A Comparative Study of Various Load Balancing Techniques in Cloud Computing

Reena Panwar
M.Tech CSE Scholar
Department of Computer Science and Information Technology
Galgotias College of Engineering and Technology,
Greater Noida, India

Abstract—**In recent years, Cloud Computing has been an emerging computing model in the IT industry. Cloud computing opens up the world of computing to a broader range of uses and increases the ease of use by giving access to anything and anywhere through any internet connection. Load Balancing is major concern in the cloud virtual environment Now a day's many clients all over the world are demanding for various services, Cloud computing provide services by dynamically configure its servers and these servers may be present physically or virtually in the computing environment. Efficient load balancing schemes ensures efficient resource utilization by provisioning of resources to cloud users on demand basis in pay-as-you-say manner. This paper describes various load balancing approaches in cloud environments.**

Keywords—Cloud Computing,Load Balancing,Cloud Data Center,Virtual Machines,Resource Utillization.

I.   INTRODUCTION

Cloud Computing has become one of the popular technology adopted by both industry academia providing a flexible and efficient way to store and retrieve files. However the main issue is to schedule the incoming request in a way so with minimum response time, efficient resource utilization and at the same time resources should not be underutilized. Cloud Computing" is a term, which involves virtualization, distributed computing, networking, software and web services. For cloud environment various load balancing techniques have been proposed such as honeybee based load balancing technique, Active Clustering, Round Robin, Active Clustering ,Random sampling, Active Monitoring Load Balancer, Throttled Load Balancer, WCAP, JIO, CLBVM.

Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to the clients requirement at specific time. With the users and the types of application on the cloud computing platform increasing, it becomes a critical problem about how to use the resources in the system effectively to ensure service level agreements (SLA).

A cloud consist of several data centers, servers, clients which are interconnected in an efficient way. It includes fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services etc. With the rapid development of internet technology and network technology, more and more people use the internet to obtain information, shopping and entertainment. Server needs to response client request in order to improve user satisfaction.

II.   CLOUD COMPUTING SERVICES

A .Iaas (Infrastructure as a service) .

Ihe most basic cloud-service model & according to the IETF(Internet Engineering Task Force), providers of IaaS offer computers – physical or (more often) virtual machines – and other resources.

B. Saas (Software as a service)

In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis. SaaS providers generally price applications using a subscription fee.

C. Paas(Platform as a service)

In the PaaS models, cloud providers deliver a computing platform, typically including operating system, programming language execution environment, database, and web server.
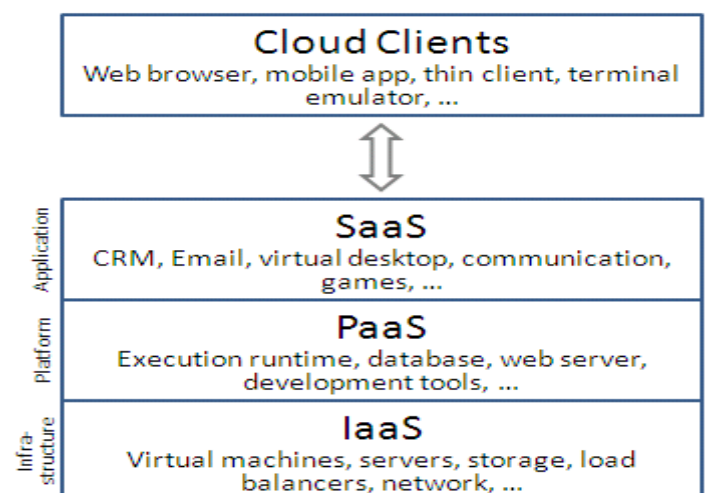

Fig1. Cloud Computing Services

III. CLOUD COMPUTING DEPLOYMENT MODEL

Based on the domain cloud is divided into three parts:

A. *Public Cloud*

Public clouds may be used by individuals or an organization based upon their requirements and necessities. It is type of cloud which can be access from anywhere in the world and can be accessed by anyone. Examples of this cloud are Amazon's or Google's cloud which are open to all after specific SLA between user and provider.

