Volume 3, Issue 7, July-2016, pp. 371-374 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

# A Survey on Load Balancing Algorithms in Cloud

T.Deepa[1], S Sharon Amulya Joshi[2]
[1] Assistant Professor, CSE Department
Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh.
[2] Assistant Professor, CSE Department
Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh.

**Abstract**: - In the cloud computing environment, load balancing is one of the important issues, with great increase in the users and their requirements for different services on the cloud computing platform, efficient usage of resources in the cloud environment is very important. An efficient load balancing algorithm ensures efficient resource utilization by providing resources to user's on-demand in pay-per-use manner. Load Balancing uses scheduling for prioritizing users. The performance indicators of load balancing algorithms in cloud are response time and waiting time. This paper presents various load balancing schemes in cloud environments.

**Key words**: cloud computing, Load balancing, Load balancing Algorithms.

## 1. INTRODUCTION

Cloud Computing is Internet-based computing, where resources, software and information are shared to computers and other devices on-demand. Cloud computing is one of the popular technologies adopted by both industry and academia, which provides a flexible and efficient method to store and access files.

Cloud provides a facility for users to access information at any time and from anywhere. It is not required for the user to be in the same location as the hardware that stores data. When the user has the internet connection they can access the services of the cloud. It delivers all the services dynamically through the internet according to the user requirements.

Cloud computing environment consists of two components: the front end and the back end connected through a virtual network or the internet. The front end is visible to the user(client). It consists of interfaces and applications that are required to access the cloud computing platforms.

- **Mrs. T. Deepa** currently working as assistant Professor in Computer Science and Engineering Department in Ravindra college of Engineering for Women, Kurnool. *E-mail: deepayasoda@gmail.com*
- **Miss. S. Sharon Amulya Joshi** currently working as assistant Professor in Computer Science and Engineering Department in Ravindra college of Engineering for Women, Kurnool. *E-mail: sharonsrigiri@gmail.com*

The back end represents a service provider. It consists of interfaces and applications that are required to access the cloud computing platforms.

Cloud services are categorized into these types.

a. Infrastructure as a Service (IaaS): It provides access to computing resources in a virtualized environment. Users use the fundamental computing resources like processing, storage, networking etc. Example: Google Compute Engine, Windows Azure.

b. Platform as a Service (PaaS): Users can hire programming and infrastructure tools provided by the vendors to develop and run the applications.
Examples: Google App Engine, Windows Azure

c. Software as a Service (SaaS): It is a software distribution model in which Users can use the software provided by the vendors.
Examples: Google Apps, Microsoft Office 365.

The deployment model defines the type of access provided to the cloud user. Cloud provides four deployment models:

a. Public Cloud: Public Cloud allows systems and services to be easily accessible to general public. The IT giants such as Google, Amazon and Microsoft offer cloud services via Internet.

b. Private Cloud: Private Cloud allows systems and services to be accessible within an organization. The Private Cloud is operated only within a single

 

organization.

c. Community Cloud: Community cloud allows system and services to be accessible by group of organizations. It shares the infrastructure between several organizations from a specific community.

d. Hybrid Cloud: Hybrid Cloud is a mixture of public and private cloud. Non-critical activities are performed using public cloud while the critical activities are performed using private cloud.

## 2. LOAD BALANCING

Load balancing is the process of assigning the total load to the individual nodes of the distributed system to do the work faster and efficient utilization of the resources by avoiding a situation where some of the nodes are heavily loaded while other nodes are doing little work or being idle. Load balancing makes sure that all the nodes in the network will perform approximately equal amount of work. The process of load balancing is given in figure1.A load balancer receives tasks from different clients and it takes a decision to transfer the job to the remote server for load balancing.

