

A Comparative Study on Load Balancing Algorithms in Cloud Computing

Seema Nagar ^[1], Ajeet KR Bhartee ^[2]

M.Tech ^[1], Assistant Professor ^[2]

Department of Computer Science and Engineering
Galgotias College of Engineering & Technology
Greater Noida
UP-India

ABSTRACT

Cloud computing is fast growing domain where new advancements are being done every day. Computing services are provided to users via internet. There are mainly three types of services provided over the cloud i.e. Software as a service, Platform as a service and Infrastructure as a service. The service providers will manage all the aspects of cloud services related to implementation, installation, configuration, troubleshooting and the various other issues related to services provided. The cloud service providers will provide the services as per the usage and will charge only for the amount of service used. Cloud computing thus provides flexibility to users on the amount of service to be used and when to use. In this paper we will discuss various types of load balancing algorithm in cloud computing, and what are the various options available related to simulation of the various load balancing algorithms.

Keywords:- Cloud, Load balancing, Virtual machine, Cloud services, Distributed, Simulators, CloudSim

I. INTRODUCTION

The cloud computing is an internet based computing mainly involves using web browser on your computer to access the services over the internet. The actual installation and processing is done by remote servers that can be located on some distant computer. There are generally three categories of services which are provided by the cloud namely software as a service for eg. Google docs, platform as a service for eg. Development platform can be used a service on which software developers can develop their code, infrastructure as a service for eg. Servers can be used as a service.

Cloud computing basically consists of three entities i.e. Clients, Distributed servers and Data centres. The clients request services and send requests to the data centres. A Data centre is a collection of the different servers which hosts the applications. A distributed server is a collection of servers which are scattered over the internet and provides access to services. Refer to fig. 1. Now a days there has been a huge demand of the cloud computing services and thus no. of clients accessing the services is increasing, so there is a major concern to balance the load of the distributed servers. Client request to access services is randomly generated hence some nodes might get heavily loaded while some might be lightly loaded. To properly balance the load on virtual machines, the load balancers are used which migrate some load from heavily

loaded virtual machine to lightly loaded machine so that the load is evenly distributed.

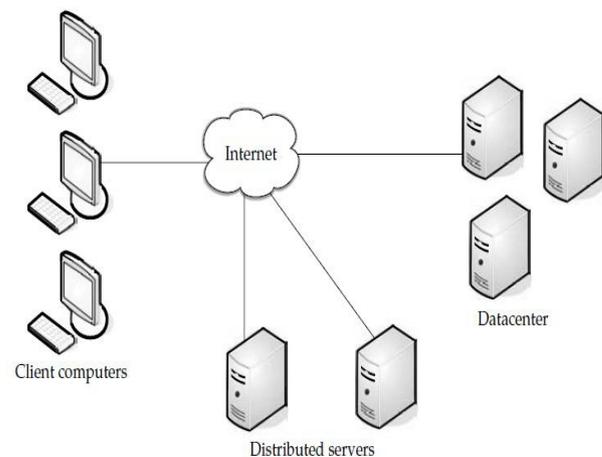


Fig. 1 Cloud components ^[1]

Load balancing algorithms can be broadly categorized in two types: static and dynamic load balancers. A static algorithm uses predefined information of the system to balance the load of the system. Dynamic algorithms use current state of the system to balance the load of the system. Further there are three ways to dynamic load balancing i.e. Global, Cooperative and Non-cooperative. In global method, there is single machine which balances the load of the system. In cooperative approach, few machines work in a cooperative

way to balance the load of the system. In non-cooperative method of load balancing, every machine balances its load by its own.

II. LOAD BALANCING ALGORITHMS

There are different load balancing algorithms which are available to balance the load on different nodes in the system. They evenly distribute the load in the system so that all the resources are effectively utilized. The choice of load balancing algorithm depends on the programmer as well as well the context in which cloud services are being used. Some of them are discussed below.

A. Round Robin Load Balancer

It is a static algorithm which assigns the client requests to the virtual machines as per the round robin scheduling. A static algorithm is low in flexibility as the algorithm does not consider the current load condition parameters of the system. In round robin the first request is assigned to the vm which is selected randomly and later the requests are assigned to the virtual machines on a circular order where each virtual machine is given a fixed quantum of time.

