International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Survey on Various Load Balancing Techniques in Cloud Computing Environment

Ravisha Sadhu[1], Jignesh Vania[2]

[1, 2]L.J Institute of Engineering and Technology, Gujarat Technological University, Sarkhej, Ahmedabad, India

Abstract: ***Load balancing is the major issue over cloud. Many techniques have been proposed to improve load balancing in cloud computing environment. The main objective of load balancing technique is to improve performance of system in terms of response time and throughput. The goal of this paper is to provide survey of various load balancing techniques used in cloud computing environment. This paper is divided into three parts. First part gives introduction of cloud computing and load balancing. Second part gives survey of various load balancing techniques proposed by some researchers. Third part gives conclusion based on survey of various load balancing techniques. This paper provides concept behind various load balancing techniques, their merits and demerits.***

Keywords: Cloud Computing,Load balancing,Logarithmic Least Square Method,Genetic Algorithm,Virtual Machine, Selection, Crossover, Mutation

1. Introduction

1.1 Cloud Computing

Cloud computing definition proposed by NIST(National Institute of Standards and Technology)says "Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. network, server, storage and applications and services) that can be rapidly provisioned and released with minimal management efforts or service provider interaction".

Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers.

1.1.1 Service Models

Cloud-computing providers offer their "services" according to different models as shown in fig. 1.

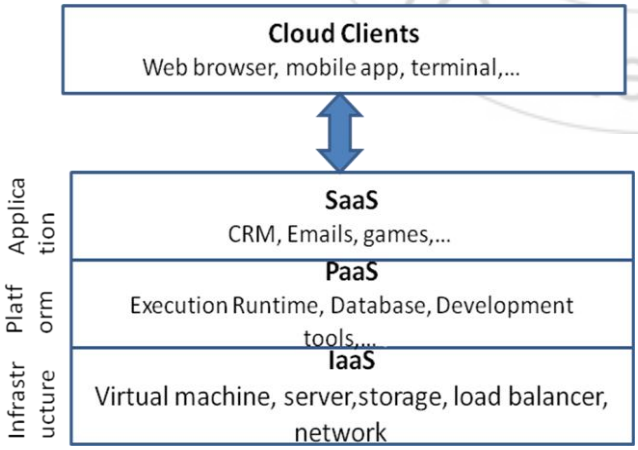

Figure 1: Service Models

1.1.1.1 Infrastructure as a service (IaaS): Infrastructure as a service delivers infrastructure on demand in form of virtual hardware, storage and networking. Virtual machine instances are created on provider's infrastructure. Users are given tools and interfaces to configure software installed in virtual machine. Virtual storage is delivered in form of raw disk space or object store. Virtual networking identifies collection of services that manage networking among virtual instances and their connectivity towards Internet.

1.1.1.2 Platform as a service (PaaS): In the PaaS models, cloud providers deliver a computing platform, typically including operating system, different tools, programming-language execution environment, database, and web server. PaaS provides all the resources to the customers that are required for building applications and create environment where applications are deployed and executed.

1.1.1.3 Software as a Service (SaaS): Saas provides all the application to the consumer which are provided by the providers.Used when existing SaaS services fit user's need. So minimum customization is required.Applications are running on a cloud infrastructure. Interfaces (web browser) are used access the applications. The consumer does not control the internal functions of applications.Applications such as office automation, document management, photo editing, customer relationship management software etc are provided by SaaS.

