International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 02 | Feb-2018 www.irjet.net p-ISSN: 2395-0072

A Review on different load balancing Algorithm in cloud computing

Pooja R. Kathalkar[1], A.V. Deorankar[2]

[1]*M. Tech. Scholar Department of Computer Science and Engineering Government College of Engineering Amravati (MH) India*

[2]*Head Department of Information Technology Government College of Engineering Amravati (MH) India*

Abstract - Load balancing in the cloud computing environment has an important impact on the performance of cloud . Good load balancing makes cloud computing more efficient and improves user satisfaction. Efficient task scheduling and resource management is a challenging problem of distributed computing but it is still in its infant stage in spite of exhaustive research in recent years. Genetic algorithms can be use to enhance the performance of load balancing approaches. Conventional scheduling algorithms such as Round Robin, First Come First Serve, Ant Colony Optimization etc. have been widely used in many cloud computing systems. Cloud receives clients tasks in a rapid rate and allocation of resources to these tasks should be handled in an intelligent manner. Aside the prominent cloud issues in the areas of scheduling, resources allocation and security, cloud computing now highlights additional pressing issues seeking for attention in the area of fault tolerance in executing tasks and also virtual machines (VM) failure. These types of problems in a broad outline are called NP-hard (non-deterministic polynomial time), which means that there is no exact solution and no quick solution to it.

Key words: Bio-Inspired Algorithms, Load Balancing, Resource Management, Task Scheduling, Virtual Machines, Fault Tolerance.

I. INTRODUCTION

Cloud computing is the emerging technology in distributed environment consisting of several data centers, servers, virtual machines, load balancers etc. which are connected intelligently. Further, the cloud deals with many things such as storing and retrieving of documents, sharing of multimedia, lending the related resources on pay-as-you go model and much more. Even though there is much advancement in the era of computers and Internet of Things (IoT) with respect to responsiveness, reliability and flexibility, still there is a room for improvement in scheduling, optimal resource allocation and management algorithms since these algorithms come under NP-hard and NP-complete complexity classes. Hence, there is a need to address these set of challenging problems using different techniques. Efficient task scheduling and resource management is a challenging problem of distributed computing but it is still in its infant stage in spite of exhaustive research in recent years.

The objectives of the study of the load balancing techniques are as follow:

- Improve both resource deployment and job response time while also avoiding a situation where some of the nodes are having a huge amount of load while other nodes are doing nothing or idle.
- To reduce operational cost, better performance in terms of response time and data processing time, maintain the system consistency.

Load balancer supports multiple load balancing algorithms. five common load balancing algorithms which are use in load balancing mechanism are as follow :

1. ROUND ROBIN:

Round Robin is a very famous load balancing algorithm, in which the processes are divided between all processors. The process allocation order is maintained locally independent of the allocations from remote processors. In round robin fixed quantum time is given to the job. Main emphasis in round robin is on fairness and time limitation.

2. WEIGHTED ROUND ROBIN

Weighted Round Robin (WRR) scheduling is used to facilitate controlled sharing of the network bandwidth. WRR assigns a weight to each queue; that value is then used to determine the amount of bandwidth allocated to the queue. The round robin aspect of the scheduling allows each queue to be serviced in a set order, sending a limited amount of data before moving onto the next queue and cycling back to the highest-priority queue after the lowest-priority queue is serviced.

3. LEAST CONNECTIONS

The least-connection scheduling algorithm directs network connections to the server with the least number of established connections. This is one of the dynamic scheduling algorithms; because it needs to count the number of connections for each server dynamically to estimate its load. The load balancer records the connection number of each server, increases the connection number of a server when a new connection is dispatched to it, and decrease the connection number of a server when a connection finishes or timeouts.