B. *Private Cloud*

In this type of cloud the specific organization's or company's employee can only get access and it will be accessible only within organization's premises and by authenticating each and every user, it is not open to all.

C. *Hybrid Cloud*

Hybrid Clouds (combination of both private and public clouds)-This types of cloud are combination of both public as well as private cloud.

IV. LOAD BALANCING

Load Balancing is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded Load balancing. Load balancing in cloud computing provides an efficient solution to various issues residing in cloud computing environment set-up and usage. Load balancing has two major tasks, one is the resource provisioning or resource allocation and other is scheduling in distributed environment. Efficient provisioning of resources and scheduling of resources as well as tasks will ensure:

- Resources are easily available.
- Resources are efficiently utilized under condition of high/low load.
- Cost of using resources is reduced.
- Load balancing achieves high system throughput and short response time.

According to the initiation of process load balancing algorithm can be of three categories:

- Sender Initiated: If the load balancing algorithm is initialized by the sender.
- Receiver Initiated: If the load balancing algorithm is initiated by the receiver.
- Symmetric: It is the combination of both sender initiated and receiver initiated

Depending on the current state of the system, load balancing algorithms can be divide into two categories as given in :

- Static: It doesn't depend on the current state of the system. Prior knowledge of the system is needed
- Dynamic: Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach.

In the cloud computing environment various load balancing algorithm have been proposed such as Honeybee-based load balancing technique, Active Clustering, Random Sampling, Active Monitoring Load Balancer, Throttled Load Balancer, WCAP, JIQ, CLBVM. The main focus is to achieve high throughput and short response time.
Generally load balancing algorithm is of two types:

A. *Static load balancing algorithm*

B. *Dynamic load balancing algorithm*

A. *Static load balancing algorithm*

Static algorithms are mostly suitable for homogeneous and stable environment but whenever attributes are dynamically changing, results produces are not good. This algorithm allocate the tasks of a parallel program to workstations based on either the load at the time nodes are allocated to some task, or based on an average load of our workstation cluster . Algorithms proposed to achieve load balancing in static environment cannot adapt to the run time changes in load.

B. *Dynamic load balancing algorithm*

Dynamic algorithm are more flexible and cloud providers install heterogeneous environment and cannot rely on prior knowledge.

In a distributed system, dynamic load balancing can be done in two different ways: distributed and non-distributed. In the distributed one, the dynamic load balancing algorithm is executed by all nodes present in the system and the task of load balancing is shared among them. The interaction among nodes to achieve load balancing can take two forms: cooperative and non-cooperative. In the first one, the nodes work side-by-side to achieve a common objective, for example, to improve the overall response time, etc. In the second form, each node works independently toward a goal local to it, for example, to improve the response time of a local task.

Dynamic load balancing algorithms of distributed nature usually generate more messages than the non-distributed ones because each of the nodes in the system needs to interact with every other node. In non-distributed type, either one node or a group of nodes do the task of load balancing. Non-distributed dynamic load balancing algorithms can take two forms: centralized and semi-distributed. In centralized the load balancing algorithm is executed only by a single node in the whole system: the central node. This node is solely responsible for load balancing of the whole system. The other nodes interact only with the central node. In semi-distributed form, nodes of the system are partitioned into clusters, where the load balancing in each cluster is of centralized form. A central node is elected in each cluster by appropriate election technique which takes care of load balancing within that cluster.

V. EXISTING LOAD BALANCING TECHNIQUES

In the cloud computing environment various load balancing algorithm have been proposed such as Honeybee-based load balancing technique, Active Clustering, Random Sampling, Active Monitoring Load Balancer, Throttled Load Balancer, WCAP, JIQ, CLBVM. The main focus is to achieve high throughput and short response time.

