Overview of Load Balancing Algorithms in Cloud Computing

T.Kiruthika
MPhil Research Scholar
SNMV College of Arts & Science

M.Chandra Kumar
HOD, Department of BCA
SNMV College of Arts & Science

Abstract—Cloud computing is a computing model. Cloud provides convenient, on-demand access to a shared pool of computing resources such as networks, servers, storage, applications and so on. Allocation of cloud resources to users on demand gives rise to the problem of load balancing. If workload is not distributed properly, then some nodes in cloud will be heavily loaded and some nodes will be under loaded. In the same way if the resources provided by the cloud are not allocated efficiently, it leads to delay in providing service to the users .For cloud environment various load balancing approaches have been implemented to provide efficient distribution of load among available machines. Such as Round Robin load balancing, Throttled load balancing, Min-Min load balancing, Min-Max load balancing, Honey Bee and Ant Colony behavior based load balancing, etc. For effective load balancing, a single load balancing algorithm is not sufficient. Hence there is requirement of algorithm which combines features of two or more load balancing algorithms.

Keywords—**Cloud computing, Load Balancing, Deployment models, Virtualization, Service models.**

1. INTRODUCTION

Cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualized computer. Cloud is a platform providing dynamic pool resources and virtualization. It also facilitates the scalable IT resources such as services, applications and infrastructure. These resources work with internet on pay-per-use basis which helps in adjustment of capacity in a fast way.

Service models in cloud

- Software as a Service (SaaS),
- Platform as a Service (PaaS),
- Infrastructure as a Service (IaaS).

Deployment models in cloud

- Private Clouds,
- Community Clouds,
- Public Clouds,
- Hybrid Clouds.

Characteristics of cloud

- on-demand self-service,
- broad network access,
- resource pooling,
- rapid elasticity, and
- measured service.

Virtualization

Virtualization is the creation of a virtual version of something, such as hardware platform, operating system, storage device or network resources. One physical hardware can run multiple OS and applications through a hypervisor. A hypervisor is the virtualization manager on a physical hardware. Cloud cannot exist without virtualization.

2. LOAD BALANCING

Load balancing is technique to distribute job or task or load to different nodes to execute the process on shortest time on distributed network. [4]There are several different issues of cloud computing that last from years ago, one of that issue is load balancing. Cloud is very high flexible and provides great retrieval of data as per user's requirements and other services, apart from this to handle large amount of data several techniques to optimize load is workload which is total time required by the processor to execute all assigned processes are estimation of load, load comparison, different system stability, system performance, interaction between the nodes, nature

of work to be transferred, selecting of nodes and many other ones to consider while developing such algorithm [5]. The goal of load balancing is minimizing the average response time. The user should not wait in queue.

3. NEED OF LOAD BALANCING ALGORITHM IN CLOUD COMPUTING

Load balancing in clouds provide distributing the excess dynamic local workload evenly across all nodes. The main aim of load balancing is to distribute the traffic among the node equally in the cluster for the better performance of network.

4 TYPES OF LOAD BALANCING ALGORITHM

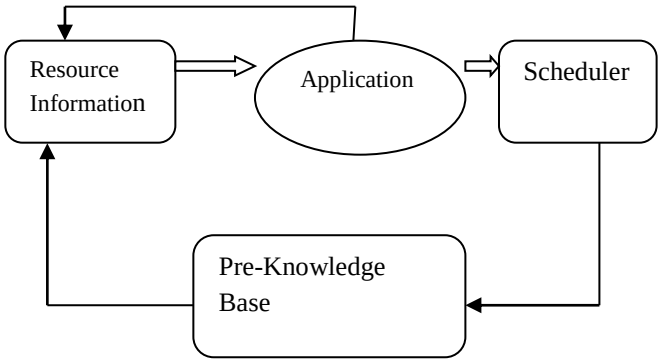
There are two types of Load Balancing Algorithm. They are

- Static Load Balancing Algorithm
- Dynamic Load Balancing Algorithm

Static load balancing algorithm

In static algorithms [6], prior knowledge about the system is already known which includes processing power, memory, performance and data about user's requirements. These algorithms do not need the information regarding current state of the system. Static load balancing algorithms allocate the tasks of a parallel program to workstations based on either the payload at the time nodes are allocated to some task, or based on an ordinary burden of our workstation cluster. The decisions linked to load balance are made at compile time when resource demands are calculated.

