Load Balancing on the Cloud Environment

Shaivya Jindal[1*], Neeta Sharma[2]

[1*]Department of Computer Science, School of Engineering and Technology, Noida International University, India
[2]Department of Computer Science, School of Engineering and Technology, Noida International University, India

Abstract— Cloud Computing is the growing technology. It is a means of having multiple computing processes, making do and delivering software and services. Cloud Computing is basically a collection of resources and services of computation integrated together and is provided to the end users on pay-as-needed basis. cloud computing is a well know structured model that provides services, where resources and data are retrieved from cloud service provider through a platform of internet web-based tools and application. As multiple users of the cloud generates multiple request for resources present on the cloud which may cause a deadlock. So in order to avoid deadlocks on the cloud it is important to divide the loads generated, of all the VMs among themselves. Load balancing actually means the same. So the aim of this paper is to discuss load balancing and have a comparative study of different load balancing algorithms and there implementations.

Keywords—Load Balancing, Cloud computing, Data centers , Resource allocation

I.   INTRODUCTION

Cloud computing is the one of the fastest growing technology in the world. Cloud Computing provide a virtualized network to all its users to access to all the services and resources available on the cloud.

It allows the clients to use the application without it hardware/software set up and access their own files and data from anywhere all around the globe with a connection with internet The geographical location of the end user does not matter until it is connected to the web. When multiple requests generated together on any system for services then many a times it ends up on a deadlock, creating a state of hault serving none of the request, a situation of deadlock. Load balancing is the answer to this problem. There are various technique which can be used to solve this problem. There are soft computing algorithms like Particle Swarm Optimization (PSO).

Load balancing is one of the current methods that will assist the networks and resources by providing a high throughput and least response time. In cloud computation environment, resource allocation or we can say load balancing takes place broadly in two steps. Firstly the load balancer assigns the requested to physical computers at the time of uploading or raising the request.

When an application or resources receives multiple requests on the cloud, each of these requests must be allocated to a particular application instance to maintain a balance of computational load among the instances of that specific application or resources. Working with different algorithms results in different results.

II.   RELATED WORK

This section deals with the discussion focused on the researches literary work for load balancing in cloud computing. Different researchers deals with different view of solving the problem of load on the cloud some of the renounced researchers theory are discussed.

Seyed Mohssen Ghafari et al. proposed a load balancing algorithm Bee-MMT (artificial bee colony algorithm-Minimal Migration Time) for power consumption management in cloud computation This algorithm is the combination of two individual algorithms Firstly Artificial Bee Colony algorithm (ABC) to detect over-weighted hosts and Then MMT algorithm to transfer the load from the over weighted virtual machines to the others to decrease their load. While doing this it can detect under-weighted hosts and

then transfer load from all virtual machines allocated to these hosts and then toggle them to the sleep mode.

Shridhar G. Damanal et al. has introduced a modified throttled algorithm for load balancing in cloud computing [2]. This modified throttled algorithm basically deals with the fact that how incoming jobs are assigned to the available virtual machines on the cloud effectively and efficiently. This works on maintaining an index table of virtual machines and their states on the cloud. In this modified version of algorithm makes an attempt to improve the response time and to achieve efficient usage of all the available virtual machines. Virtual Machines are being selected on the basis of the state of Virtual machine. If Virtual machines is available the request is approved

else denied by returning -1 to the data center. When the other request is generated virtual machine chose the next of already assigned.

III. METHODOLOGY

Relevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory). The most important focus of load balancing is to distribute computing resources and services. Whatever services needed by the user is based on the usage. Resources available on the cloud not only share the service and required resources but also reallocated at run time. Client is unaware about the format of service delivering system and also cloud automatically mange all the resources present on it. So, a number of distributed host machines are grouped in a cloud. And as multiple servers are present on the Cloud computing environment and also multiple virtual machines, datacenters and storage devices etc. as resources are also present and there interconnected between themselves in a reliable approach. Whenever a request is generated on the cloud by the client then the host available on the cloud automatically creates a virtual machine to respond the request of the client or the end user. So each host machine variable on the cloud, according to the load, creates virtual machines randomly on end users' demand. Some host machines may get flooded with the load and some remain light-weighted. Load can be of any type. Now Load balancing ensures distribution of cloud resources and service efficiently and effectively. Loads are distributed evenly, most of the cases, with the help of multiple load balancing algorithms. Thereby we can maximize resource utilization and reduce the waiting time which is one of the most crucial issues at any cloud computing environment.

 Goals of Load Balancing as discussed by multiple authors includes improvement in performance ,Increase the flexibility of the system to engulf modification and respond accordingly, Building a fault tolerant system by creating backups so the data and the services provides cannot able to delete the data of any user all-around the globe.