B. First Come First Serve

It is one of the simplest algorithms used for task scheduling. The implementation is done with the help of FCFS queue. When a job enters the system it will be placed in the queue. From the queue jobs are assigned to the virtual machines in the order in which the jobs arrived i.e. jobs are scheduled on the basis of their arrival time. The job which arrives first will be allotted the virtual machine. FCFS is easy in implementation but as it has large average waiting time hence the throughput is less for this algorithm.

C. Throttled Load Balancer

This algorithm is based on virtual machine. It works by finding the most suitable virtual machine for a job and then assigns the job to the virtual machine. The load balancer will check if a suitable virtual machine is available to fulfil the job request. In case a virtual machine is found which matches the criteria, and then the job is assigned to the virtual machine, if there is no virtual machine which is available, it will wait for the virtual machine to be free. When the vm is set to free, the job is assigned to it. Refer to fig.2.

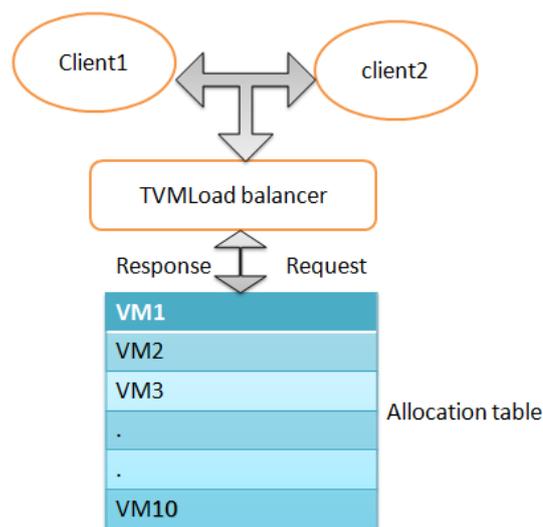


Fig.2 Throttled Algorithm^[2]

D. ESCEL Algorithm

This algorithm initially assigns the task randomly to the virtual machines which are currently available. It is based on the concept of current execution load. This algorithm calculates the current load on each virtual machine and then distributes the load from heavily loaded virtual machine to the lightly loaded virtual machine. It is dynamic in nature as it considers the current load condition of each virtual machine.

E. Active Clustering

This is based on the concept of clustering in cloud computing. In clustering we group the data as per certain predefined parameters. In the similar way we can group the nodes in the cloud based upon certain criteria such as resource types etc. Whenever there is an incoming request, it will be first checked that to which group it belongs and once it is found, the request will be assigned to the node identified. This will reduce the response time of the jobs as now the searching of node will take less time.

F. Honeybee Algorithm

This algorithm is based upon the behaviour of honeybees. It is used to balance the load on the system, where some of the nodes may be lightly loaded and some may be overloaded. The algorithm requires that node to maintain its separate queue. For each incoming request each node will first need to compute the profit which causes the additional overhead. Due to the computation overhead involved in maintaining a separate queue, this algorithm does not show much significant improvement in the throughput.

G. Biased Random Sampling

It is a dynamic load balancing algorithm in which a virtual graph is constructed, where each vertex represents the node in the directed graph. Every node has certain number of free nodes which are represented by the in-degree. Whenever a new job is assigned to a node, its in-degree is decremented by one. Whenever a job is completed, the in-degree of that job is increased by one.

H. Particle Swarm Optimization

This algorithm is inspired from the behaviour of flocking of birds. To understand assume there is a group of birds which is searching for food in an area and there is only a single piece of food in that area. All the birds are not knowing the location of food. In such a case the best approach is to trace the bird which is nearest to the food and updating the position and velocity of each bird accordingly. PSO generates many simultaneous candidate solutions. During each iteration of the algorithm each candidate solution is checked and updated. PSO can also be applied to task scheduling in cloud computing. It provides the good performance in variety of scenarios. PSO algorithm has many limitations such as the rate of convergence of PSO is low in case of solving large scale problems related to optimization and PSO easily moves into the defects of local optima due to the high randomness involved in it. This convergence issue can be addressed by applying PSO with simulated annealing algorithm and it eventually increases its convergence rate.

III. SIMULATORS

The simulator is widely used to carry out the implementation for the proposed work by the researcher. In many cases it is not feasible to test the outcome of the work done in a real environment. In such cases a simulator is used to test the implementation work done. It provides many in-built features to test the results by slight variation in parameters. In this section we shall be discussing some very useful simulators used for a cloud environment.

