Abstract

Cloud Computing is a methodology for distributing and accessing applications across the network. The various parameters considered in Cloud are: fault tolerance, high availability, scalability, and flexibility, reduced overhead for users, reduced cost of ownership, on demand services. The set of rules and policies that control the order by which various jobs are executed in a system form the basis for scheduling. Load balancing algorithms attempts to improve the response time of the user's submitted applications by ensuring maximum utilization of available resources. Load balancing deals with the way by which the various tasks are assigned to the resources thereby improving the system performance. Scheduling of various tasks to the resources in a Cloud environment is an active research area. Resources are dynamic in nature so the load of resources varies with change in configuration of Cloud and hence Load balancing of the tasks in a Cloud environment can significantly influence the Cloud’s performance. The hierarchical load balancing concept uses the tree data structure to make decision regarding the placement of tasks on Virtual Machine. In order to utilize the resources efficiently and to satisfy the QoS requirement of the users, several hierarchical load balancing algorithms have been
proposed by researchers for various applications. This paper deals with the overview of the load balancing concepts in Cloud with an assessment of the various hierarchical load balancing algorithms.

References

3. Han Xiangchun, Chen Duanjun, Chen Jing, "one centralized scheduling pattern for dynamic load balance in Grid", 2009 international forum on Information Technology and Applications.
14. V. P. Narkhede, Prof. S. T. Khandare, “Fair Scheduling Algorithm with Dynamic Load


**Index Terms**

Computer Science

Distributed Computing

**Keywords**

Hierarchical; Virtual Machine(VM); OLB; Load Balancer, LBMM