1st International Conference on Applied Soft Computing Techniques
22 & 23.04.2017
In association with
International Journal of Scientific Research in Science and Technology

An Efficient Load Balancing Scheme in Cloud Computing

D. Pradhiba*[1], R.Kanniga Devi[2]
*[1]P.G Student, Department of Computer Science and Engineering, Kalasalingam University, Srivilliputtur,Tamilnadu, India[1]
[2]Assistant Professor, Department of Computer Science and Engineering, Kalasalingam University,Srivilliputtur,Tamilnadu,India[2]

ABSTRACT

In computer science domain, cloud computing is one of the booming technologies. In cloud computing platform the clients and cloud data centers are distributed across the world. The biggest challenge of cloud data centers is servicing and handling millions of request from the clients. Here the major challenge is load balancing the cloud datacenter. Load balancing is a technique which is used to spread the load of system fairly across all the servers and avoiding overloading. The prime goal of balancing the load of system is to minimize fail-over, attain better resource utilization and enabling scalability etc. In this paper, an algorithm is proposed to distribute load of the system uniformly across cloud data center for balancing the load which decreases average execution time of cloud data centers.

Keywords: Cloud computing, Cloud Datacenter, Load balancing

I. INTRODUCTION

The cloud computing is one of the most trending technologies in IT domain. It is a technique of handling and pooling services like servers, data base, storage, software and more over the internet based on the user's need or demand. Users can get the resources from the data centers as per their requirements from anywhere through an internet connected computer or hand held devices. One of the challenging task in cloud computing is load balancing used to allocate work load among the data centers. Datacenters are physical machines that has the responsibility to complete the request and demand of cloud users. So load balancing is required to manage the load across data centers, reduce the overload, improve performance, minimize average execution time and provide better resource utilization. Load balancing can minimize the response time and maximize the user's satisfaction. It also increase the source utilization and limit the energy consumption. The classification of load balancing algorithms are of two types: static and dynamic. Load balancing is static when it needs previous data of system. It is dynamic when it requires current data of system.

In static load balancing algorithm [1], load is assign to machine as per their capacity of processing. It has preliminary knowledge of the system like performance, processing energy, memory and data about client requirement. It doesn't consider the dynamic changes at run time. Popular static algorithms like Round Robin algorithm, Weighted Round Robin algorithm are used for balancing the load.

In dynamic load balancing algorithm,[1] preliminary information about the system is not required, because runtime condition of system is collected for assign load to every machine. Popular dynamic algorithms like Least Connection algorithm, Weighted Least Connection algorithm are used for balancing the load. The dynamic algorithm gives higher performance and fault tolerance, but it is difficult.

Dynamic load balancing helps to improve the efficacy, maintaining the stability of the system and adjusting system modification. In this paper, an algorithm is proposed to distribute load of the system uniformly

across cloud data center for balancing the load which decreases average execution time of cloud data centers.

II. RELATED WORK

An Gulshan soni, et al.[1] proposed the central load balancer technique that has the responsible to take decision for balancing the workload among virtual machines. Central Load Balancer (CLB) have VM table which contain VM state, priority of VM .The allocation of request to VM is based on the state and priority of the VM. It tried to avoid the situation of under loading and over loading of virtual machines.

Fei Ma ,et al.[2] used TOPSIS method for allocation of virtual machine in cloud data centers to achieve load balancing. In this system, two decision have to be taken for balancing the load when the system is over loaded or under loaded. The VM monitor observes the over and under loaded of the system. The PM have to be migrated to selected VM and which VM have to be migrated when the PM is problematic. In large scale of cloud environment this system achieve better load balancing with less migration time.

Vasudha Arora, et al.[3] analyzed the performance of round robin algorithm, equally spread execution load and throttled load balancing policies across virtual machine. The throttled load balancing policy reduces average response time and data processing time of data centers. But it's not suitable when data centers are reconfigured for route the traffic dynamically.

