International Journal of Advanced Technology in Engineering and Science
Vol. No.5, Issue No. 05, May 2017
www.ijates.com
ijates
ISSN 2348 - 7550

WEB SERVER LOAD BALANCING BASED ON CLOUD COMPUTING

Arun K Lekshmon[1], Deepa S Kumar[2]

[1,2]Department of Computer Science,College of Engineering Munnar,kerala (India)

ABSTRACT

Internet is growing rapidly. The web server processing capacity is limited due to the resource availability. The networking technology is improving its speed and as a result the number of clients accessing web server increases. Due to the limited capacity of the server, large number of clients accessing the server become a bottleneck. To overcome this problem we use web server clusters and different load balancing techniques. In this paper we focus on how cloud computing can be used to reduce the server load and improve the performance of a web server. Cloud computing provide computing resources through Internet. Cloud uses efficient resource sharing techniques and round robin load balancing methods for improve the performance of the web servers.

Keywords: ***Cloud computing; Load balancing; Web servers;***

I. INTRODUCTION

Cloud computing is emerging technology, which provides scalable resources to the users through network services. The resources includes computing power, memory(RAM), storage, and networking. Cloud provide these resources to the user as a payable service. It include the strategy that pay for what you use. Cloud computing technology is growing rapidly due to increasing demand of computing resources. Now a days computing technology needs large number of resources for fast and efficient processing of services that needed by the user. Traditional computing technologies cannot handle such large number of resources. According to NIST, cloud computing is a model that provide access to shared pool of resources through networking services [1,2]. In this paper we focus on how cloud computing can be used to improve the performance of a web server. Traditional web servers consist of number of machines with sufficient computing power and memory. Due to the increase in network technology and number of users,traditional servers cannot handle large number of requests from clients. Cloud computing uses efficient use of resources and load balancing techniques to improve the performance of web servers.

II. CLOUD COMPUTING

Cloud computing fast growing technology ,which offers computation is done in virtual computers rather than local machines. Cloud provide computation services over virtual machines, which can be accessed by the user anywhere from the network. The main feature of cloud computing includes scalability of resources according to the demand of the

users. On-demand self-service, broad network access, elasticity, measured service [3]. Cloud provides three service models to the users [4].

A. Infrastructure as a Service (IaaS)

Infrastructure as a service provide fundamental computing resources to users, which includes storage, network, computing capabilities etc. In this model users can use their own operating system, storage techniques and applications.

B. Platform as a Service (PaaS)

Platform as a service provide a developmental platform for the users. By using this platform, users can develop their own application and deploy it. In this model users cannot control over the cloud infrastructure.

C. Software as a Service(SaaS)

Software as a service provide application services over Internet. Applications are hosted by service providers and make available to the users. Google apps social applications like twitter Facebooketc. are examples of SaaS.

III. WEB SERVER LOAD BALANCING

A web server is a computer system that accept request from clients and process the request and send information back to the clients [5]. Client uses software called web browsers to connect to the server. The communication between client and server uses hypertext transfer protocol. A web server can handle a limited number of concurrent clients. When the number of clients exceeds the maximum limit then the web server is said to be overloaded. When the server is overloaded, it affects the normal working of the server. Overload may cause delayed service of client requests, server returning HTTP error code or does not respond to the client requests.

Different techniques are used to avoid the web server overload, which includes monitoring the network traffic, adding more hardware resources to the server, adding security to avoid denial of service attacks, add multiple servers and use load balancing technologies to manage the traffic from different clients.

Load may comprise of amount of memory used, CPUload, network load, delay load etc. Load balancing is the technique used improving the efficiency of computing by distributing the workload across multiple computing resources. Efficient load balancer can helps in utilizing the available resources optimally and minimize the resource usage. A load balancer divide the work among different servers. Load balancer can be a software or a hardware device.

A. Traditional Load Balancing Techniques

In traditional load balancing approach uses different methods to distribute the incoming client requests to different servers. Based on position of the balancer situated in the system, we can group balancing methods into four classes [6].

1) Client side load balancing

In this method balancing is done in the client side. Client based approach uses two methods for server balancing. First method uses a web client to select the appropriate server. In this method web browser select the appropriate web server from cluster of servers. Netscape browser is an example of this kind of approach, when a user access the Netscape

homepage then the browser select one of the server from the number of servers to retrieve the page. Problem related with this method is when ever the server architecture is changed then it must be updated in the browser. It is not generally applicable method because it is not possible to own a browser for each website.