1.1.2 Types of Clouds

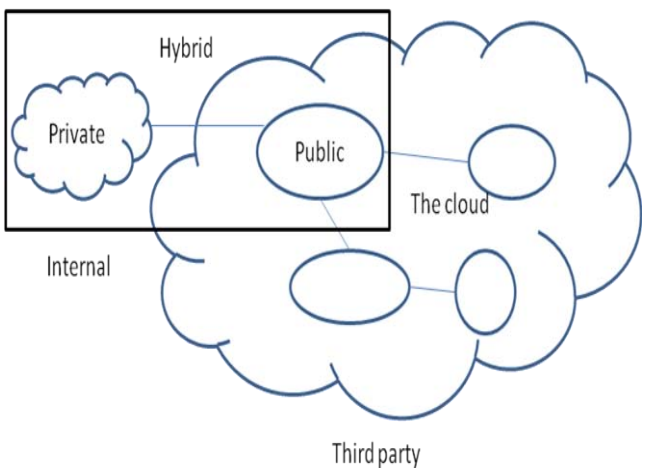

Figure 2: Types of clouds

1.1.2.1 Private cloud: To keep confidential information within organization, institutes or enterprises build and use their own cloud. They may use it to store and manage data of their organization or to provide enough resources on demand basis to its team of employees or clients. Organization owns the hardware and software infrastructure, manages the cloud and controls access to its resources. They offer greatest level of security. Open Stack, VMware and Cloud Stack are some private clouds.

1.1.2.2 Public cloud: Infrastructure is developed by third party service provides and it is provided to consumer on subscription basis. User's data and application are deployed on data centers on vendor's premises.Confidentiality is the major security issue in using public cloud. They are more vulnerable than private clouds. Amazon web services, Google Compute Engine, Microsoft Azure, HP cloud are some of the public clouds.

1.1.2.3 Hybrid cloud: A hybrid cloud is a combination of public and private cloud.When private cloud capacity is not sufficient to meet organization's need, some public cloud resources are leased. So privately owned infrastructure and public cloud resources together serve organization's need.The downside is that the complexity of overall management increases along with security concerns.

1.2 Load Balancing

Load balancing is a technique which is used to equally distribute workload over all processors so that all processeor does the same amount of work and no processor is overloaded.Thus load balancing increases throughput and reduces response time.Load balancing is done by load balancer which accepts multiple requests from users and distribute them across servers on cloud such that no server is overloaded.

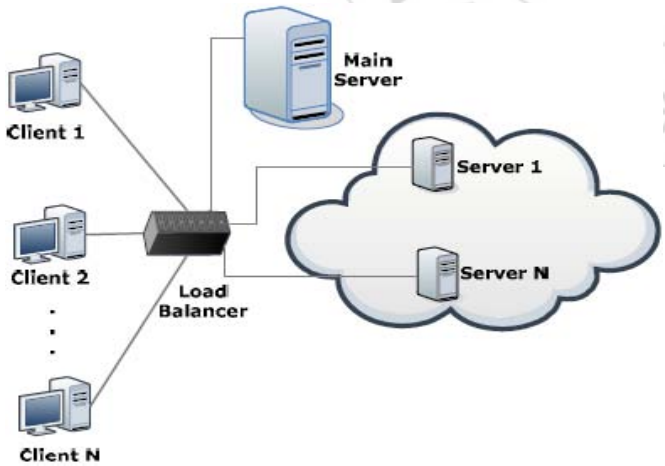

Figure 3: Working of load balancer [7]

1.2.1 Categories of load balancing algorithms:

1.2.1.1 Depending on who initiates the process:
Sender-Initiated: Sender or client identifies the need for load balancing and sender initiates the execution of load balancing algorithm.

a) **Receiver-Initiated:** Receiver or server identifies the need for load balancing and receiver initiates the execution of load balancing algorithm.
b) **Symmetric:** It is the combination of sender-initiated and receiver-initiated types.

1.2.1.2 Depending on current state of system
Static algorithm: Current status of all nodes and their properties are known in advance and based on this prior knowledge the algorithm works.

Dynamic algorithm: Algorithm works according to dynamic changes in state of nodes.

1.2.2 Existing Load balancing algorithms:

1.2.2.1 Round-Robin Algorithm: It is the static load balancing algorithm which uses the round robin scheme for allocating job. It selects the first node randomly and then, allocates jobs to all other nodes in a round robin fashion. Without any sort of priority the tasks are assigned to the processors in circular order.

1.2.2.2 Opportunistic Load Balancing Algorithm: It is a static algorithm. This algorithm deals quickly with the unexecuted tasks in random order to the currently available node. Each task is assigned to the node randomly. The task will process in slow in manner because it does not calculate the current execution time of the node.

1.2.2.3 Min-Min Load Balancing Algorithm: This is static load balancing algorithm. The cloud manager identifies the execution and completion time of the unassigned tasks waiting in a queue. The cloud manager first deals with the jobs having minimum execution time by assigning them to the processors according to the capability to complete the job in specified completion time. The jobs having maximum execution time has to wait for the unspecified period of time.The assigned tasks are updated in the processors and the task is removed from the waiting queue.

