Cloud Load Balancing

Christy Babu
Integrated MCA
Amal Jyothi College of Engineering
Kanjirappally, Kerala, India

christybabu@mca.ajce.in

Meera Rose Mathew
Associate Profssor Amal Jyothi College
of Engineering Kanjirappally,
Kerala,India

meerarosemathew@amaljyothi.ac.in

ABSTRACT

Cloud computing is a comprehensive approach that allows individuals and organisations to acquire the services they demand based on their specific requirements. The concept provides a variety of services, including storage, deployment platforms, and easy access to online services, among others. Load balancing is a typical problem in the cloud that makes it difficult to maintain application performance while measuring Quality of Service (QoS) and adhering to the Service Level Agreement (SLA) contract as required by cloud providers to organisations. Cloud service providers struggle to distribute workload evenly among servers. By effectively leveraging the resources of VMs, an effective LB approach should optimise and assure high user happiness.

Keywords: *Cloud Computing, load balancing*

I.   INTRODUCTION

Cloud computing is a technology model based on internet which provides the services and products through internet and the pricing based on the principle of pay per use. Cloud computing is entirely dependent on virtualization, which means that instead of purchasing large physical servers or computers, clients can utilise virtual servers and computers to access cloud computing services. One of the main cloud service providers is Amazon and the cloud platform of Amazon is AWS. In this research paper the main topics that cover is the what is the cloud load balancing and which are the main algorithms that helps in the load balancing of cloud computing.
.

LOAD BALANCING OF CLOUD COMPUTING

The traffic In the web is increasing day to day because of the heavy traffic a single server is unable to maintain the traffic of the web.It is necessary to maintain the load of a server to avoid the problems of overloading a server. One of the main problems that cloud computing faces is the problem due to the balancing of the load. Load balancing means the distribution of workload to multiple networks or servers inorder to achieve better performance and quality of service(QOS).

The load can be balanced in two ways:
1.   Upgrading the server to a high speed and high capacity single server
2.   Using multiple servers to manage the traffic of particular location or specifications.

The front end and the backend are the two elements of cloud computing. The front end is the user interface or platform, whereas the backend is the background procedures that provide the entire service and are not visible to the user, such as cloud services and models. This is based on the concept of virtualization. Virtualization is a technology that makes it easier to allocate server or resource resources to a request. Load Balancing is a technique for optimization of the resources of Virtual Machine in the cloud computing.
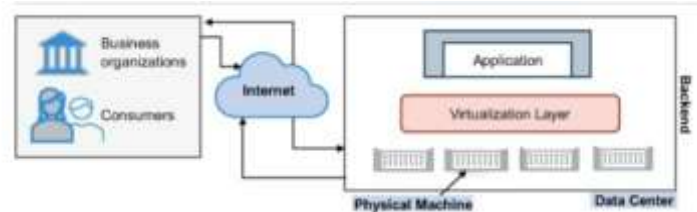

Fig 1. Cloud Computing Architecture

II.   TYPES LOAD BALANCING ALGORITHM

The load balancing algorithms are mainly classified into three categories:

1.   Dynamic Algorithms: Dynamic load balancing algorithms are a type of cloud load balancing algorithm that is flexible and adaptable and is dependent on the system's past status. It's more well-known for its adaptability and versatility.

2.   Static Load Balancing : Static Load Balancing: This approach is based on a prior understanding of the system and its capabilities. It is not flexible to changes in the system or server, for example. The system or server will fail if any unexpected changes occur.

3.   Nature-inspired Load Balancing: These are algorithms that are solely based on biological or genetical processes that are based on human nature. These algorithms were created by a combination of mathematical and machine learning techniques. It is the most precise algorithm, capable of performing precisely in complex and dynamic scenarios.
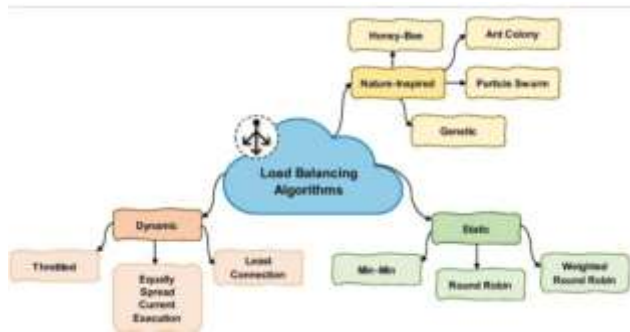
Fig 2. Load Balancing Algorithms

