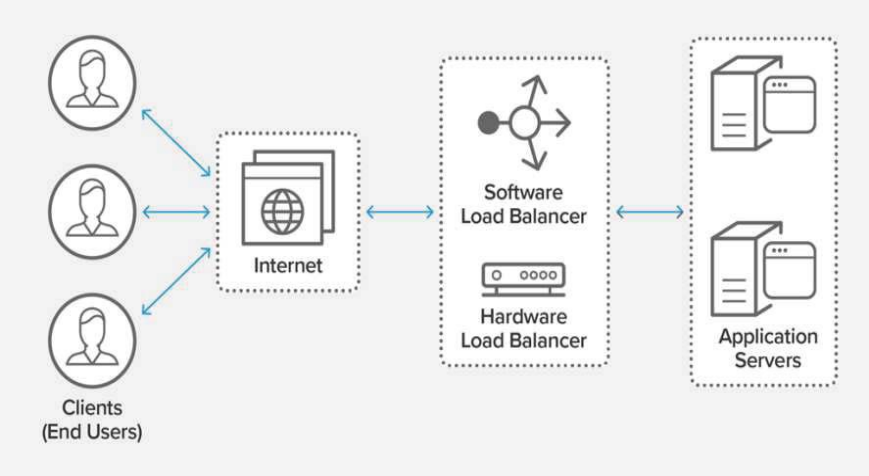

FIGURE 1. Cloud Load Balancing

CHALLENGES OF LOAD BALANCING

Load balancing has many challenges:

1. **Scalability:** Permit people to get to assets for quick down scaling or scale-up whenever [5].
2. **Data handling:** Permits consumers to heterogeneously hold the information, with no control issues [6].
3. **Energy management:** Power economizing is the most important thing in cloud load balancing [5].
4. **Complexity of algorithms:** Algorithms in cloud computing ought to be simple to achieve [5].
5. **Single point of failure:** LB choices are made by the centralized node. On the off chance that the key gadgets break down this will affect the general registering system [6].
6. **VM migration** :
   Virtual machine migration takes place in three steps:
   1. Load adjusting which inspect the current load on machine asset
   2. Resource disclosure which recognize another suitable resource
   3. Workload migration which moves additional assignments to accessible resources.

GOALS OF LOAD BALANCING

1. Improvement in System Performance
2. Secure against framework disappointments
3. Strength support of the framework
4. Keep up System immovability and dependability
5. Capacity to deal with unexpected traffic spikes
6. High asset availability [5]
7. Expanding asset use
8. Throughput ought to be maximum
9. Least reaction time
10. Improvement in unwavering quality
11. Lessen holding up time

ASSIGNING OF JOBS/TASKS

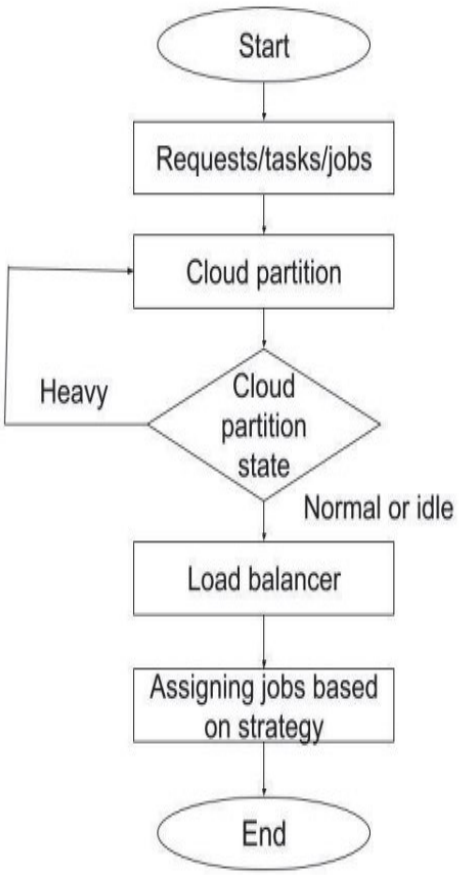

FIGURE 2. Jobs Scheduling

VARIOUS SORTS OF LOAD BALANCING TECHNIQUES

In View of Condition of the System

Load balancing techniques categorized into 3 norms [6]

- Static load balancing: A static load balancing algorithm takes the choice of moving the heap doesn't rely upon the continuous condition of the system. Cloud requisite prior information like server capacity, power of processing, Memory, network, performance, user requirements.
- Dynamic load balancing: Dynamic algorithms take the continuous condition of the system into account while making the decision of shifting load. It permits processes to move from overloaded nodes to under loaded nodes.
- Hybrid load balancing: The static and dynamic methods are combined in it. It is best for balanced distribution of figuring undertakings.