1.2.2.4 Max-Min Load Balancing Algorithm: Max Min algorithm works same as the Min-Min algorithm except the following: After finding out the minimum execution time, it deals with tasks having maximum execution time.The assigned task is removed from the list of the tasks that are to be assigned to the processor and the execution time for all other tasks is updated on that processor.

1.2.2.5 The two phase scheduling load balancing algorithm: It is the combination of OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) Scheduling algorithms to utilize better execution efficiency and maintain the load balancing of the system.OLB scheduling algorithm keeps every node in working state to achieve the goal of load balancing and LBMM scheduling algorithm is utilized to minimize the execution time of each task on the node thereby minimizing the overall completion time. This algorithm works to enhance the utilization of resources and enhances the work efficiency.

1.2.2.6 Honeybee Foraging Behavior load balancing Algorithm: It is a nature inspired load balancing technique which helps to achieve load balancing across heterogeneous virtual machine of cloud computing environment and maximize the throughput. First the current workload of the VM is calculated, then it decides the VM states whether it is over loaded, under loaded or balanced. The priority of the task is taken into consideration after removed from the overloaded VM which are waiting for the VM. Then the task is scheduled to the lightly loaded VM. It reduces the response time of VM and also reduces the waiting time of task.

1.2.3 Metrics for Load Balancing

1.2.3.1 Throughput: The total number of tasks that have been executed is called throughput. A high throughput is required for better performance of system.

1.2.3.2 Associated overhead: The amount of overhead that is produced by execution of load balancing algorithm. Minimum overhead is expected for successful implementation load balancing algorithm.

1.2.3.3 Fault Tolerant: The ability of system to perform correctly even at time of failure at any node in the system.

1.2.3.4 Migration Time: The time taken in migration or transfer of a task from one machine to another.Minimum migration time is required for better performance of system.

1.2.3.5 Response Time: The minimum time that a distributed system executing a specific load balancing algorithm takes to respond. Response time should be as minimum as possible for better performance of system.

1.2.3.6 Resource Utilization: It is the degree to which the resources of the system are utilized.A good load balancing algorithm provides maximum resource utilization.

1.2.3.7 Scalability: It determines the ability of the system to accomplish load balancing algorithm with a restricted number of processors or machines.

1.2.3.8 Performance: If all the above parameters are optimal then it will improve the performance of the system.

2. Literature Survey

Ajit et al.[1] proposed VM level load balancing approach in cloud environment in which load assignment factor is calculated for each host based on its configuration. First VMs are mapped on host in which first VMs are mapped on host with highest load assignment factor, then on host having load assignment factor lower than that and so on. After that individual requests are assigned to VMs in which the requests are first assigned to VM mapped on host having highest load assignment factor, then on hosts having load assignment factor lower than that and so on. The proposed algorithm reduces average response time.But drawback is that when new task arrives during all VMs are busy, then task goes into waiting state.

Soni et al.[2] proposed a novel approach for load balancing in cloud data centers in which Central Load Balancer is connected to all users and virtual machines and it will balance load among all virtual machine. The Central Load Balancer calculates priorities of all virtual machines based on their CPU speed(MIPS) and memory. The Central Load Balancer assigns requests first to virtual machine having highest priority and which is available to execute request, then it assigns requests to virtual machine having next lower priority and which is available and so on. This approach provides quick and reliable load balancing and minimizes response time.But drawback is that if no virtual machine is available, the arrived request goes into queue.

Domanal et al.[3] proposed novel VM-assign algorithm in which VM-assign Load Balancer maintains a index/assign table which stores information about number of requests currently allocated to each VM. When requests arrive, index table is parsed. If VM with least number of requests is found and was not used in previous assignment, then requests are allocated to that VM, otherwise next least-loaded VM is searched in index table and request is allocated to it and so on. This algorithm prevents both under-utilization and over-utilization of virtual machines.