Zhao et al, proposed a distributed load balancing algorithm COMPARE_AND_BALANCE based on sampling to reach an equilibrium solution which decreases the virtual machine migration time .

Menasce et al. discussed the concept of cloud computing, its advantages, disadvantages and described several existing cloud computing platforms.

Randles et al. [8], introduced three load balancing algorithm in large-scale complex systems are an extended honeybee foraging algorithm, a biased random sampling on a random walk procedure and Active Clustering.

Kuo-Qin Yan et al. [3] proposed a scheduling algorithm. Their algorithm combined the capabilities of both OLB (Opportunistic Load Balancing) [15] and LBMM (Load Balance Min-Min) [16] scheduling algorithms, and is comparatively more efficient and maintain the load balance of the system.

Zhang Bo et al. [5], proposed an algorithm which adds capacity to the dynamic balance mechanism for the cloud. The experiments demonstrate that the algorithm has obtained a better load balancing degree and used less time in loading all tasks

Round Robin algorithm provides load balancing in static environment. In this the resources are provisioned to the task on first-cum-first-serve (FCFS- i.e. the task that entered first will be first allocated the resource) basis and scheduled in time sharing manner.

Radojevic [9] proposed an improved algorithm over round robin called CLBDM (Central Load Balancing Decision Model) [14]. It calculates the execution time of task.

Hemant S. Mahalle, Parag R. Kaveri and Vinay Chavan[12] proposed a "Active monitoring load balancer algorithm" which maintains information about VMS and the number of requests allocated to each VM. When a request is come, it will assigned to the Vm which is least loaded, if more then first one is selected. Active VM load balancer return id to Data Center Controller data center controller sends the request to the VM identified by that id.

Shridhar G. Domanal and G.Ram Mohana Reddy [8] have proposed a "Modified throttled algorithm" which is a local optimized load balancing approach for distributing incoming job request uniformly between the servers or virtual machines. The performance is analyzed using CloudAnalyst simulator and compared with existing Round Robin and Throttled algorithm. In throttled algorithm the client first requests the load balancer to find a suitable Virtual Machine to perform the required operation, then load balancer check the VM which can access the load easily and perform task given by client.

Recently Shridhar G. Domanal and G. Ram Mohana Reddy have proposed a VM-Assign load balancing algorithm which allocates the incoming requests to all the available virtual machines in an efficient manner. The performance is analyzed using Cloudsim Simulator and then compared with existing Active-VM load balance algorithm.

Table1. Comparison Table of Load Balancing Algorithms in Cloud Computing Environment

Algorithm	Static environment	Dynamic environment	Distributed balancing
Round- robin	YES	NO	NO
Active Clustering	NO	YES	YES
OLB	YES	NO	NO
CLBDM	YES	NO	NO
Biased Random Sampling	NO	YES	YES
MaxMin	YES	NO	NO
MinMin	YES	NO	NO

Also there are different metrics for the comparision of load balancing algorithm.

Throuhput: It can be estimated by the total number of task executed successfully within a fixed span of time.

Overhead: It includes the overhead due to task movement, inter-processor and inter-process communication.

Fault Tolerance: It measures the capability of node in spite of link failure.

Migration Time: Time required to migrate jobs from one node to another node.

Response Time: It can be measured as, the time interval between sending a request and receiving its response. It should be minimized to boost the overall performance.

Scalability: It is the ability of an algorithm to perform load balancing for a system with any finite number of node.

Performance: It is used to check the efficiency of system.

VI. CONCLUSION

Load balancing is very essential task in cloud computing to achieve efficient utilization of resources. The goal of load balancing is to increase client satisfaction and maximize resource utilization and increase the performance of the cloud system thereby reducing the energy consumed and the carbon emission rate. In this paper we discussed various load balancing algorithm. A comparision has been done based on the different cloud computing environment. Here static load balancing schemes provides easiest simulation and monitoring of environment but it is not suited for heterogeneous clouds whereas dynamic load balancing schemes are best suited for heterogeneous environment. Existing load balancing schemes reduces overhead, response time and improve performance.
Here, dynamic load balancing schemes in distributed environment provide better performance. However, performance of the cloud computing environment can be further maximized, if dependencies between tasks are modeled using workflows and if response time minimum resource utilization is improved.

