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.

4. WEIGHTED LEAST CONNECTIONS

The Weighted Least Connections algorithm does to Least Connections what Weighted Round Robin does to Round Robin. That is, it introduces a "weight" component based on the respective capacities of each server. Just like in the Weighted Round Robin, you'll have to specify each server's "weight" beforehand.

5. RANDOM

This algorithm matches clients and servers by random, i.e. using an underlying random number generator. In cases where in the load balancer receives a large number of requests, a Random algorithm will be able to distribute the requests evenly to the nodes. So like Round Robin, the Random algorithm is sufficient for clusters consisting of nodes with similar configurations

II. LITERATURE REVIEW

From the overall study it is seen that the above topic put emphasis on some important methods or techniques. Dynamic algorithms are well suited in cloud computing environment because they distribute work at run time and assign suitable weights to the servers. A lightest weight server is search in network and preferred by this algorithm . Now a day's Bio-inspired dynamic load balancing algorithms are widely getting importance in load balancing techniques.

1) Ant colony algorithm

Different ant colony algorithms also introduce to balance the load applying ant behaviour for searching food.[10]Larger weight means that resource has high computation power. Load balancing ant colony optimization (LBACO) not only balance the load but also minimizes make span. All tasks are assumed to be mutually independent and computationally intensive.

2) Honey bee foraging algorithm

This algorithm is basedforaging behaviour of honey bees. When an under loaded VM assigns a task, it updates number of priority tasks and load of VM to other tasks in waiting list. This approach helps other processes to choose their VM [59]. If a task has high priority, then it selects a VM having minimum number of priority tasks. It does not take into consideration only load balancing but also keeps track of priorities of tasks which currently removed from heavy loaded machines. It increases throughput and minimizes response time.[9]

3) Throttled load balancing

This algorithm depends upon the theory of suitable search of virtual machine.[8] The task manager makes a list of virtual machines. By using the list, client request allotted to the relevant machine. If the size and capability of the machine is suitable for request, then the job is given to that machine. This algorithm is better than round robin algorithm.

4) Pareto based fruit fly optimization algorithm

A Pareto based fruit fly optimization algorithm (PFOA) is use to solve the task scheduling and resource allocating (TSRA) problem in cloud computing environment.[2] First, a heuristic based on the property of minimum cost initialize the population. Second, a resource reassign operator is used to generate non dominated solutions. Third, a critical path based search operator is used to improve the exploitation capability.

5) Multi objective Scheduling cuckoo algorithm

CSA mimics the breeding behavior of cuckoos, where each individual searches the most suitable nest to lay an egg (compromise solution) in order to maximize the egg's survival rate and achieve the best habitat society.[4] Fuzzy set theory is used to create the fuzzy membership search domain where it consists of all possible compromise solutions. CSA searches the best compromise solution within the fuzzy search domain simultaneously tuning the fuzzy design boundary variables. Tuning of fuzzy design variables eliminate the requirement of expertise needed for setting these variables.

6) Min-Min algorithm in cloud environment.

Load Bbalancing Min Min algorithm has a three level load balancing framework.[11] In first level LBMM architecture is the request manager which is responsible for receiving the task and assigning it to service manager, when the service manager receives the request; it divides it into subtask and assigns the subtask to a service node based on node availability, remaining memory and the transmission rate which is responsible for execution the task.

III. CONCLUSION

In this paper we have done survey on different load balancing techniques use in load balancing mechanisms in cloud computing and different bio-inspired load balancing algorithms proposed by various authors.

REFERENCES

[1] D. Huang, C. Zhu, H. Zhang, and X. Liu, "Resource intensity aware job scheduling in a distributed cloud," China Communications, vol. 11, no. 14, pp. 175–184.

[2] X.-l. Zheng and L. Wang, "A pareto based fruit fly optimization algorithm for task scheduling and resource allocation in cloud computing environment," in Evolutionary Computation (CEC), 2016 IEEE Congress on. IEEE, 2016, pp. 3393–3400.

[3] M. Rasti-Barzoki and S. R. Hejazi, "Pseudo-polynomial dynamic programming for an integrated due date assignment, resource allocation, production, and distribution scheduling model in supply chain scheduling," Applied Mathematical Modelling, vol. 39, no. 12, pp. 3280–3289, 2015.

[4] M. Akbari and H. Rashidi, "A multi-objectives scheduling algorithm based on cuckoo optimization for task allocation problem at compile time in heterogeneous systems," Expert Systems with Applications, vol. 60, pp. 234–248, 2016.

[5] P. S. Pillai and S. Rao, "Resource allocation in cloud computing using the uncertainty principle of game theory," 2014.

[6] X. Xu, W. Dou, X. Zhang, and J. Chen, "Enreal: an energy-aware resource allocation method for scientific workflow executions in cloud environment," 2015.

[7] S.-C. Wang, K.-Q. Yan, W.-P. Liao, and S.-S. Wang, "Towards a load balancing in a three-level cloud computing network," in Computer Science and Information Technology (ICCSIT),20103rdIEEE International Conference on, vol. 1. IEEE, 2010, pp.

[8] Shridhar Donamal., "Load balancing in cloud computing using modified throttled algorithm", 2013.

[9] Dhinesh Babu, "Honey bee behavior inspired load balancing of tasks in cloud computing environments" ,Volume 13, Issue 5, May 2013

[10] Hajara Idris., "An improved ant colony optimization algorithm with fault tolerance for job scheduling in grid computing systems", May 17, 2017

[11] Huankai Chen, "User priority guided min min scheduling algorithm for load balancing in cloud computing", 2013.