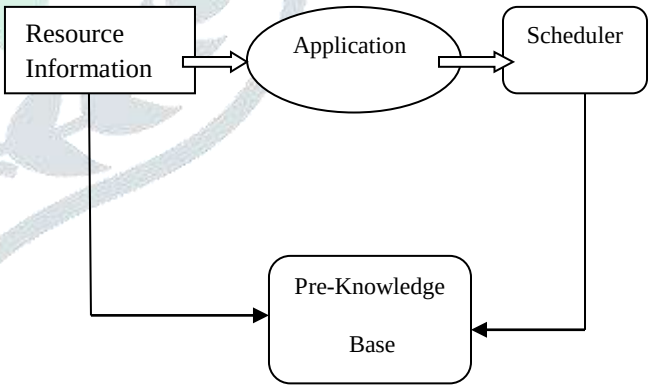
The main concept behind this algorithm is that each server has been assigned a weight and then, server having the highest weight receives more connections. In equal weighted condition, servers will receive balanced traffic. This approach is generally defined during the design or implementation period of the system.



N.4.1 Static Load Balancing

Dynamic Load Balancing algorithm

Dynamic load balancing algorithms are those algorithms which search for the lightest server in the network and then designated appropriate load on it. In this, work load is distributed among the processors at runtime. The algorithms in this category are considered complex, but have better fault tolerance and overall performance. The load depends on the current state and no demand for prior knowledge. The advantage is if any node fails, it does not end the operation. The goal of dynamic load balancing is to allocate job dynamically [7,8].



4.2 Dynamic Load Balancing

COMPONENTS OF LOAD BALANCING ALGORITHM

- Transfer policy
- Selection policy
- Location policy
- Information policy
- Transfer policy

- **Transfer policy**

Transfer policy is responsible to determine when a task should be transferred from one node to the other node.

- **Selection policy**

Selection policy focuses on choosing the processor for load transfer so that the overall response time and throughput may be improved.

- **Location policy**

Location policy determines the availability of essential resources for providing services and makes a selection based on location of resources.

- **Information Policy**

Information policy acquires workload related information about the system such as nature of workload and the average load on each node.

5 EXISTING LOAD BALANCING ALGORITHMS IN CLOUDCOMPUTING

In cloud computing environment, there are various Load Balancing Algorithms which are closely analyzed.compared on the bases of some predefined metrics, such as throughput, response time, Overhead, performance, fault tolerance, migration time, resource utilization, and scalability the load balancing algorithm contains many types.

Round Robin Algorithm

Round robin works in the round manner where each customer is allotted with a time slice and it has to wait for their turn. The time is divided and the interval is allotted to each node. The process is going to occur in round robin until the all processes complete their task such that a balance technique is implemented in order to balance the process in a group.

Honey Bee Foraging Algorithm

This algorithm is utilized to allocate job dynamically. This algorithm also balances the priorities of tasks on the machines in such a way that the amount of waiting time in the waiting line is minimal. This load balancing technique works well for heterogeneous cloud computing systems [9].

Load Balance Min-Min (LBMM):

LBMM scheduling algorithm [10] and new optimized Load Balancing Max-MinMax (LB3M) [11] had main objective to minimize execution time of each task, also avoid unnecessary replication of task on the node thereby minimizing overall completion time. The advantage is reliable task assignment to clients. The disadvantage is slower than other algorithms because the task must pass three layers to process.

Max-Min Load Balancing

It works on the Opposite strategy as compare with min-min approach where the maximum value is consider to execute first. Max-Min [3] is almost same as the min-min algorithm except that 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. Then the ready time of the node is updated by adding the execution time of the assigned task. A maximum time taken process is shifted one by one.

