Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

A Comparative Study on Load Balancing Algorithms in Cloud Computing

A.JoiceShakila,
Asst.Professor, Department of Computer Science, Sanghamam College of Arts and Science, Annanamangalam, Tamil Nadu, India

Abstract-- Cloud computing has revolutionized the way in which computer resources and services were delivered over the internet. It represents the ability to shift loads rapidly across the internet. Load Balancing in Cloud Computing has been gaining a great deal of attention and so many researches are being done because good balancing enhances the performance and user satisfaction. Millions of requests to the cloud servers demand a technique to ensure that all the servers are equally loaded. Workload balancing between cloud servers is done by means of hardware and software algorithms. This article makes a survey on the various algorithms used in the public cloud.

Keywords-- *Cloud Computing; Load Balancing; Load Balancer;*

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction[1]. Cloud Provider offers services that can be grouped into three categories such as Software as a Service, Platform as a Service and Infrastructure as a Service. Enterprises can choose to deploy applications on Public, Private or Hybrid clouds.

Though, the cost of computation, application hosting, content storage and delivery is reduced significantly with the advent of cloud computing, we cannot deny that it has got some serious issues like data protection, Data Recovery and availability, Management capabilities and Regularity and Compliance Restrictions which are to be addressed. One of those challenges is Load Balancing.

II. LOAD BALANCING

Load Balancing is a technique in which the workload on the resources of a node is shifted to respective resources on the other node in a network without disturbing the running task[2]. A load balancer acts as the "traffic police" sitting in front of your servers and routing client requests across all servers capable of fulfilling those requests in a manner that maximizes a. speed and capacity utilization and ensures that no one server is overworked, which could degrade performance. Cloud environment must serve millions of concurrent requests from users or clients and return the correct text, images, video, or application data, all in a fast and reliable manner. Load Balancer ensures that no server is heavily loaded or lightly loaded.

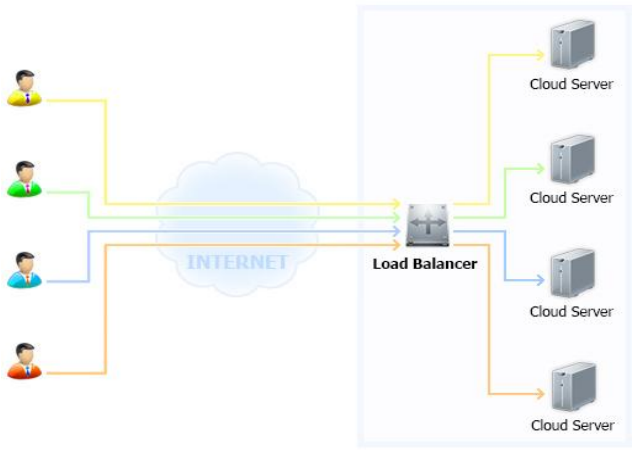

Figure 1: Load Balancer in Cloud

III. TYPES OF LOAD BALANCERS

Load balancers typically come in two flavours: hardware-based and software-based. Hardware load balancers are extremely acute. A hardware load balancer offers more deployment capabilities. It has fewer flaws compared even to strictly software based balancers. Software load balancers are based on algorithm. Software Load Balancing is classified into Static and Dynamic algorithms.

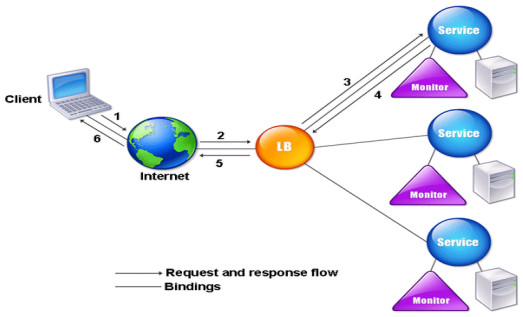

Figure 2: Architecture of Load Balancing

A. Static Load Balancing

In this approach load balancing is achieved by providing priori information about the system. The performance of the node is determined at the commencement of execution. Nodes calculate their allotted work and submit the result to remote node. Then depending on the performance work load is distributed in start without considering the current load. It is non-preemptive. E.g. Round Robin, Min-Min,Max-Min,Max-Min-Max, Central Manager, Threshold algorithm and Randomized algorithm.

b. B. Dynamic Load Balancing

These algorithms monitor changes on the system work load and redistribute the work accordingly. This algorithm usually composed of three strategies: transfer strategy, location strategy and information strategy. Dynamic Load Balancing

algorithm is Preemptive. E.g. Central Queue, Local Queue, Least Connection, Throttled, Honeybee Foraging and Biased Random Sampling.

IV. STATIC LOAD BALANCING ALGORITHMS

A. Round Robin Algorithm

It is one of the simplest scheduling technique that utilize the principle of time slices. Here the time is divided into multiple slices and each node is given a particular time slice or time interval i.e. it utilizes the principle of time scheduling. Each node is given a quantum and its operation. The resources of the service provider are provided to the requesting client on the basis of time slice. In this algorithm some cloud servers may be heavily loaded but some are not[2].

B. Min-Min Algorithm

This algorithm begins with a set of all unassigned jobs. Minimum completion time for all jobs is calculated and the job with minimum completion time is selected. Then, the node which has the minimum completion time for all jobs is selected. Finally, the selected node and the selected job are mapped. This process is repeated until all the unassigned jobs are assigned. In Min Min algorithm, a request with long processing time may wait for a long time[3].