Based on process orientation load balancing algorithms are classified as:

Firstly Sender Initiated the request in which the client or the end user generates the request until a receiver on the cloud is assigned to it to receive his request. Secondly Receiver Initiated the request, the receiver sends a request to acknowledge a client or end user who is ready to share his request. Lastly the Symmetric which is a combination of both sender and receiver initiated request of load balancing algorithm.

IV. PROPOSED APPROACH: (PSO)

It should include important findings discussed briefly. Wherever necessary, elaborate on the tables and figures without repeating their contents. Interpret the findings in view of the results obtained in this and in past studies on this topic. State the conclusions in a few sentences at the end of the paper. However, valid colored photographs can also be published. This section includes a review of several algorithms which focus on load balancing in cloud computing environment. Load balancing algorithms have also been used in grid Computing, task scheduling, distributed computing etc. It is already known that task assignment, load balancing has been found to be NP-complete problems. The swarm optimization algorithm is able to get schedule features than other genetic algorithm. L. many other researchers have talked about PSO algorithm in grid computing and has got the better results which can further enhance the working on the cloud. ACO in distributed system [3] provides a less efficient result than PSO algorithm. This algorithm improves the quality as well as running time than others.

So, Particle Swarm Optimization is suggested for the load balancing problem in cloud computing for better results. The PSO is based on the principle of bird flocking which search their food randomly on the search space. During their search many things get changed. But they keep on working on the operation to achieve its goal. In PSO two parameters are considered for its evaluation. Firstly Pbest i.e. the best position obtained so far and secondly Gbest i.e. best value obtained. During their searching each node work according to their direction and speed based on current position, velocity, Pbest and Gbest. Performance of the whole sistym can also be measured.

V. CONCLUSIONAND FUTURE SCOPE

 In cloud computing , dynamic load balancing is one of the most important field for researchers as it can enhance the better performance on the server.

 One of the main goals of load balancing is to increase end user satisfaction and maximize resource utilization and minimize the waiting time and response time increase the throughput substantially increase the performance of the cloud system and also minimize the request rejection thereby reducing the energy consumed.

 In this paper we studied various algorithms for load balancing in cloud computing. A study about the PSO algorithm in different research areas generated the best result. This study focused on the different aspect of PSO that can be used for optimize load balancing in cloud computing.

Therefore, In future work, we can work on PSO in a more better way on the clod itself.

REFERENCES

[1] Seyed Mohssen Ghafari, Mahdi Fazeli, Ahmad Patooghy, Leila Rikhtechi, "Bee-MMT: A Load Balancing Method for Power Consumption Management in Cloud Computing," in proc. 6th International Conference on Contemporary Computing (IC3), IEEE, pp. 76-80, August 2013

[2] Shridhar G. Domanal, G. Ram Mohana Reddy, "Load Balancing in Cloud Computing Using Modified Throttled Algorithm," in proc. International Conference on Cloud Computing in Emerging Markets (CCEM), IEEE, pp. 1-7, October 2013.

[3] A. Salman, "Particle Swarm Optimization for Task Assignment Problem," Microprocessors and Microsystems, pp. 363-371, Nov 2002.

[4] Abhijit A Rajguru, S.S. Apte, "A Comparative Performance Analysis of Load Balancing Algorithms In Distributed Systems Using Qualitative Parameters", International Journal of Recent Technology and Engineering, Vol. 1, Issue 3, August 2012

[5] David Escalnte and Andrew J. Korty, "Cloud Services: Policy and Assessment", EDUCAUSE Review, Vol. 46, July/August 2011.

[6] JianzheTai,JueminZhang,JunLi,WaleedMeleis and NingfangMi "A R A: Adaptive Resource Allocation for Cloud Computing Environments under Bursty Workloads" 978-1-4673-0012-4/11 ©2011 IEEE.

Authors Profile

Shaivya Jindal pursed her Bachelor of Technology from ABES engineering college, ghaziabad. She have been working on networking as well as cryptography and published multiple research papers in International journels recognized by IEEE, Springer, EIA etc . titled as "Prevention of man in the middle attack by using honeypot" and "An efficient and secure micro-payment transaction using shell cryptography" currently I am pursuing my Master of technology from SET , Noida International University.

Neeta Sharma pursed her bachelor of technology, Master of Technology as well as Ph.D. She is currently working as Assistant Professor in Department of computer science, Noida International University, India since 2013. She is a efficient member of IEEE & IEEE computer society since 2013.. Her main research work focuses on cloud Cloud Computing, Cloud Security and Privacy, Big Data Analytics, Data Mining etc. She has 10 years of teaching and research experience .