Feilong tang et al.[4] proposed a novel dynamical load-balanced scheduling (DLBS) approach to maximize the network throughput by balancing workload dynamically. The results indicate that this DLBS approach is able to balance transmission traffic dynamically and globally so that it improves throughput and bandwidth utilization ratio.

Surbhi kapoor et al.[5] The problems of existing throttled and modified throttled algorithms are not considering the resource specific demands of the tasks and also they are not suitable for heterogeneous VMs environment. These problems have been addressed in this proposed approach by clustering the VMs.

Youssef fahim, et al.[6] proposed the hybrid algorithm to improve the availability, performance and maximize the VM utility. This hybrid algorithm takes into account the current load of VM and current task status of VM during execution to avoid probable blocking.

Ekta gupta, et al.[7] proposed a load balancing technique which is based on Ant Colony Optimization .The under loaded and over loaded servers are identified and load balancing operation are performed by Ant Colony Optimization between identified datacenter server to increase the throughput and availability of resource.

Joseph doyle, et al.[8] considered carbon emission, average service request time and electricity cost as balancing factor. The voronoi partitions method was used to identified which request is allocated to which data centers as per balancing factors in order to balance the workload of system.

Chung-Cheng Li, et al.[9] proposed a novel neural network-based dynamic weighted round-robin scheduling algorithm and SLA-aware two-level decentralized load balancer (tldlb) architecture to achieve dynamic load balancing in cloud datacenter. To distribute the large number of incoming request to different VM a novel neural network-based dynamic load balancing algorithm was developed and SLA violation rate was reduced by using a novel decentralized load balancing architecture.

G.Shobana, et al.[10] proposed preemptive task scheduling which follows the foraging behavior of honey bees to achieve load balancing in cloud datacenter. Tasks are allocated to VMs based on the foraging behavior of honey by preemptive task scheduling which reduces the make span. It also considered task priority to minimize latency and maximize throughput.

Hitesh A. Bheda, et al.[11] used the flexible load sharing algorithm which partitioned the system into domain for load balancing in cloud computing. It overcomes the message loss problem and it provides unlimited resource capacity. Biased random selection method was used by this FLS algorithm for the scalability and performance improvement.

Ankit Kumar et al.[12] proposed a modified active monitoring load balancer which distributes the incoming jobs among the VM efficiently. Whenever a request comes from client it has sent to modified active monitoring load balancer which have VM table. Then it checks the VM table for allocating the requests to VM. If VM is available it allocates the request to the VM. Otherwise it waits until the VM became available state. The VM table contains VM ids, state of VMs, VM loads and memory of VMs. This algorithm does not consider the VM reliability and energy awareness.

J.Octavio, et al.[13] used the distributed problem solving techniques for balancing the heterogeneous loads optimally. The CProtocol and EProtocol was used by agent for manage the load in distributed manner, it reduce the cost of energy consumption.

III.PROPOSED WORK

In this section, we will be discussing our proposed approach. In this proposed approach, cluster heads were elected from group of data centers depends upon the higher number of connection with other data centers. The data centers were clustered based on the threshold value or capacity which is assigned by the cloud service provider. The input of each client is assigned to cluster head based on the shortest path between client and cluster head. In previous load balancing schemes there is a variation in load of each data centers (some data center may have high load and some have low load) because of shortest path. To overcome this our proposed approach assigns capacity to each data center. If the capacity of data center exceeds then the remaining client job has transferred to the nearest data center which has a capacity to serve. The cluster heads were communicated with other to know the status and count of services provided. By this new approach data centers can achieve uniform load balancing.

This proposed scheme is implemented in c platform by using graph theory concept. The step-wise procedure of the proposed approach is given as follows:

Steps to achieve uniform load balancing:

Clustering data centers:

Step 1: Datacenter Network is consider as a undirected weighted graph G= (V, E, W) .Where V is datacenter, E is connection between datacenters and W is weight age of the connection between datacenters.

Step 2: In order to identified the cluster head (CH) the degree value should be calculated for each datacenter (V)

Step 3: The datacenters which are having maximum degree value have to be considered as cluster heads (CH).