Chandrasekaran et al.[4] proposed load balancing approach for virtual machine resources in cloud using genetic algorithm which calculates load of node before deploying VM on node and finds solution which gives best load balancing.GA has three steps: Selection, Crossover and Mutation. In selection, the fitness of individuals in current population is determined and individuals having highest fitness are kept in child population. Then selection probability of individual is determined based on fitness function. So chromosome with high probability will have more chances to get selected. The crossover combines parents selected in selection step to form new solution which keeps same VMs from two parents and different VMs from two parents are distributed to least loaded physical machine set until all VMs are distributed. Mutation selects two physical machines based on mutation probability and swaps VMs between those two machines to form new solution. This algorithm minimizes number of VM migrations and provides better load balancing. The drawback is that depending upon problem instance size, population size should be within certain range to get optimal solution, otherwise algorithm gives suboptimal solution or takes more time to provide optimal solution.

Pilavare et al.[5] proposed a novel approach for improving performance of load balancing using genetic algorithm in cloud computing environment. First pairwise comparison matrix is generated which compares all VMs in terms of cost. Then for assigning priorities to VM, Logarithmic Least Square Method is used in which first all values in each row of pair wise comparison matrix are multiplied, then nth root of this product is taken and finally all values in each row are normalized, so this resultant matrix is called priority matrix which shows priorities of all VMs. Then these prioritized VMs are given as input to genetic algorithm. This algorithm minimizes response time and make span of given task set.

3. Conclusion

This paper gives survey on various techniques for load balancing in cloud computing environment. All these techniques improve performance in terms of response time.But Genetic algorithm also minimizes number of VM migrations and gives better load balancing. This genetic algorithm performs better than round-robin and greedy algorithm under stable and variable load conditions. By assigning the priority to the VM's and giving the prioritized input to the genetic algorithm can improve the response time of the system and can also minimize the make span of given task set. Genetic algorithm performs better than other load balancing algorithms because it uses natural strategy to find a solution.

References

[1] Mr. M. Ajit ,Ms. G. Vidya , "VM Level Load Balancing in Cloud Environment",4th ICCCNT , IEEE – 31661,Tiruchengode, India, July 4-6,2013.
[2] Gulshan Soni, Mala Kalra2, "A Novel Approach for Load Balancing in Cloud Data Center", IEEE International Advance Computing Conference (IACC) , ISBN- 978-1-4799-2572-8,pg. 807-812,2014.
[3] Shridhar G.Damanal, G. Ram Mahana Reddy, "Optimal Load Balancing in Cloud Computing By Efficient Utilization of Virtual Machines", ISBN- 978-1-4799-3635-9,IEEE -2014.
[4] Chandrasekaran K.and Usha Divakarla,"Load Balancing of Virtual Machine Resources in Cloud Using Genetic Algorithm", ICCN 2013, Elsevier Publications 2013,pp. 156–168.
[5] Mr. Mayur S. Pilavare , Mr. Amish Desai ,"A Novel Approach Towards Improving Performance of Load Balancing Using Genetic Algorithm in Cloud Computing",IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems-2015,ISBN- 978-1-4799-6818-3,IEEE-2015.
[6] Rajwinder Kaur, Pawan Luthra, "Load Balancing in Cloud Computing",DOI: 02.ITC.2014.5.92 , Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing, ITC ,2014.
[7] Sukhvir Kaur,Supriya Kinger,"Review on Load Balancing Techniques in Cloud Computing Environment", International Journal of Science and Research (IJSR), Paper ID: 02014812,Volume 3,Issue 6,June 2014, pg. 2499-2504.
[8] Dharmesh Kashyap, Jaydeep Viradiya,"A Survey Of Various Load Balancing Algorithms In Cloud Computing", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,VOLUME 3,ISSUE 11,ISSN 2277-8616,NOVEMBER 2014,Pg. 115-119.

Author Profile

Ravisha Sadhu received the B.E degree in Information Technology Engineering from L.J Institute of Engineering and Technology in 2014.She will complete her M.E in Information Technology Engineering from L.J Institute of Engineering and Technology in 2016.

Jignesh Vania received the B.E. degrees in Information Technology from U. V. Patel Collage of Engineering & Technology in 2003 and received Master Degree in Information Technology from Shantilal Shah Engineering Collage, Bhavnagar under Gujarat Technological University.Currently he is working as assistant professor at L.J Institute of Engineering and Technology