A. Cloud SIM

CloudSim goal is widely used for performing simulations in a cloud. In case of a cloud to access the infrastructures you need to pay some amount to the service provider hence using a simulator is easier as it allows users to perform repeated simulations with change in different parameters, free of cost for the application, infrastructure and platform services.

There are several features of CloudSim such as: support for modelling and instantiation of large scale cloud computing infrastructure including data centres on a single physical computing node and Java virtual machine; a self-contained platform for modelling data centres, service brokers, scheduling and allocation policies, availability of virtualization engine, which aids in creation and management of multiple, independent and co-hosted virtualized services on a data center node; and flexibility to switch between space-shared and time shared allocation of processing cores to virtualized services.^[3]

B. Green Cloud Simulator

This is a simulator which is used for developing solutions for monitoring, allocation of resources, scheduling of the workload on data centres in a cloud. It efficiently provides modelling of the energy consumed by the data centres and IT related equipments such as servers, routers, network related equipments etc. It supports independent energy models for each type of resource. It has a user friendly GUI interface and is an open source tool.

C. Network Cloud

Network Cloud is an extension of CloudSim and is capable of implementing a network layer in CloudSim, reads a BRUTE file and generates a topological network. Here, we have a topology file which contains the number of nodes along with the various entities involved in simulation. In this simulation tool, each entity is to be mapped to a single BRUTE node so that network CloudSim can work properly. Network CloudSim can be used to stimulate network traffic in CloudSim^[4].

D. Virtual Cloud

It is a simulator to perform the functionalities of a cloud in a simulated environment. Virtual cloud provides the researchers to model and test their policies and to utilize the cloud computing resources efficiently. It has a multilayered architecture and it helps to test new approaches developed and find out where the improvement can be done^[3].

E. SPECI

SPECI, Simulation Program for Elastic Cloud Infrastructures, is responsible for analyzing the various scalability and performance aspects of future Data centers. It is assumed that when data centers are made to grow big, then they do so in a non-linear fashion, so there is a need to analyze the behaviour of such data centers. To analyze such a scenario SPECI is used^[4, 6].

IV. CONCLUSION

In our paper we have discussed the various load balancing algorithms which are used in cloud computing to evenly distribute the load on the virtual machines in the system. It improves the resource utilization of the nodes and hence the overall system efficiency. We have also discussed some important simulators available for the cloud. Any of the simulators can be used as per the requirement of the work involved and the ease of use of the developer. In future we will carry out analysis of the other simulators for diverse platforms.

REFERENCES

- [1] Pragati Priyadarshinee, Pragya Jain, “Load Balancing and Parallelism in Cloud Computing”, June 2012, *IJEAT*
- [2] Subasish Mohapatra, K.Smruti Rekha and Subhadarshini Mohanty, “A Comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing”, *International Journal of Computer Applications* (0975 – 8887), Volume 68– No.6. April 2013.
- [3] Shridhar G.Domanal and G.Ram Mohana Reddy, “ Load Balancing in Cloud Computing Using Modified Throttled Algorithm”, National Institute of Technology, Karnataka, Suratkal, Mangalore, India.
- [4] Parveen Kumar and Anjandeeep Kaur Rai, “An Overview and Survey of Various Cloud Simulation Tools”, *Journal of Global Research in Computer Science*, Volume 5, No. 1, January 2014.
- [5] Malhotra Rahul, Jain Prince, “Study and Comparison of CloudSim Simulators in the Cloud Computing”, *The SIJ Transactions on Computer Science Engineering & its Applications* (CSEA), Vol. 1, No. 4, September-October 2013, pp.111-115
- [6] Sriran Ilango, “SPECI, a simulation tool exploring cloud-scale data centres”
- [7] Solmaz Abdi, Seyyed Ahmad Motamedi, and Saeed Sharifian, “Task Scheduling using Modified PSO Algorithm in Cloud Computing Environmen” , *ICMLEME*, 2014, Jan. 8-9, 2014 Dubai (UAE).
- [8] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi, “A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms”, 2012, *IEEE*.
- [9] Shanti Swaroop Moharana, Rajadeepan D.Ramesh And Digamber Powar, “Analysis Of Load Balancers In Cloud Computing”, May 2013, *Ijcse*.