Step 4: Then the remaining datacenters (V') have to be allocated to cluster heads (CH) based on the shortest path distance, the shortest path distance is calculated by weight age of the connection between datacenter (V') and cluster heads.

Step 5: The limit of datacenters (V') which are assigned to the cluster heads(CH) are based on the threshold value.

$$\text{Threshold} = \frac{\text{Number of datacenters (V)}}{\text{Number of clusters (K)}}$$

Where the number of clusters (K) should be defined by cloud provider

Client assignment to clusters:

Step 6: Each client has to be allocated at most uniformly to the clustered datacenters (CL) based on the threshold value.

$$\text{Threshold} = \frac{\text{Number of clients(C)}}{\text{Number of clusters (k)}}$$

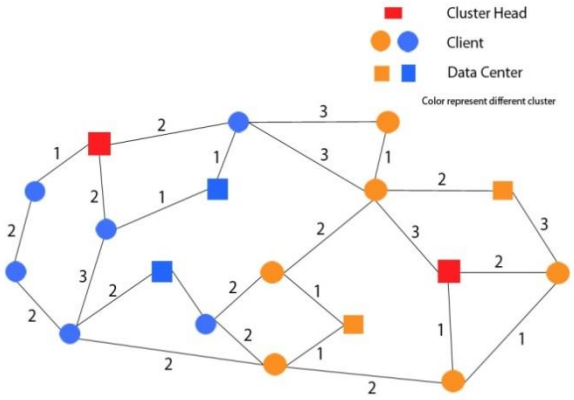

Figure 1: Model of proposed system

IV. EXPERIMENT AND RESULTS

The experiment has been carried out by implementing the proposed algorithm and the existing algorithm in C. A random task generator have been used. There is no limit to the number of data centers generated. Implementation parameters have been mentioned in table 1. Our proposed algorithm has been compared with K-Nearest neighbor algorithm and shared nearest neighbor clustering algorithm on the basis of execution time and response time.

Table 1 Implementation Parameters

Number of Data centers	9-100
Number of Clusters	3
Number of Clients	18-300

Comparison of proposed and other two algorithms have been shown in table 2. By varying the data centers and clients the comparison table is shown for the three algorithms on the basis of execution time and response time.

Table 2 Comparison Table

Clients	**Neighbor size K-value**	**Cluster Time (sec)**		
		Proposed algorithm	**K-Nearest Neighbor algorithm**	**Shared Nearest Neighbor algorithm**
69	23	184	212	201
135	45	320	351	335
165	55	452	485	471
196	32	500	528	515
291	97	579	616	599

Figure 2 Comparison of existing and proposed algorithm

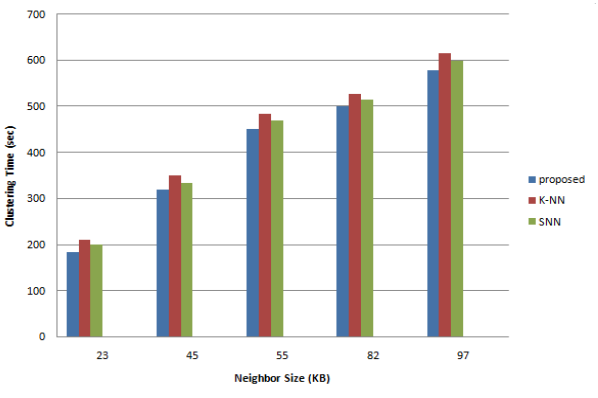

Figure 2 shows the comparison of existing algorithms with proposed algorithm. It clearly shows that the clustering time is lesser than the previous algorithms.

V.CONCLUSION

The biggest challenge in cloud computing is balancing the load of cloud datacenters. So load balancing is required to manage the load across data centers to avoid overload of datacenter, improve performance of datacenters, minimize average execution time and provide better resource utilization. we present a load balancing algorithm to distribute load of the system uniformly across cloud data center for balancing the load and decreases average execution time of cloud data centers.