Second method for client side load balancing is, usage of a client side proxy server. Proxy server implemented with Web Location and Information service, which keep track of replicated web servers and route client requests to appropriate server. This method can't be used for large scale [7].

2) DNS based load balancing

DNS based approach, load balancing is done at DNS server. Each naming domain maintains its own local address and hostname information. When a user request an address, browser send it to dns server to convert canonical name to IP address. DNS server uses different load balancing algorithm to select the server.

3) Server side load balancing

Server side load balancing uses a server side proxy server. Proxy server will not process the HTTP requests, it remap client requests to most suitable server,that is server with low load in it. Draw back of this method is that, in most cases the proxy server it self become a bottleneck and the method cannot be scaled .

4) Network based load balancing

Network side balancing can be done at two levels, First is in the network layer and next is in between network and link layer. Network layer mapping consist of a HTTP scheduler, it done the operation similar to the normal routing and forwarding operations . Each replicated server has a unique IP address and all requests from clients goes to a http Scheduler which has a logical IP address and scheduler remap the request to lowest load server. Drawback of this method is that the IP packet must be modified In the second method remapping is done between network and link layers, in method all replicated web servers and logical server has the same IP address, remapping is done depending on the different network technology used. MAC address or port number can be used for remapping.

Monitoring the load in server scheduler send Internet control Message Protocol (ICMP) echo request to replicated servers, and the response time gives an indication of the load in that server. Another method is to monitor the traffic or response time form each server.

B.Load Balancing Algorithms

Load balancing algorithm can be classified into to classes. Static load balancing algorithms have prior knowledge related to system resources and details of all tasks, and execution depends on these prior knowledge, and don’t depend on the current state which the system is in. These types of algorithms are easy to design and Implement e.g. Round robin Algorithm, Randomized Algorithm [8,9].

Dynamic algorithms take decisions depending on the current state of the system. Dynamic algorithms are complex, but perform better than static algorithms.

1) Round robin Algorithm

Round robin algorithm is a static load balancing algorithm, which does not consider any system state information. Algorithm assigns time slices to each sever in a circular manner. Round robin Scheduling is easy to implement and simple. Drawback of this algorithm is that, it will not take the current state of the system and make poor assignment decisions [10].

2) Weighted Round robin algorithm

Weighted round robin is the modified version of round-robin algorithm. In this algorithm each server is assigned a weight based on the load capacity. It assigns time slices to each server in a circular manner according to the weight in the server. It select the next lowest weight server. This will improve the resource allocation [10,11].

3) Biased random sampling

In this algorithm, load on the server is represented by connected virtual graph. Each node in the graph represent a server and each in-degree represents servers free resources. Load on the server is determined by analyzing edges in the graph. An edge is removed from the node, whenever it takes a new job, which indicate load on the server and add an inward edge when a job is completed, indicating resources are available [9,12].

4) Active clustering

Active clustering is considered as a self aggregation algorithm, this is done by grouping similar type instances together. This load balancing algorithm works well, if each node is aware similar nodes and delegate workload to them. It forms a similar type node cluster who can share the workload. Algorithm first select a random node as initiator and select a matchmaker from its neighbours. Matchmaker node forms link between nodes that is similar to the initiator node [9].