REFERENCES

[1] Haozheng Ren, Yihua Lan, and Chao Yin, "The Load Balancing Algorithm in Cloud Computing Environment",,IEEE ,2nd International Conference on Computer Science ,China 2012.

[2] N. S. Raghava and Deepti Singh," Comparative Study on Load Balancing Techniques in Cloud Computing" OPEN JOURNAL OF MOBILE COMPUTING AND CLOUD COMPUTING,In Press.

[3] S.-C. Wang, K.-Q. Yan, W.-P. Liao, and S.-S. Wang, "Towards a load balancing in a three-level cloud.computing network," in Computer Science and Information Technology (ICCSIT), 2010 3rd IEEEInternational Conference on, vol. 1, pp. 108–113, IEEE, 2010.

[4] MR.Manan D. Shah, MR.Amit A. Kariyani, MR.Dipak L. Agrawal2010 "Allocation Of Virtual Machines In Cloud Computing Using Load Balancing Algorithm", IRACST - International Journal of Compute Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555, Vol. 3, No.1, February 2013.

[5] Z. Bo, G. Ji, and A. Jieqing, "Cloud loading balance algorithm," in Information Science and Engineering (ICISE), 2nd International Conference on, pp. 5001–5004, IEEE,2010.

[6] Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-ofthe-art and research challenges. Journal of Internet Services and Applications, 1(1), 7–18. doi:10.1007/s13174-010-0007-6.

[7] Xu, Y., Wu, L., Guo, L., & Chen, Z. (2011). An Intelligent Load Balancing Algorithm Towards Efficient Cloud Computing, 27–32.

[8] Shridhar G.Domanal and G.Ram Mohana Reddy, "Load Balancing in Cloud Computing Using Modified Throttled Algorithm", IEEE Randles, M., Bendiab, A. T. & Lamb, D. (2008). Cross layer dynamics in self-organising service oriented architectures. IWSOS, Lecture Notes in Computer Science, 5343, pp. 293-298, Springer.

[9] Radojevic, B & Zagar.M, "Analysis of issues with load balancing algorithms in cloud environment. In proceedings of 34th International Convention on MIPRO, IEEE.

[10] Randles, M., Bendiab, A. T. & Lamb, D. (2008). Cross layer dynamics in self-organising service oriented architectures. IWSOS, Lecture Notes in Computer Science, 5343, pp. 293-298, Springer

[11] S C., Zhang, S., Liu, Q., Xie, J. & Hu, J. (2009). Independent Tasks Scheduling Based on Genetic Algorithm in Cloud Computing.

[12] Dr. Hemant S. Mahalle, Prof. Parag R. Kaveri, Dr.Vinay Chavan, LoadBalancing On Cloud Data Centres" in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, issue 1, January 2013.

[13] Load Balancing computing[EB/OL], http://en.wikipedia.org.

[14] Cloud Computing, http://en.wikipedia.org/wiki/Cloud_computing.

[15] T. D. Braun, H. J. Siegel, N. Beck, L. L. B¨ol¨oni, M. Maheswaran, A. I. Reuther, J. P. Robertson M D. Theys, B. Yao, D. Hensgen, et al., "A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems," Journal of Parallel and Distributed computing, vol. 61, no. 6, pp. 810–837, 2001.

[16] T. Kokilavani and D. Amalarethinam, "Load balanced min-min algorithm for static meta-task scheduling in grid computing.," International Journal of ComputerApplications, vol. 20, no. 2,2011.

[17] D. A Menasc´e and P. Ngo, "Understanding cloud computing: Experimentation and capacity planning," in Computer Measurement Group Conference, 2009.

.