VI.REFERENCES

[1] Gulshan Soni, Mala Kalra,"A Novel Approach for Load Balancing in Cloud Data Center", IEEE International Conference on Advance Computing, pp. 807-812,2014.

[2] Fei Ma, Feng L1U, Zhen L1U, "Distributed Load Balancing Allocation Of Virtual Machine In Cloud Data Center", IEEE International Conference On Computer Science And Automation Engineering, pp. 20– 23.

[3] Vasudha Arora, S. S.Tyagi, "Performance Evaluation of Load Balancing Policies across Virtual Machines in a Data Center", International Conference on Reliability, Optimization and Information Technology,pp.384-387,2014.

[4] Feilong Tang, Laurence T. Yang, Can Tang, Jie Li, Minyi Guo, "A Dynamical And Load-

Balanced Flow Scheduling Approach For Big Data Centers in Clouds", IEEE Transactions On Cloud Computing, Volume: 1, Issue: 99, 2016.

[5] Surbhi Kapoor, Dr. Chetna Dabas, "Cluster Based Load Balancing In Cloud Computing", IEEE Eighth International Conference On Contemporary Computing, pp. 76 - 81, 2015.

[6] Youssef Fahim, Elhabib Ben Lahmar, El Houssine Labriji Saea Ouahabi, "The Load Balancing Improvement of A Data Center By A Hybrid Algorithm in Cloud Computing", IEEE Third International Colloquium In Information Science And Technology,pp.141-144,October 2014

[7] Ekta Gupta, Vidya Deshpande, "A Technique Based On Ant Colony Optimization for Load Balancing in Cloud Data Center", IEEE 13th International Conference On Information Technology, pp. 12-16, 2014 .

[8] Joseph Doyle, Robert Shorten, Donal O'Mahony, "Stratus: Load Balancing the Cloud for Carbon Emissions Control", IEEE Transactions On Cloud Computing, Vol. 1, No. 1,pp. 2168-7161, June 2013.

[9] Chung-Cheng Li, Kuochen Wang, "An SLA-Aware Load Balancing Scheme for Cloud Data Centers", IEEE International Conference On Information Networking, pp. 58 - 63,2014

[10] G.Shobana, M.Geetha, .R.C.Suganthe, ",Nature Inspired Preemptive Task Scheduling for Load Balancing in Cloud Datacenter", IEEE Information Communication and Embedded Systems, 2014 International Conference on , pp. 1-6, 2014.

[11] Hiren H. Bhatt, Hitesh A. Bheda, "Enhance Load Balancing Using Flexible Load Sharing In Cloud Computing", IEEE 1st International Conference On Next Generation Computing Technologies, pp.72-76, September 2015.

[12] Ankit Kumar, Mala Kalra, "Load Balancing in Cloud Data Center Using Modified Active Monitoring Load Balancer", IEEE International Conference on advances in Computing, Communication and Automation, pp. 1-5, Apr 2016.2016

[13] J.Octavio, Gutierrez-Garcia, Adrian, Ramirez-Nafarrate, "Collaborative Agents for Distributed Load Management in Cloud Data Centers Using Live Migration of Virtual Machines", IEEE Transactions On Services Computing, Vol. 8, No. 6,pp. 1939-1374, November/December 2015.

[14] Wang Yong, Tao Xiaoling, He Qian, Kuang Yuwen, "A Dynamic Load Balancing Method Of Cloud-Center Based On SDN ",IEEE Transactions On Network Technology And Application, Volume: 13, Issue: 2, pp. 130 – 137, February 2016 .

[15] Jia Zhao, Kun Yang, Xiaohui Wei, Yan Ding, Liang Hu, "A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment," IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, pp. 1045-9219, February2016

[16] Lei Yu, Liuhua Chen, Zhipeng Cai, Haiying Shen, Yi Liang, Yi Pan, "Stochastic Load Balancing for Virtual Resource Management In Datacenters", IEEE Transactions On Cloud Computing, Vol. , No. , November 2014.