III.   TYPES OF STATIC ALGORITHMS

1.   Round Robin Algorithm

The Round Robin algorithm is the simplest of all algorithms. It is a scheduling technique that does not maintain the state or priority of any job. The works or jobs are assigned a time duration in the round robin process, and the job is automatically halted when the time span expires. If the job is not completed within the timeframe, it is placed in a waiting list and must wait until the next scheduled time to resume work. This algorithm just keeps a list of planned works; it is ineffective since all jobs are treated equally and no priority is assigned to any work.
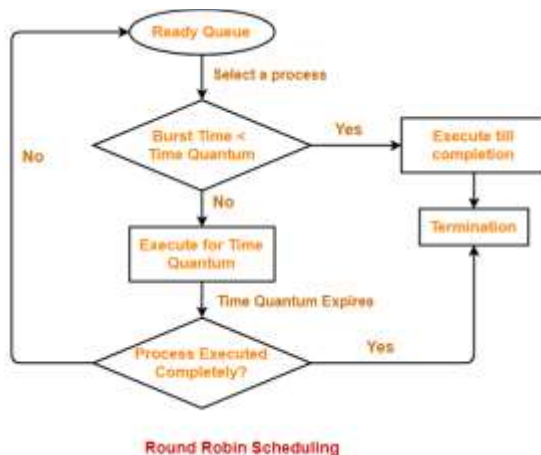

Fig 3. Round Robin Algorithm

2.   Weighted Round Robin Algorithm

The Weighted Round Robin Algorithm is similar to the round robin algorithm in that each work is assigned by weight rather than a time span, and the job is completed based on the weight. The highest-weighted task will be done first.
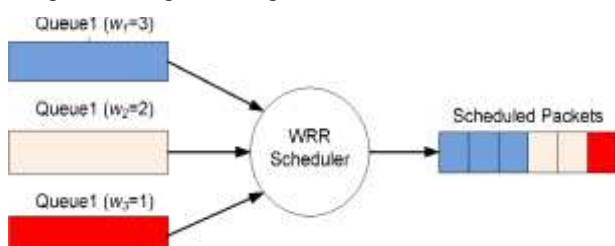

Fig 4. Weighted Round Robin Algorithm

3.   Min-Min Algorithm

The Min-Min Algorithm is an algorithm in which the tasks that take the least amount of time to accomplish are performed first, and the tasks that take the most time are completed last. The biggest disadvantage of this approach is that it does not take into account the task's priority. For example, sometimes the task with the least time is a low priority task and the task with the highest priority is a higher priority task, yet the task with the least time (low priority) will be completed first and the task with the highest priority would be completed last.
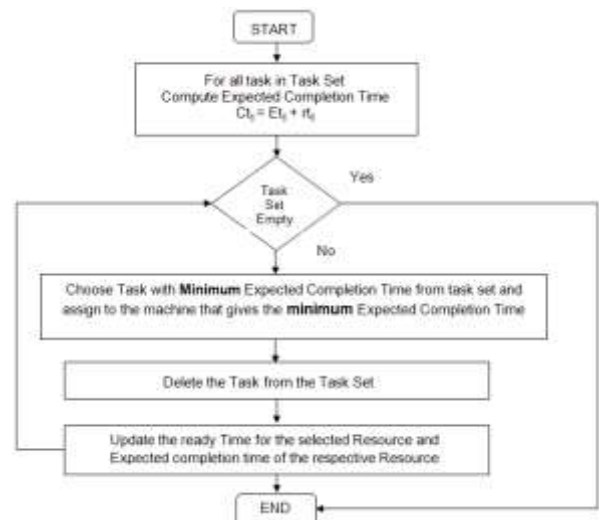