C. Max-Min Algorithm

Max-Min algorithm calculates the minimum completion time for all jobs. After finding out minimum completion time of jobs, the maximum value is selected. The machine that has the minimum completion time for all the jobs is selected. Finally the selected node and the selected job are mapped[3].

D. Central Manager Algorithm

In this algorithm, acentral processor selects the host for new process. The minimally loaded processor depending on the overall load is selected when process is created. Load manager selects hosts for new processes so that the processor load confirms to same level as muchas possible. From the on hand information on the system load state central load manager makes the load balancing judgment. This information is updated by remote processors, which send a message eachtime the load on them changes.

E. Threshold algorithm

According to this algorithm, the processes are assigned immediately upon creation to hosts. Hosts for new processes are selected locally without sending remote messages. Each processor keeps a private copy of the system's load. The load of a processor can characterize by one of the three levels: underloaded, medium and overloaded. A disadvantage of the algorithm is that all processes are allocated locally when all remote processors are over loaded[8].

V. DYNAMIC LOAD BALANCING ALGORITHMS

A. Central Queue Algorithm

In this algorithm, whenever a new request arrives, the queue manager inserts it into the queue. When request for an activity is received by the queue manager it removes the first activity from the queue and sends it to the requester. If no ready activity is present in the queue the request is buffered, until a new activity is available. When a processor load falls under the threshold then the local load manager sends a request for the new activity to the central load manager.

B. Least Connection Algorithm

The default method, when a virtual server is configured to use the least connection, it selects the service with the fewest active connections.

C. Throttled Algorithm

Throttled algorithm is completely based on virtual machine. In this client first requesting the load balancer to check the right virtual machine which access that load easily and perform the operations which is given by the client or user. This ensures that only a pre-defined number of internet cloud-lets are allocated to a single VM at any given time. If more request groups are present than the number of available VMs at a data centre, some of the request will be queued until the next VM becomes available [5].

D. Honeybee Foraging Algorithm

The idea of Honeybee Foraging algorithm is derived from the behaviour of honeybees – Finders and Reapers. The servers are grouped as virtual server and each virtual server have a process queue. Each server, after processing a request from its queue, calculates the profit. If profit is high, the server stays else, it returns to the forage. But there is no significant improvement in throughput [9].

E. Biased Random Sampling

In Biased Random Sampling, a virtual graph is constructed. Each node is represented as a vertex in a directed graph and each in-degree represents free resources of that node. On request, the load balancer allocates the job to the node which has at least one in-degree. After allocation, the in-degree of that node is decremented by one. After Job completion, it increments the in-degree by one. The performance is degraded as the number of servers increase due to additional overhead for computing the walk length [4].

Table 1: Comparison of Load Balancing Algorithms

Algorithm	Nature	Environment	Merits	Demerits
Round Robin[7]	Static	Decentralized	All servers are treated equally	Not ideal for tasks with varying execution time
Min-Min	Static	Centralized	Better performance with small requests	Starvation may occur.
Max-Min	Static	Centralized	Efficient with concurrent execution	Requests with long execution time may wait for a long time.
Central Manager	Static	Centralized	Useful when dynamic tasks are created	Bottleneck may arise

Threshold[7]	Static	Decentralized	Remote memory access is minimized	Execution time is increased
Central Queueing	Dynamic	Centralized	Useful for heterogeneous nodes	Less resource utilization
Least Connection	Static	Decentralized	Useful for homogeneous nodes	Not useful for tasks of different durations
Honeybee Foraging	Dynamic	Decentralized	Response time and Waiting time are reduced significantly	Throughput may be decreased when more resourced are added
Biased Random Sampling[7]	Dynamic	Decentralized	All cloud servers are utilized	Servers may be heavily loaded

CONCLUSION

Cloud computing provides everything to the user as a service which includes platform as a service, application as a service, infrastructure as a service. . Load balancing is necessary in cloud computing if efficient and maximum utilization of resources needs to be achieved.

This paper made a comparative study on the Static and Dynamic Load Balancing algorithm.

This analysis of various algorithm helps in designing the new enhanced load balancing technique or algorithm.

References

[1] M. Katyal and A. Mishra, "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment," vol. 1, issue 2, pp. 5-14, 2013. (2)

[2] https://www.rackspace.co.uk/

[3] T. Kokilavani and D.I. George Amalarethinam “Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing” International Journal of Computer Applications, Volume 20– No.2, April 2011.

[4] D.Saranya, L.SankaraMaheshwari“ Load Balancing Algorithms in Cloud Computing:A Review” Volume 5, Issue 7, July 2015

[5] ShipraGoyal,Manoj K Verma “Load Balancing technique in Cloud Environment-A Review” IJARCSE, Volume 6, Issue 4, April 2016

[6] ShridharG.Domanal and G.RamMohana Reddy, “Load Balancing in Cloud Computing Using Modified Throttled Algorithm”, IEEE Randles, M.Bendiab, A. T. & Lamb, D. (2008).

[7] MayankaKatyal, Atul Mishra“ A comparative study of Load Balancing algorithms in Cloud Environment.

[8] Daniel Grousa, Anthony T. ,” Non-Cooperative load balancing in distributed systems”. Journal of Parallel and Distributing Computing, 2005.

[9] Y. S. Sheeja and Jayalekshmi “Cost effective load balancing based on honey bee behaviour in cloud environment” Journal of IEEE, Dec. 2014.