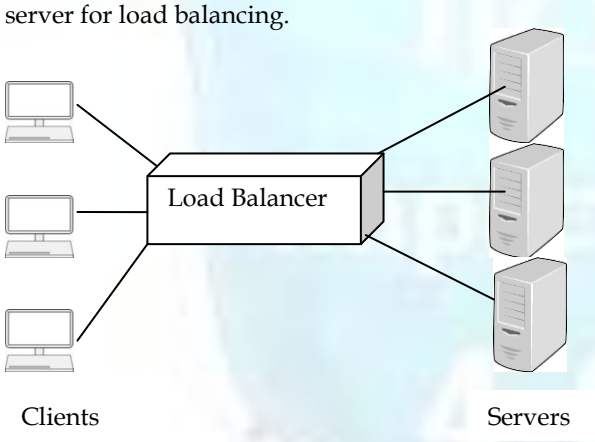


Figure1. Process of load balancing

### 2.1 Need of Load Balancing in Cloud Computing

The main concentration of load balancing in the cloud environment is distributing the load dynamically among the nodes in order to achieve maximum resource utilization and make sure that all the nodes will have the approximate load. An ideal load balancing algorithm helps in making use of the available resources most efficiently, by ensuring no node is over loaded or under loaded. Goals of load balancing involve [3]:

1. Optimum resource utilization
2. Maximum throughput
3. Maximum response time
4. Avoiding overload

### 2.2 Metrics for Load balancing

Various parameters are considered in the existing load balancing algorithms

a. Resource Utilization: This parameter is used to check the utilization of resources. An efficient load balancing algorithm the resource utilization must be optimum.

b. Performance: An efficient load balancing algorithm must have high performance.

c. Scalability: The ability of an algorithm to balance the load of the system with a finite number of nodes. This metric should be improved.

d. Throughput: It represents the number of tasks whose execution is completed. For better performance throughput must be high.

e. Response time: It is the amount of time a particular load balancing algorithm takes to respond in a distributed system.

f. Overhead associated: It determines the overhead involved while implementing a load balancing algorithm. The movement cost, inter process communication are the causes for overhead.

## 3. EXISTING LOAD BALANCING ALGORITHMS

There are many methods [1],[3],[4],[5] to balance the load in cloud computing are available. Some of them are presented in this paper. The algorithms are divided into two types based on the current system state and base on who initiated the process as shown in figure 2.

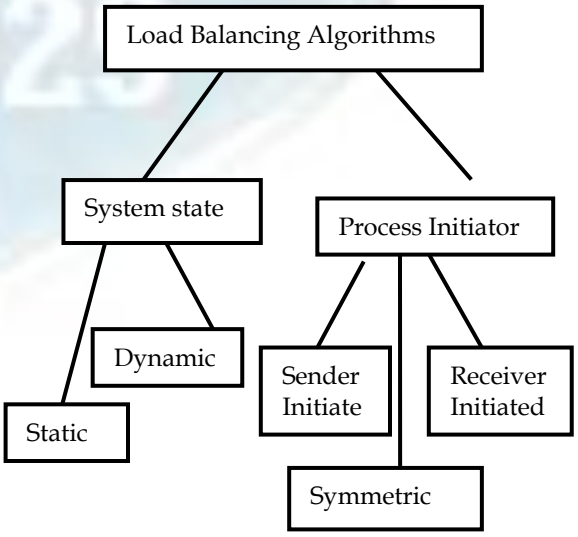


Figure2: Classification of load balancing algorithms

Based on the initiator of the process load balancing algorithms are further classified in to three types.

 **DOI 05.2016-75251336/IJCERT.2016.3703** 

a. Sender Initiated: In this sender identifies that nodes are over loaded, so it initiates the execution of load balancing algorithm

b. Receiver Initiated: If any imbalances in load is identified by the receiver/server in cloud then the server initiates the execution of load balancing algorithm.

c. Symmetric: It is a combination of both sender initiated and receiver initiated algorithms.