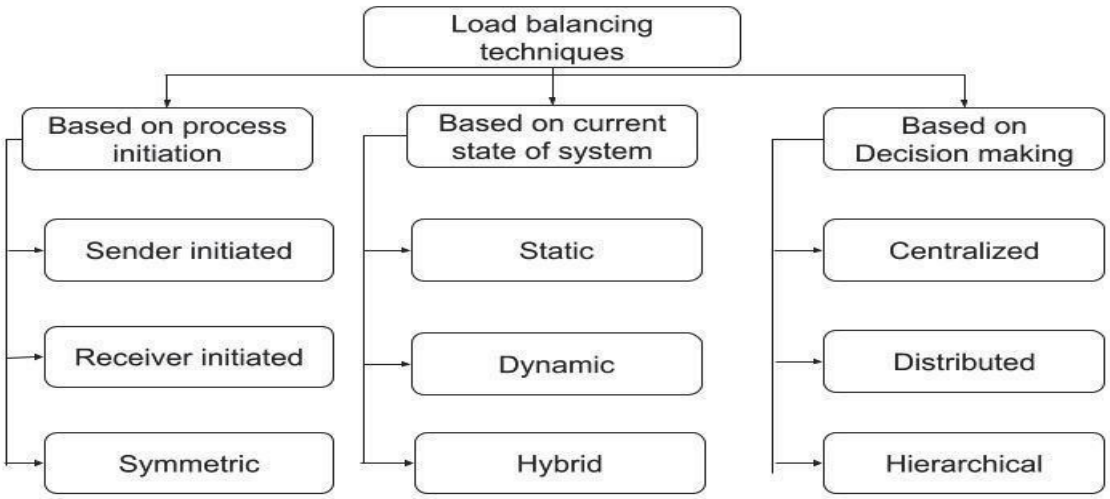

FIGURE 3. Load Balancing Technique

Sender Initiated Load Balancing

In the sender-initiated algorithm, the sender instigates the process. The sender transmits the request messages until it gets a receiver that can allow the workload.

FIGURE 4. Sender Initiation

- **Receiver initiated load balancing:** The receiver starts the interaction. The receiver sends a solicitation message to perceive a sender who is set up to share the responsibility.

- **Symmetric load balancing:** Sender initiated +receiver initiated

FIGURE 5. Receiver Initiation

Based on Decision Making Node

- **Centralized load balancing:** The responsibility dissemination is done from brought together nodes to various cycles.

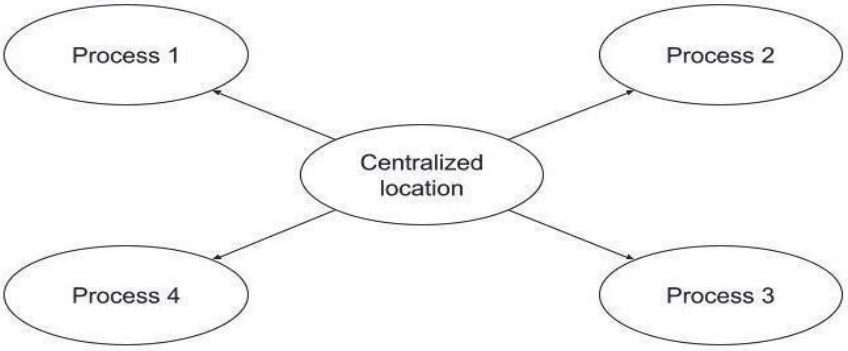

FIGURE 6. Centralized Load Balancing

- **Circulated load adjusting:** Each node takes close by information of the cloud association and ensures productive distribution of tasks.

- **Hierarchical load balancing:** A tree structure is made to address levels of clouds. Parent node settles on choices dependent on data gathered by slave nodes.