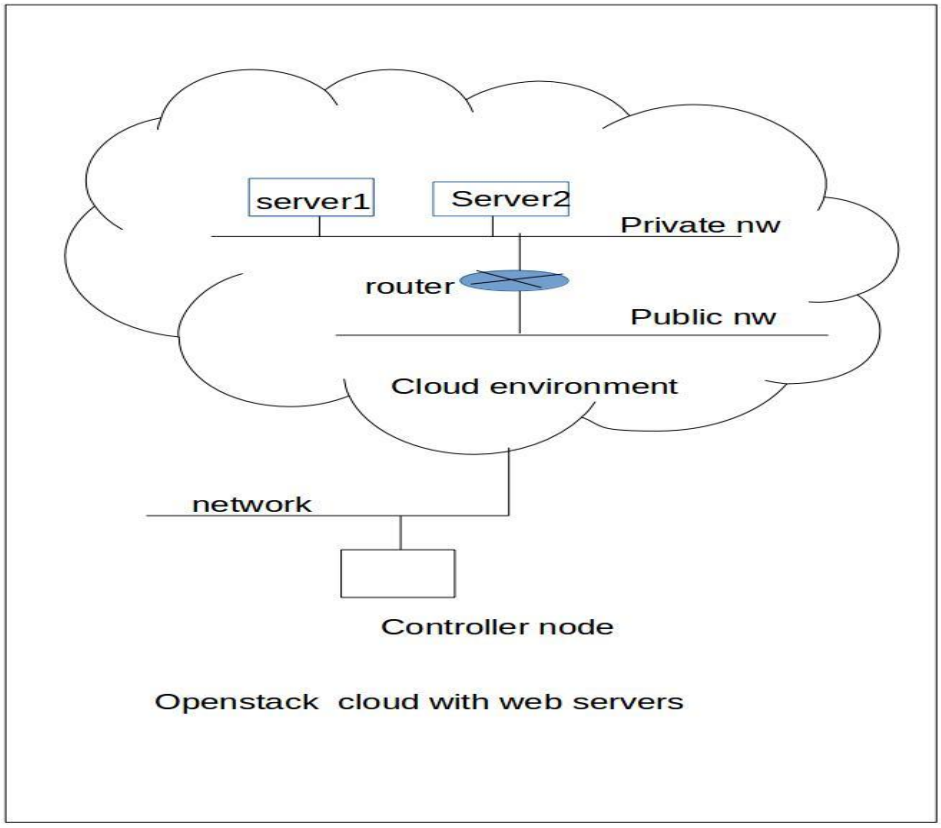
IV. CLOUD BASED IMPLEMENTATION

Cloud computing provide hardware and software resources through Internet according to the users need. Cloud based web server balancing consist of mainly three steps, implementing a cloud platform, creating number of mirror web servers inside the cloud using virtual machines and implementing the load balancer method.

Fig 1: Openstack Cloud with servers

Creating a web server inside the cloud environment consistof launching number of virtual machines with apache serviceinstalled in each VM. Each server is assigned a unique IPaddress.

Cloud platform is implemented using OpenStack cloud. It is an open source platform, mainly deployed as Infrastructure as a service. OpenStack cloud computing platform is developed by Rackspace and NASA. OpenStack consist of modular architecture. Its core modules are , Nova which is a computing service used to manage and automate



Openstack cloud with web servers

computing resources. Neutron is the networking service responsible for managing network. Keystone is the identity service used for authentication and control the access. Cinder which is responsible for managing block storage. Glance service is an images service used to store disk and server images. Implement the OpenStack platform we install all the core services.

Implementing load balancer includes creating a pool ofweb servers. Each server is assigned with a virtual IP andassociating a floating IP with these virtual IP's. Floating IP isthe IP address used by the clients to connect to the server.