Based on the current state of the system load balancing algorithms are classified into static and dynamic algorithms

**3.1 Static Algorithms:** The static algorithm does not consider the current state of the node. All the nodes and their properties are known in advance, the algorithm works based on this previous information.

a. Round robin algorithm[6]: It uses the principle of time slices. Here time is divided into slices and each node is given a time interval, in which the node has to perform the task. If the processing is not completed within the time quantum, it has to wait for the next slot.

b. Min-Min algorithm [2]: In this algorithm first the minimum completion time of all the nodes is calculated. A task with minimum completion time is selected and assigned to the corresponding node. This process is repeated until all the tasks are assigned to the nodes.

c. Max-Min algorithm [2]: It a static algorithm, in this first minimum completion time of all the tasks is calculated and a task with maximum completion time is selected and it is assigned to the corresponding node.

d. Opportunistic Load Balancing algorithm [4]: This algorithm does not calculate the execution time and the current load of the node. It assigns the tasks randomly to the nodes. The task processing takes long time because it does not calculate the execution time of the node.

e. The two phase scheduling load balancing algorithm [4]: It is a combination of opportunistic load balancing (OLB) and Load Balance min- min algorithms to provide better execution efficiency and to balance the load. OLB scheduling makes every node in working state to balance the load and LBMM minimizes the execution time of each task.

**3.2 Dynamic algorithms:** Dynamic algorithm depends on the current state of system. The algorithm works based on the changes in the state of the nodes dynamically. At any time if a node is having heavy load it is transferred to a node with light load.

a. Ant Colony optimization technique [7]: In this algorithm, when a request is initiated the ant starts its movement in forward direction visiting the nodes one by one and checking whether a node is overloaded or under loaded and records the data. If the ant finds an overloaded node it starts backward movement to the previous under loaded node to share data.

b. Honey Bee Foraging Algorithm [3]: It is a nature inspired decentralized load balancing method that helps to balance the load across heterogeneous nodes of the cloud. In this algorithm first current load of the nodes is calculated then it decides whether the node is over loaded, under loaded or balanced. A task from the heavy loaded node is removed and by considering its priority it is assigned to a lightly loaded node.

c. Biased Random Sampling Algorithm [3]: In this algorithm the network is represented as a virtual graph. The servers are represented as nodes and the in-degree represents the free resources available to the node. On the basis of in-degree the load balancer assigns the tasks to the node. When a task is assigned the in-degree is decremented and it is incremented when the job gets executed.

d. Resource Allocation Scheduling Algorithm (RASA) [3]: In this algorithm first virtual nodes are created. The expected response time is calculated for all the virtual nodes. According to least loaded node criteria, efficient virtual node is found. If the number of resources are odd then Min-Min strategy is applied, otherwise Max-Min strategy is applied.

## 4. CONCLUSION

Cloud computing provides many services to the user over the network. Load balancing is the major issue in cloud. Load balancing is required to distribute the dynamic load evenly to all the nodes. Overloading of the system will lead to poor performance. For efficient resource utilization an efficient load balancing algorithm is required. We discussed many metrics for an efficient load balancing algorithm. In this paper we surveyed multiple load balancing algorithms which are already proposed by various researchers.

 **DOI 05.2016-75251336/IJCERT.2016.3703** 

## REFERENCES


1. [1] Rajwinder Kaur,and Pawan Luthra "Load Balancing in Cloud Computing" Association of Computer Electronics and Electrical Engineers, 2014.
2. [2] "An in-depth analysis and study of Load balancing techniques in the cloud computing environment" 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
3. [3] Martin Randles ; Sch. of Comput. & Math. Sci., Liverpool John Moores Univ., Liverpool, UK ; David Lamb ; "A. Taleb-BendiabA Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing "
4. [4] "A Comparative Study of Load Balancing Algorithms in Cloud Computing", International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 24, May 2015
5. [5] Anand Chaudhari "Load Balancing Algorithm for Azure Virtualization with Specialized VM's" International Journal of Innovations in Engineering and Technology (IJIET) Vol. 2 Issue 3 June 2013
6. [6]"Efficient and Enhanced Algorithm in Cloud Computing" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013
7. [7] "Load Balancing of Nodes in Cloud Using Ant Colony Optimization". In proc. 14th International Conference on Computer Modeling and Simulation (UKSim), IEEE, pp: 3-8, March 2012.
8. [8] "Static Load Balancing Algorithms In Cloud Computing: Challenges & Solutions" International Journal Of Scientific & Technology Research Volume 4, Issue 10, October 2015


 **DOI 05.2016-75251336/IJCERT.2016.3703** 