Fig 4. Flow Chart of Min-Min Algorithm

IV.   TYPES OF DYNAMIC ALGORITHMS

1.   Throttled Algorithm

Throttled Algorithm is a form of cloud computing algorithm that belongs to the dynamic category of load balancing algorithms. The throttled algorithm's principal goal is to assign the appropriate virtual machine to the client's request. The Throttled Algorithm keeps track of every virtual machine and their index values. When a client makes a request, the server looks for a virtual machine to serve them. If one is available, the task is accepted and the virtual machine is assigned. The algorithm returns -1 if no virtual machine is available to serve the client.
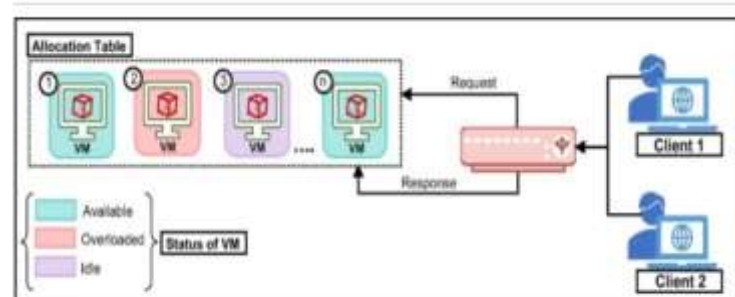
Fig 5. Throttled Algorithm

2. Modified Throttled Algorithm

Modified The throttled algorithm is a modified version of the throttled algorithm that improves cloud computing performance and economy. The improved throttled algorithm first determines whether or not the virtual machine is available. The algorithm gives -1 if the virtual machine is busy. If the virtual machine is available, the algorithms look up the available machine from the list and assign it to the client request based on the index value of the machine.
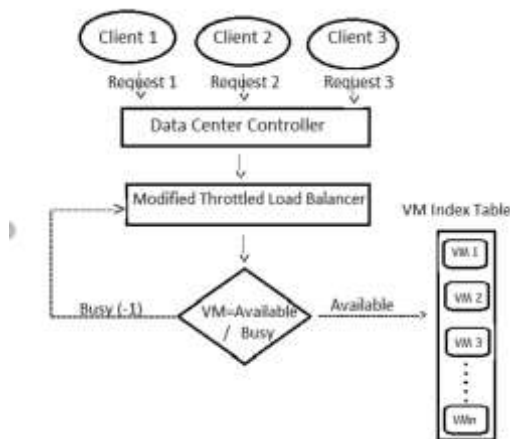
4. Least Connection Algorithm

The least connection algorithm assigns a job or piece of work to the virtual machine with the fewest open connections. If a request is made, the available virtual machines are checked from the index list for the virtual machine with the least workload or active connection, and that virtual machine is assigned to the work.

Fig 6. Modified Throttled Algorithm

Fig 8. Least Connection Algorithm

3. Equally Spread Current Execution

This is a load balancing mechanism that is dependent on the amount of work that the client has requested. In this approach, the priority is given to the workload or size of the work. When a client makes a request, the amount of the task is noted, and the virtual machine that is available with light works is checked, and the work is distributed randomly to the virtual machine. One issue with this technique is the possibility of virtual machine overloading if the index table is modified after the virtual machine has been allocated for a task.
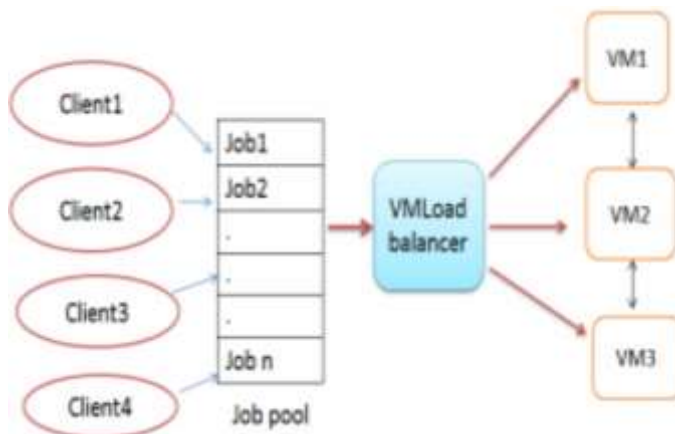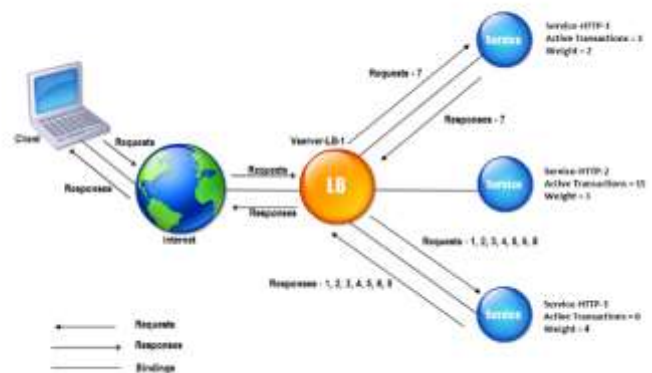
V.   TYPES OF NATURE INSPIRED BASED ALGORITHM

