Volume 6, Issue 8, August 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Proposed Model for proper Balancing of Load in the Public Cloud

G. Thejesvi, E. Hari Prasad
Academic Consultant, Department of Computer Science, Dravidian University, Kuppam, Chittoor Dist, Andhra Pradesh, India

Abstract: ***The erroneous development in the computer and communication technology leads to use various web based software applications using with the Internet. Cloud computing is an emerging technology where millions of clients and individuals use various cloud services like storage, software's and infrastructure on rental basis. Tremendous increase in the number of users has led to some issues and problems. One of the main issues is balancing the work load and increasing the performance of the system. To balance the load equally to all nodes, various static and dynamic algorithms have been proposed and these algorithms consider various parameters like performance, response time, fault tolerance, high availability, cost parameters, number of services, scalability, flexibility, reduced overhead for users, etc. Each algorithm has its advantages and disadvantages. Hence there is a need to do more research work in this area i.e. load balancing. In this paper, we proposed a model for proper balancing of load in the public cloud.***

Keywords: ***Public cloud, Load balancing, main controller, balancers***

I. INTRODUCTION

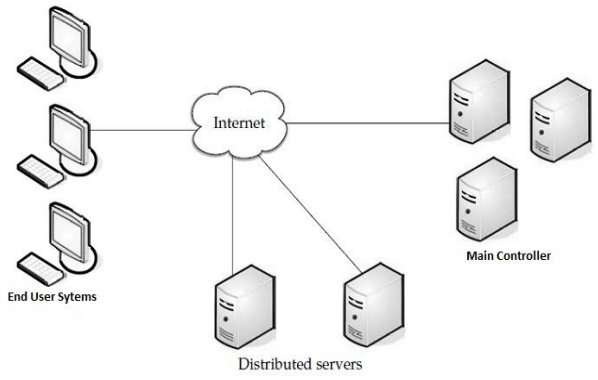
Cloud computing is a technology where one can avail various services like storage, software usage and also some other shared equipments based on requirement of the users on rental basis. Generally cloud computing can work with a normal system and the Internet only. Cloud computing is a new concept. Public cloud is the service model which has huge number of clients. Processing in public cloud is delayed because of high traffic in this cloud environment. Cloud technology is best and most advantageous technology now a day. But the maintenance of various nodes is a big issue here. Hence, resolving issue regarding the balancing the load is an interesting area of research.

Many algorithms were developed for balancing the load in the cloud computing. The load balancing algorithms use prior knowledge of user requirements, capacity of nodes, processing power, memory, and performance. The early algorithms are static which consider the previous state of the cloud environments only. These algorithms are not adaptable to the run time changes of load. These algorithms do not distribute the load properly, thus resources cannot be utilized optimally to the maximum possible extent. Dynamic load balancing algorithms take into account the current state of the system. It uses run time statistics to balance the load in the system. Though dynamic load balancing algorithms are difficult to simulate, they are highly adaptable to cloud environment.

In this dissertation, a dynamic load balancing algorithm called "Enhanced round robin algorithm" is proposed. The proposed algorithm distributes the load based on two parameters – node status (idle, normal, or overloaded) and time stamping. The performance of the proposed algorithm is compared with throttled load balancer algorithm and least connection scheduling algorithm the comparison results showed that proposed algorithms performed better than other algorithms.

II. CLOUD COMPONENTS

The cloud architecture generally consists of users, central controller, and distributed servers. These are the components of a cloud computing and each component in the cloud architecture is very important and each has its own functionality.



Cloud Computing Architecture

Users:

Generally end users are the users who use these cloud services. The end users can also be called as clients. These clients who interact with the service provider can be categorized into three types as follows in [1]:

- **Mobile users**: users who use the smart devices like Tabs, Smart watch, iPod and smart phones like a Blackberry to access the cloud services.
- **Thin Clients:** These types of users will display the processed data and information. These users do not perform any computation work. Based up on the query, the clients retrieve the data, but the internal processing is done at cloud server only. Even any internal memory is also not available for these users.
- **Thick Clients**: The users of this type generally connect to the service provider's web portal using an internet connection and various browsers like Firefox, Internet Explorer or Google Chrome.

Thin clients are generally cheaper to buy, maintain and replace than their thick clients counterparts and consume less power.

III. MAIN CONTROLLER

The Overall functionality of a cloud computing depends upon the main controller only. Generally main controller is a set of servers with high capacity and performance where the end users can access different applications. When the end user request for a cloud service, the main controller checks the request and forwards the request to the appropriate balancer to finish the task. The service provider can have multiple branches and they can maintain different controllers in different geographical locations based on cloud partition. The main controller in Cloud Computing works on logic of virtualization, where high utilization is achieved by allowing one server to compute several tasks concurrently.

Distributed Servers:

The other remaining component of cloud computing architecture is Distributed server. This server shares a single instance of data and/or applications to the various hosts. The users feel that they are using data and/or applications from their own machine.

Load Balancing:

Load balancing is a process of distributing the overall load to the all available nodes in the cloud system. The main objective of balancing the load is to utilize the resource maximum, to increase the response time and to perform efficiently. Distribution of load in the cloud depends on load balancing algorithm. While developing a load balancing algorithm, one has to consider the number of tasks submitted to cloud, the various resources required to complete the task, and load status of all available nodes etc.