Active Clustering Algorithm

It works as based on the clustering concept. The same type of nodes is grouped and working as the group. It increases the throughput of the system. This algorithm works on the principle of grouping similar one's and working on them group wise. The performance of the system is enhanced with high resources thereby increasing the parameter outcome using the algorithm. This algorithm is degraded with an increase in system diversity.

Fuzzy Active Monitoring Load Balancing (FAMLB)

This algorithm uses two parameters processor speed and load on virtual machine. In [2], the authors have introduced a new fuzzy logic based dynamic load balancing algorithm with additional parameters like memory usage, bandwidth usage, disk space usage, virtual machine status and named it as fuzzy Active Monitoring Load Balancer (FAMLB)

Join Idle Queue:

Join Idle Queue load balancing algorithm [1] is applied for dynamically scalable web services. This technique involves a dispatcher to whom processors informs at the time of their idleness, without interfering with job arrivals.

6 Conclusion

Cloud computing allows wide range of users to access distributed, scalable, virtualized, hardware and software resources over the Internet. Load balancing is one of the most important issue of cloud computing. It is a mechanism which distributes workload evenly across all the nodes in the whole cloud. Through efficient load balancing, we can achieve a high user satisfaction and resource utilization. Load balancing algorithms are good in distributing load equally to all virtual machines and perform task by increasing response time. A large number of parameters and different types of soft computing techniques can be included in the future for the better utilization and needs of the user.

References

[1] Yi Lua, Qiaomin Xiea, Gabriel Kliotb, Alan Gellerb, James R. Larusb, Albert Greenbergc, " Join-Idle-Queue: A Novel Load Balancing Algorithm for Dynamically Scalable Web Services" Volume 68 Issue 11, November, 2011, pp:1056-1071, Elsevier Science Publishers, 2011.

[2] Z. Nine, M. SQ, M. Azad, A. Kalam, S. Abdullah and R. M. Rahman, "fuzzy logic based dynamic load balancing in virtualized data centers" In fuzzy system (FU)

[3] Shahrzad Aslanzadeh, Venkatesh Mahadevan, Christopher Mcdermid," Availability and Load Balancing in Cloud Computing" International Conference on Computer and Software Modeling IPCSIT vol.14 IACSIT Press, Singapore 2011.

[4] Mayanka Katyal, Atul Mishra "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment"IJDCC, 2013.

[5] 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.

[6] Nidhi Jain Kansal, Inderveer Chana, "Cloud Load Balancing Techniques: A Step Towards Green Computing", IJCSI, Vol. 9, Issue 1, January 2012.

[7] Shiny, "Load Balancing in cloud computing", IOSRJournal of Computer Engineering", Vol 15, issue 2,2013,pp22-29.

[8] Yatendra Sahu, R. K. Pateriya, "Cloud computing overview with load balancing Techniques", International Journal of Computer Applications, 2013, pp 40-44.

[9] L. D. Dhinesh Babu, P. Venkat Krishna,"Honey bee behavior inspired load balancing of tasks in cloud computing environments", Elsevier, 2013.

[10] S. Wang, K. Yan, W. Liao, and S. Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, September 2010, pages 108-113.

[11] Che-Lun Hung, Hsiao-hsi Wang and Yu-Chen Hu "Efficient Load Balancing Algorithm for Cloud Computing Network", International Conference on Information Science and Technology (IST 2012), April 28-30, pp; 251-253.

[12] Nitin Kumar Mishra, Nishchol Mishra," Load Balancing Techniques: Need, Objectives and Major Challenges in Cloud Computing- A Systematic

Review", International Journal of Computer Applications (0975 – 8887) Volume 131 – No.18, December 2015.

[13] Sushil Kumar, Deepak Singh Rana," Various Dynamic Load Balancing Algorithms in Cloud Environment: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 129 – No.6, November2015.

[14] R. Ramya, M. Krishsanth, L. Arockiam," A State-of-Art Load Balancing Algorithms in Cloud Computing", International Journal of Computer Applications (0975 – 8887) Volume 95– No.19, June 2014.

[15] Shikha Gupta, Suman Sanghwan," Load Balancing in Cloud Computing: A Review", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 6, June 2015.