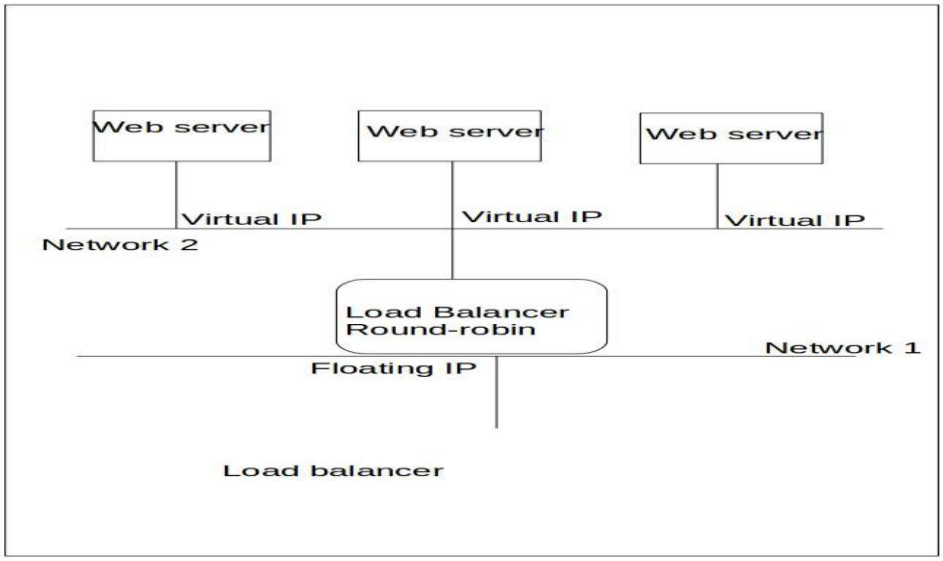


Fig 2: Load Balancer

Floating IP and virtual IP's are in two different networks.Load balancer service is installed in floating IP network whichconsist of round-robin method to select the virtual IP. Round robin algorithm select virtual IP's in circular manner to avoidoverload in the servers.

IV. CONCLUSION

The paper discusses the different load balancing techniques. And present a design for cloud based web load balancing. Cloud based web servers provide better performance by avoiding overload conditions and efficient use of resources and load balancing techniques.

REFERENCES

[1] Kumar, Pawan, and Rakesh Kumar. "Optimal resource allocation approach in cloud computing environment." *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on*. IEEE, 2016.

[2] Jadeja, Yashpalsinh, and Kirit Modi. "Cloud computing-concepts, architecture and challenges." *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*. IEEE, 2012.

[3] Dillon, Tharam, Chen Wu, and Elizabeth Chang. "Cloud computing: issues and challenges." *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. Ieee, 2010.

[4] Ghaffari, Fariba, Hossein Gharaee, and Mohammad Reza Forouzandehdoust. "Security considerations and requirements for Cloud computing." *Telecommunications (IST), 2016 8th International Symposium on*. IEEE, 2016.

[5] Arlitt, Martin F., and Carey L. Williamson. "Internet web servers: Workload characterization and performance implications." *IEEE/ACM Transactions on networking* 5.5 (1997): 631-645.

[6] Bryhni, Haakon, Espen Klovning, and Oivind Kure. "A comparison of load balancing techniques for scalable web servers." *IEEE network* 14.4 (2000): 58-64.

[7] Cardellini, Valeria, Michele Colajanni, and Philip S. Yu. "Dynamic load balancing on web-server systems." *IEEE Internet computing* 3.3 (1999): 28-39.

[8] Ragmani, Awatif, et al. "A performed load balancing algorithm for public Cloud computing using ant colony optimization." *Cloud Computing Technologies and Applications (CloudTech), 2016 2nd International Conference on*. IEEE, 2016.

[9] Singh, Amritpal. "A Review of Existing Load Balancing Techniques in Cloud Computing." *International Journal of Advanced Research in Computer Engineering & Technology* 4.7 (2015).

[10] Radojević, Branko, and Mario Žagar. "Analysis of issues with load balancing algorithms in hosted (cloud) environments." *MIPRO, 2011 Proceedings of the 34th International Convention*. IEEE, 2011.

[11] Zongyu, Xu, and Wang Xingxuan. "A predictive modified round robin scheduling algorithm for web server clusters." *Control Conference (CCC), 2015 34th Chinese*. IEEE, 2015.

[12] Randles, Martin, David Lamb, and A. Taleb-Bendiab. "A comparative study into distributed load balancing algorithms for cloud computing." *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*. IEEE, 2010.

[13] Huo, Jiuyuan, Hong Qu, and Ling Wu. "Design and implementation of private cloud storage platform based on OpenStack." *Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on*. IEEE, 2015.

https://www.openstack.org

Books:

[2] R.E. Moore, Interval analysis (Englewood Cliffs, NJ: Prentice-Hall, 1966).(10, Times New Roman)

Note that the title of the book is in lower case letters and italicized. There is no comma following the title. Place of publication and publisher are given.

Theses:

[3] D.S. Chan, Theory and implementation of multidimensional discrete systems for signal processing, doctoral diss., Massachusetts Institute of Technology, Cambridge, MA, 1978. (10, Times New Roman)

Note that thesis title is set in italics and the university that granted the degree is listed along with location information **Proceedings Papers:**

[4] W.J. Book, Modeling design and control of flexible manipulator arms: A tutorial review, Proc. 29th IEEE Conf. on Decision and Control, San Francisco, CA, 1990, 500-506 (10, Times New Roman)