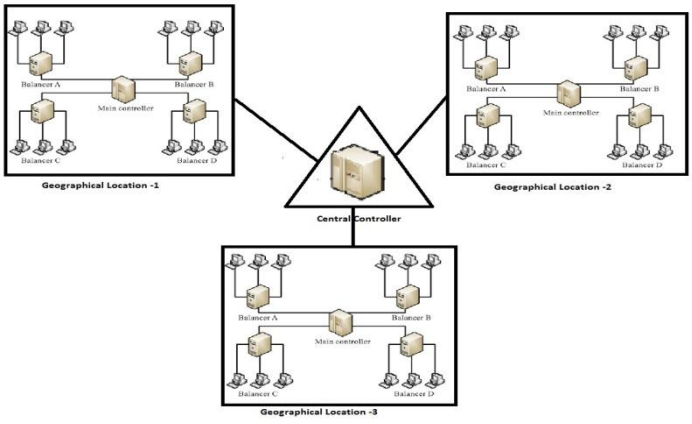
Load balancing Objectives:

The following are some of the objectives of a load balancing algorithm have to achieve. [15]

- To increase the overall system performance efficiently.
- To be robust when any failure of the system occurs
- To keep a backup of data when system crashes or partially effected.
- To keep up the system powerful and stable.
- To support scalability
- To be easy to implementing the algorithm
- To reduce the cost of the system.

Proposed Cloud Architecture for proper balancing of load:

The public cloud contains many nodes which are placed at different geographical locations. To manage these nodes of large public cloud in a better way, the concept of cloud portioning is used. Cloud portioning is nothing but a grouping of nodes in a subarea of cloud i.e. groups of nodes are formed on geographical area. The proposed algorithm is based on cloud portioning concept. The overall architecture of the proposed system is shown in the below figure.



Proposed Cloud Architecture

In this architecture, central controller selects an appropriate cluster for load distribution. Each cluster has a main controller that selects appropriate balancer for job allocation from group of balancers under its control. When a job arrives, the central controller selects an appropriate cluster for resource allocation to this job and this job is sent to the main controller of the selected cluster. The main controller is responsible for selecting an appropriate balancer within the cluster. The balancer selects a node to be allocated to the job from a set of nodes under its control. The decision to select a node to attend a job is done at two levels, first at main controller and then at balancer level.

The proposed Enhanced Round Robin algorithm is executed at main controllers and at balancers. It is assumed that the clusters are formed by the cloud provider, and an appropriate strategy is applied to send a job to the clusters' main controller.

IV. CONCLUSION

The proper balancing of load in the cloud computing is the one of the main challenges of cloud environment. The allocation of jobs to the various virtual machines should be done properly; otherwise it may affects, the overall system performance. The proper allocation of every job is aimed for client satisfaction. If the job schedule is done properly, automatically it increases the performance and stability of a system. Many algorithms have been suggested for balancing the load. But each and every algorithm has its advantages and limitations. The necessity of balancing the load, metrics need to measure the load balancing and various load balancing algorithms developed so far are explained here. The proposed model properly distributes the work load and balances all the nodes in the public cloud.

REFERENCES

[1] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach,TATA McGRAW-HILL Edition 2010.

[2] Stefan Ried, Holger Kisker, and Pascal Matzke. The evolution of cloud computing markets. Technical report, Forrester Research, 2010.

[3] Tom Jenkins. "Managing Content in the Cloud". Open Text Corporation, 2010.

[4] Mark Russinovich. http://channel9.msdn.com/Events/BUILD/BUILD2011/SAC-852F, April 2012.

[5] Ewald Roodenrijs. Private versus public cloud. Computable, March 2011.Oracle Cloud Conference, 2012.

[6] Sander Hulsman. Overheid vreest voor veiligheid in de cloud. Computable, 4, February 2012.

[7] Livny, M.; Melman, M. (2011): "Load Balancing in Homogeneous Broadcast Distributed Systems." Proceedings of the ACM Computer Network: Performance Symposium, pp. 47-55.

[8] A. Hamo, A. Saeed, "Towards a Reference Model for Surveying a Load Balancing," IJCSNS International Journal of Computer Science and Network Security, vol. 13, No. 2, 2013, pp. 42-47.

[9] Morgan Stanley, "Cloud Computing Takes Off", , May 23, 2011.

[10] T. Sharma, V.K. Banga, "Proposed Efficient and Enhanced Algorithm in Cloud Computing," International Journal of Engineering Research & Technology (IJERT), vol. 2, Issue 2 ,2013, pp. 1-6.

[11] S. Begum and C.S.R. Prashanth, "Review of Load Balancing in Cloud Computing," IJCSI International Journal of Computer Science Issues, Vol.10, Issue 1, 2013.

[12] A. Khetan, V. Bhushan and S. Ch. Gupta, "A Novel Survey on Load Balancing in Cloud Computing," International Journal of Engineering Research & Technology (IJERT) , Vol.2, Issue 2 ,2013.

[13] Ali M. Alakeel, A Guide to Dynamic Load Balancing in Distributed Computer Systems,IJCSNS International Journal of Computer Science and Network Security,VOL.10 No.6, June 2010.

[14] L. Yao, G. Wu, J. Ren, Y. Zhu and V. Li, "Guaranteeing Fault-Tolerant Load Requirement Balancing Scheme," Published by Oxford University Press on behalf of The British Computer Society, 2013, pp. 1-8.

[15] T-Y., W-T. Lee, Y-S. Lin, Y-S. Lin, H-L. Chan and J-S. Huang, "Dynamic load balancing mechanism based on cloud storage" in proc. Computing, Communications and Applications Conference (ComComAp), IEEE, pp: 102-106, January 2012.

[16] Ajaltouni, E. E., Boukerche, A., & Zhang, M. (2008). An Efficient Dynamic Load Balancing Scheme for Distributed Simulations on a Grid Infrastructure. 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, 61-68. Ieee. doi:10.1109/DS-RT.2008.38

[17] Y. Lua, Q. Xiea, G. Kliotb, A. Gellerb, J. R. Larusb and A. Greenber, "Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services", An international Journal on Performance evaluation, In Press, Accepted Manuscript, Available online 3 August 2011.