TABLE 1. Classification of Different Load Balancing Algorithms

Algorithm	Objective	Performance	Advantages	Disadvantages
1.Min-Min Load balancing algorithm	•Min-Min load Balancing algorithm usage is maximizing the resource usage.	•Min-Min load Balancing calculation is a static burden adjusting calculation. • Min -Min Load balancing algorithm performance is regarding Expected completion time. • First calculate the minimum completion time the base fruition time to every accessible node. After calculations are completed, which task has the minimum completion time is picked and assigned to the individual node. • First execute the small tasks and large tasks keep waiting until all small tasks are completed. It will finally result in the poor use of machines.	•Min-Min load balancing algorithm is a basic and quick calculation to improve execution. • It has a better make span when compared to all other algorithms. • Maximize the resource utilization	• Min-Min load balancing algorithm leading to Starvation • Quality of Service is not considered. <u>Draw Back:</u> • In the Min-Min algorithm the drawback is first to assign the small tasks.
2.Max-Min load Balancing algorithm	•Max-Min load Balancing algorithm aim is also maximizing the resource utilization	• Max-Min load balancing calculation is also Static load adjusting calculation. • It is much identical to the Min-Min load adjusting algorithm. • It is also performed based on the Expected completion time. •Initially all accessible tasks are allocated to the system and determined at the base finishing time. • In this minimum completion time of tasks. Choose which one having minimum completion time is maximum and that one is given to the system.	• Max-Min load balancing algorithm has better performance than Min-min balancing algorithm. • It also has better make span compared to all other algorithms. • Maximize the resource consumption	• Max-Min load balancing algorithm suffered from starvation. • It is also Quality of Service not considered.

Algorithm	Objective	Performance	Advantages	Disadvantages
3.Ant Colony Optimization Based Load Balancing Algorithm	• Subterrance insect settlement enhancement based burden adjusting objective is to track down the ideal way from state of insects to wellspring of food based on subterranean insect's conduct.	• Ant Colony Optimization Based Load Balancing Algorithm is Dynamic load balancing calculation. • It is performed to track down the ideal way. • In this algorithm to effectively circulate responsibility among the nodes. • The ant begins development towards the food from the head node.	•Subterranean insect Colony Optimization Based Load Balancing Algorithm great execution, to diminish the tasks. • It has high effectiveness • Quickly coupled	.• Ant Colony advancement Based Load adjusting Algorithm is untested in the genuine cloud world. • In This algorithm fault tolerance factor is not considered.
4.Round Robin Load Adjusting Algorithm.	• Round Robin Balancing calculation objective is additionally amplifying the resource usage and increment throughput.	• Round Robin Load balancing algorithm is a Dynamic and Static Load adjusting algorithm. • It assigns tasks to all machines in a round way and executes those undertakings for a specific measure of time called Time Quantum.	• It utilizes the resources in an effective manner. • It works effectively when the no. of undertakings to oversee is more noteworthy than no. Of machines (servers).	• It doesn't notice the worker load and aimlessly allots a smaller task to a high-performance server. • It does not check the failure of servers.

Algorithm	Objective	Performance	Advantages	Disadvantages
5.Shortest Job First Load Balancing Algorithm	•Removes the Starvation by executing the more modest task first, so it diminishes the waiting time of smaller tasks.	•Shortest Job First Load Balancing algorithm has both Preemptive and Non preemptive modes. • The Preemptive SJF functions admirably in parts of waiting time and response time.	• It functions admirably in the event of various responsibilities with various sizes. • It reduces the holding up time of all tasks	• Getting the earlier data about load sizes is troublesome. ' • Due to less information about upcoming loads the implementation is difficult.

CONCLUSION

As such cloud computing being a wide area of exploration and one of the significant topics of research is load balancing, the accompanying paper gives how load balancing techniques and algorithms are valuable in cloud conditions. The objective of load balancing is to build customer fulfillment, asset use, significantly, the presentation of the cloud framework, minimize the reaction time and reduce the energy burned-through.

REFERENCES

1. P. Chawla and A. Kumar, "A Systematic literature review on load balancing algorithms of Virtual Machines in a Cloud computing environment," International Conference on Innovative Computing and Communications (2020).
2. X. Fu, G. Xu and J. Pang, "A load balancing model based on cloud partitioning for the public cloud," Tsinghua Science and Technology**18**, 34-39(2013).
3. R. Zaman Khan and M. Haris, "A Systematic Review on Load Balancing Issues in Cloud Computing," Sustainable Communication Networks and Application, 297 - 303(2020).
4. P. Kumar and R. Kumar, ''Issues and challenges of load balancing techniques in cloud computing: A survey," ACM Computing Surveys **51**, 1 – 35 (2019)
5. M. AsimShahid, N. Islam, M. Mansoor Alam, Mazliham and S. Musa, "A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach," IEEE Access, **8**, 130500 - 130527 (2020).
6. Suman and Sangeeta, "Load Balancing In Cloud Computing: A Review," International Journal of Advanced Research in Computer Science **15**, 22 - 29(2018).