

A SURVEY ON LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Harshada Raut¹, Kumud Wasnik²

¹M.Tech. Student, Dept. of Computer Science and Tech., UMIT, S.N.D.T. Women's University, (India)

²Professor, UMIT, S.N.D.T. Women's University, Mumbai, Maharashtra, (India)

ABSTRACT

Cloud computing is sharing various computing resources rather than having local servers or personal devices to handle applications over internet. There is lots of data stored on cloud and multiple requests come for resources, due to this increase load on data servers and sometimes unable to provide resources on time. So to manage the load by scheduling task appropriately researchers have introduced some load balancing techniques. Job scheduling can be done by various algorithms like FCFS, Round Robin, and Priority Scheduling. In this paper we are reviewing existing static and dynamic load balancing techniques like Min-Min, Max-Min, Ant colony optimization etc.

Keywords: Cloud Computing, Dynamic Load Balancing, Load Balancing, Load Balancing Algorithms, Static Load Balancing, Task Scheduling.

I INTRODUCTION

Recently cloud computing has become most popular field for academic research area. Cloud computing is a central remote servers where a programs or applications run on server machine. Cloud computing is a system where services like IaaS (Infrastructure is provided as a Service), PaaS (Platform is provided as a Service), SaaS (Software is provided as a Service) are provided online using Internet Connection. These services are provided anywhere, anytime on the basis of “only pay for what you use”.



Fig 1: Cloud Computing System

Cloud Computing consists of Virtual machine, Host, Servers. User need not to worry about maintenance of any software, it is the responsibility of cloud service provider to provide the needed resources to user. On the basis of package the user gets the services from service provider, but to use cloud services user needs continuous internet connection. It is the responsibility of cloud Service provider to provide resources to user as per user's demand. Fig.1 shows cloud computing system where cloud services SaaS, PaaS, IaaS are provided to uses.

The cloud appears to be a single point of access for all the computing needs of consumers. It is difficult to manually assign tasks to computing resources in clouds because hundreds of thousands of virtual machines (VMs) are used [1]. Service provider should provide all cloud services to user efficiently, but sometimes due to too many requests from users to utilise resources machine may get overloaded. In results, it is unable to provide resources to users and performance falls down. To avoid this, a good scheduling algorithm is required. Scheduling algorithm helps to provide resources to all users, use all machines and make sure there should not be under loading or overloading machines.

There are 4 Types of Cloud Deployment Models:

➤ *Private Cloud*

This model is build where single organisation wants its own cloud infrastructure for private use. This may be managed by internally or by a third Party.

➤ *Public Cloud*

In this type of model a cloud infrastructure is provided to all users with less access restrictions. Service is easily available and free to all users so it is less secure.

➤ *Community Cloud*

This infrastructure is used by some specific organizations which are in the same group. This infrastructure may be managed by internally or by a third party.

➤ *Hybrid Cloud*

Hybrid is consisting of more than 1 cloud which may be combination of Private, Public or Community Cloud. These clouds would typically be created by the enterprise and management responsibilities would be split between the enterprise and cloud provider [2].

II LOAD BALANCING IN CLOUD COMPUTING

In cloud computing load balancing is a major issue. As the time passes number of users may get increase and this will also increase the load on server and decrease the performance of functions provided by cloud service provider. If there are multiple requests at a same time from different users for resource utilization then it create a long queue on server and increases a response time and to avoid this, divide the task among all available machines and balance a system. If resources are less compared to requests then it may delay the response to clients and server may not be able to fulfil all demands of clients.

Load Balancing is dividing total amount of work in all available machines for execution. To balance a load, each task should be scheduled very carefully so every user may get service in less waiting time. There are various algorithms used for load balancing like Round robin, Ant colony optimization, Particle Swarm Optimization, Max-Min, Min-Min etc.

The aim of load balancing is as follows:

- Quick response time from service provider
- The availability of resources in less waiting time
- No system failure for large task
- Provide the service to every user
- Improve performance
- Accommodate future modification[3]

Fig 2. shows a framework of under which load balancing algorithms work in a cloud computing environment. Cloudlet submits tasks to Job Manager and then Job Manager gives all jobs to Load Balancer. Load Balancer applies Load Balancing Algorithm for Submitted Tasks and schedule all task such that each Virtual Machine will get equal number of task for execution.

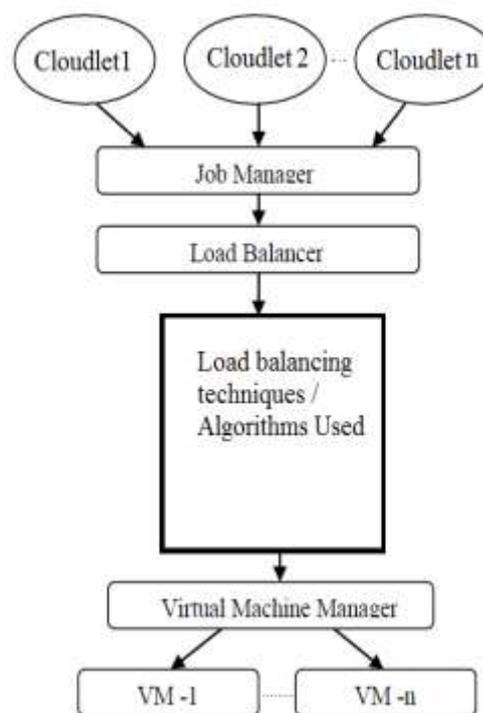


Fig 2: Load Balancing Algorithms Execution [4]

2.1 Load balancing can be done by two types

2.1.1 Static load balancing

Load is balanced by dividing the tasks before starting of execution in machine. In static load balancing, Load dividing decision does not depend upon the current state of machine.