1. Honey-Bee Algorithm

The Honey-Bee algorithm is a nature-inspired cloud load balancing method that is inspired by Honey-bees' food-gathering behaviour. In the case of Honey-Bees, the entire population is looking for food, and the honey bees that find it send the source to others to avoid overcrowding, and the others will look for food elsewhere. The cloud computing model is based on the same principle. The honey bees are the jobs in cloud computing, and the virtual machines are the food source. The tasks assigned to the virtual machine acknowledged to the datacentre that they may be assigned to other virtual machines, avoiding overloading the virtual machine and providing better performance, as well as a higher QOS in this algorithm.
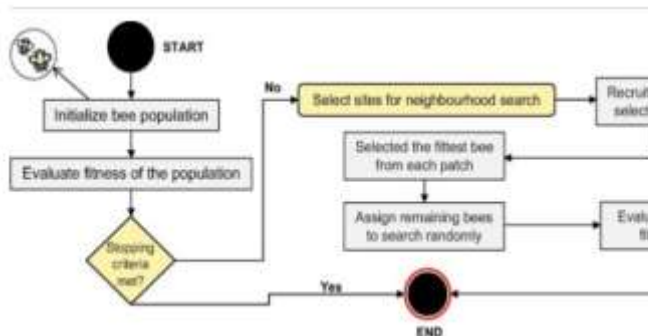

Fig 7. Equally Spread Current Execution Algorithm

Fig 9. Honey-Bee Algorithm

2. Ant-Colony Algorithm

The ant colony algorithm is a load balancing mechanism in cloud computing that is inspired by nature. The algorithm's foundation is based on how ants hunt for food. The ants wander aimlessly in search of nourishment. When they move, they expel a substance called pheromone from their bodies. The amount of pheromone released aids other ants in determining the shortest distance between the colony and the food source. The same approach is utilised in cloud computing for resource optimization, but the main difficulty with this technique is that it takes longer to obtain or generate results, and its performance is inferior to that of other algorithms.
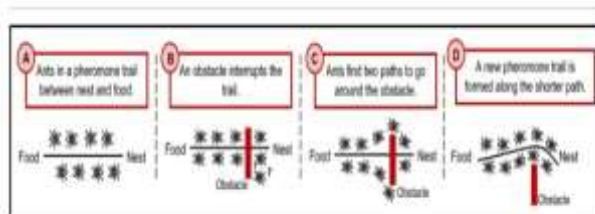

Fig 10. Ant- Colony Algorithm

VI. CONCLUSION

Load balancing is an important feature of Cloud Computing since it helps to improve workload distribution and resource utilisation, which improves the system's total response time. To handle challenges connected to load balancing, such as work scheduling, migration, resource utilisation, and so on, a variety of methodologies and algorithms have been presented. This research provided different approaches to the critical topic of load balancing in cloud computing. In the last six years, scholars have examined the challenges related with load balancing by comparing the proposed algorithms. Several ways have been offered, but several concerns in the cloud environment, such as VM migration and fault tolerance issues, have yet to be fully solved. To ensure service quality, almost all cloud platforms use dynamic load balancing algorithms at the moment. In the future, cloud platforms may use nature inspired algorithms because it is considered as the accurate method for load balancing.

VII. REFERENCES

1. Load balancing techniques in cloud computing environment:review
https://www.sciencedirect.com/science/article/pii/S131915782100046X?dgcid=rss_sd_all
2. Load Balancing in Cloud Computing
https://www.geeksforgeeks.org/load-balancing-in-cloud-computing/
3. Load balancing in cloud computing – A hierarchical taxonomical classification
https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-019-0146-7

4. Different types of cloud load balancing and algorithms
https://searchcloudcomputing.techtarget.com/answer/What-are-the-different-types-of-cloud-load-balancing
5. What is Cloud Load Balancing
https://www.nginx.com/resources/glossary/cloud-load-balancing/