Round Robin, Opportunistic Load Balancing, Min-Min, and Max-Min are the static load balancing algorithms.

2.1.2 Dynamic load balancing

Load is balanced by considering the current state of the machine and then decision is made. It checks the network and assign task to machine which is underloaded or ideal.

Ant Colony Optimization, Honeybee Foraging, Biased Random Sampling, Active Clustering are the Dynamic load balancing algorithms.

II EXISTING LOAD BALANCING ALGORITHMS

In this section we discuss various load balancing algorithms which are already implemented and are available for load balancing.

Ant colony optimization (ACO) algorithm is proposed by Marco Dorigo in 1992. It is inspired from real ants, when searching for a food ant travels randomly and in return trip they deposited some chemical pheromone. On the quantity of this pheromone other ants uses shortest path on which more pheromone value is deposited. This algorithm helps to find a minimum distance from source to destination. In [5] enhancement to ACO algorithm is proposed. Artificial ants moves forward and backward direction to find the overloaded node and update value in pheromone table. Throughput, Response time, less energy consumption is achieved, but it gives less performance.

Honeybee Foraging Behaviour Algorithm is dynamic approach. [6]Baris Yuce et. al proposed enhanced version of Bees algorithm called ANSSA. It is a decentralized honeybee-based load balancing technique. It is a nature-inspired algorithm for self-organization. In this algorithm sends bees to selected sites and evaluates a value which is best suited for that site. It gives better result for higher dimensions in local server actions. System Performance is improved with increased system diversity but throughput is not increased with an increase in system size. It is suitable for the conditions where the diverse population of service types is required.

Round Robin is a static behaviour algorithm. It uses time quantum and allocate jobs in first come first serve manner, after allocating jobs to all machine sequentially, it return to first machine and allocate next job if that machine is finishes its time quantum, jobs are not allocated priority wise. In paper [7], Nusrat Pasha et.al proposed enhancement to round robin algorithm. In this, process does not wait for total consumption of resources it move to next service. This achieves resource utilization and better response time. Sometimes more load on server if more processes trying to switch in short time.

Min Min[8] is a static load balancing algorithm. This algorithm selects job with minimum execution time and assigned it to machine which can executes that job in minimum time. This algorithm is suited where small task are more compared to large task as it executes smaller task first and gives less priority to large task in execution. More waiting time for large size task.

Max Min is static load balancing algorithm, this algorithm is overcomes the disadvantage of Min-Min algorithm. In this, select jobs with maximum execution time and assigned it to machine with min completion time [9]. In other words, larger jobs run on slower machine. The enhancement[10] to this Max-min Algorithm is instead of selecting maximum execution time task, selects an Average or nearest greater than average task then overall makespan is reduced and also balance load across resources.

III COMPARISON TABLE OF EXISTING LOAD BALANCING ALGORITHMS

Algorithms works in Environment	Algorithms	Resource Utilization	Throughput	Response Time	Waiting Time
Dynamic	ACO	Yes	Yes	Yes	No
Dynamic	Honeybee Foraging	No	Yes	No	No
Static	Round Robin	Yes	Yes	Yes	No
Static	Min Min	Yes	Yes	Yes	Yes
Static	Max Min	Yes	Yes	Yes	No

IV CONCLUSION

In this paper, we discussed and compared some load balancing algorithms. To balance a load on vast cloud system a proper scheduling of task is required. For satisfactory service a user needs flexibility in service, easy availability, less waiting time to get resource etc. Incoming jobs can be scheduled according to virtual machine's capacity so that it may avoid the unavailability of service if there are multiple requests on single machine and this will get satisfactory service to cloud users. There is a necessity of permanent solution for load balancing in cloud computing. All algorithms discussed here have some advantages and disadvantages.

REFERENCES

- [1] Qiyi, H., Tinglei, H., "An Optimistic Job Scheduling Strategy based on QoS for Cloud Computing", IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS), pp.673-675, 2010.
- [2] Rabi Prasad Padhy, Manas Ranjan Patra, Suresh Chandra Satapathy, "Cloud Computing: Security Issues and Research Challenges", *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, Vol. 1, No. 2, December 2011
- [3] Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing", *International Journal of Scientific and Technology Research*, Volume 3, Issue 11, 2014
- [4] N. S. Raghava, Deepti Singh "Comparative Study on Load Balancing Techniques in Cloud Computing", *Open Journal of Mobile Computing and Cloud Computing*, Volume 1, Number 1, August 2014.
- [5] Shagufta Khan, Nireesh Sharma, "Effective Scheduling Algorithm for Load balancing (SALB) using Ant Colony Optimization in Cloud Computing", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 4, Issue 2, February 2014
- [6] Baris Yuce, Michael S. Packianather, Ernesto Mastrocinque, Duc Truong Pham and Alfredo Lambiase "Honey Bees Inspired Optimization Method: The Bees Algorithm", *Insects* 1 July 2013; Published: 6 November 2013.

- [7] Pooja Samal, Pranati Mishra, "Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing", *International Journal of Computer Science and Information Technologies*, Vol. 4 (3) , 2013, 416-419
- [8] T. Kokilavani, Dr. D. I. George Amalarethnam, "Load Balanced Min-Min Algorithm for Static Meta Task Scheduling in Grid computing", *International Journal of Computer Applications Vol-20 No.2*, 2011
- [9] Santhosh, Dr. Manjaiah D.H, "An Improved Task Scheduling Algorithm based on Max-min for Cloud Computing", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2, Special Issue 2, May 2014
- [10] Upendra Bhoi, Purvi N. Ramanuj, "Enhanced Max-min Task Scheduling Algorithm in Cloud Computing", *International Journal of Application or Innovation in Engineering & Management (IJAIEEM)*, Volume 2, Issue 4, April 2013.