A Load Balancing Algorith for the Data Centers to Optimise Cloud Computing Application

[1]VVV Satya, [2]K.Durga Devi
[1,2]Dept. of CSE, KIET, Kakinada, AP, India

Abstract
Cloud computing is a category of network-based computing environment that provides the customers with computing resources as a service over a network on their demand. Load balancing in cloud is the process of distributing the work load among various nodes in a distributed system for better resource utilization and job response time. The load balancer calculates the value in the particular time-span and uses this value to estimate the virtual machine availability for the next time span. Load balancing ensures that all the processor in the system or every node in the network does approximately perform the equal amount of work at any instant of time. It is a process of assigning 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. It can be observed CPU utilization, throughput etc. will be improvised while balancing the load to virtual machines on the basis of utilization of resources on an instant time. This research work has proposed a novel technique to analyse the performance of optimized load balancer. The offered technique is based on condition which will provide high availability to clients, and estimating the required measures by varying the interval time. In proposed Optimized Load Balancer technique, we tried to avoid the situation of over loading and under loading of virtual machines. The Optimized Load Balancer manages load distribution among various virtual machines and assigns load corresponding to their priority and states. In this way this technique efficiently shares the load of user requests among various virtual machines.

Keywords
Cloud Computing, Load Balancing, Cloud Analyst, Efficiency, DC Time, Cost, Throughput.

I. Introduction

Cloud Computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common Internet protocols and networking standards. It is distinguished by the notion that resources are virtual and limitless and that details of the physical systems on which software runs are abstracted from the user.
In an effort to better describe cloud computing, a number of cloud types have been defined. Two different classes of clouds: those based on the deployment model and those based on the service model. The deployment model tells you where the cloud is located and for what purpose. Public, private, community, and hybrid clouds are deployment models..
In Cloud Computing, scheduling plays a vital role in Cloud computing environment is basically composed of physical servers, virtual machines, data centers, storage devices etc which are interconnected together in an efficient manner. Nowadays, computing systems heavily rely on Virtualization [1] technology as it delivers services by providing a platform for optimizing resources in a scalable manner. Thereby improves the power efficiency of the data centers and thereby enabling the allocation of multiple virtual machines (VMs) to a single physical server [2] thereby allows the effective utilization of resources.
It 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. A load balancing algorithm which is dynamic in nature does not consider the previous state or behavior of the system, that is, it depends on the present behavior of the system. The important things to consider while developing such algorithm are: estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones. This load considered can be in terms of CPU load, amount of memory used, delay or Network load. Load balancing is one of the hottest issues in cloud computing. There can be various types of load like CPU load, delay or network load and memory capacity. Load balancing in cloud is the process of distributing the work load among various nodes in a distributed system for better resource utilization and job response time. [3]. The load balancer calculates observed in the particular timespan and use this value to estimate the virtual machine availability for the next time span [4]. Load balancing ensures that all the processor in the system or every node in the network does approximately perform the equal amount of work at any instant of time [5].

There are basically two types of load balancing which are as given below:

A. Static Load Balancing

The load balancing decision is made at compile time when resource requirements are estimated. The advantage of static algorithm lies in its simplicity according to both implementation and overhead as there is no need to constantly monitor the nodes for performance statistics.

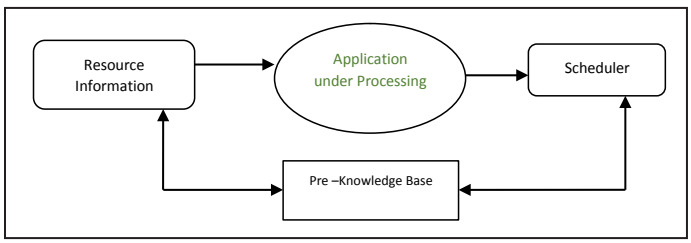

Fig. 3: Static Load Balancing

B. Dynamic load balancing

The dynamic load balancing algorithms allows the distribution of work among nodes at run-time; they use current or load information while making distribution decisions. Dynamic load balancing algorithms are distributed in nature; such algorithms frequently produce more messages than the non-distributed ones because each node is required to interact with every other node in the entire system. The advantage, of such approach is that in case

if one or more nodes in the network fail, it will not hamper the total load balancing process; instead it will influence the system performance to a little extent.

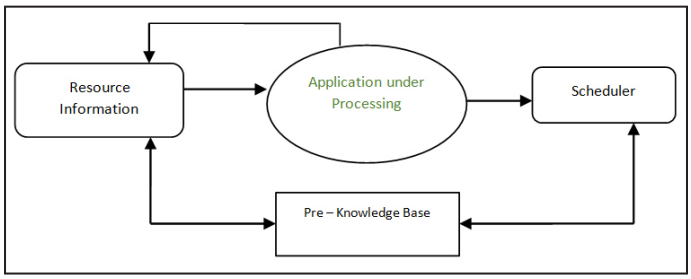

Figure 4: Dynamic Load Balancing

The rest of the paper is organized as follows: Section II describes the literature review. Section III describes the proposed methodology part. Section IV shows the Results and Discussion part. Section V brings the conclusion and future extent of the paper.

II. Literature Review

Abhishek Kumar Tiwari et al. (2016) [1] The author tried to introduce swarm intelligence algorithm which offers various types of algorithm such as particle of swarm optimization, ant colony optimization, and many more derived algorithms for optimization. The proper management of load balancing improves the efficiency of throughput. Swarm intelligence play an important role in load balancing technique. Author discussed only two divisible load scheduling algorithms that can be applied to clouds, but there are still other approaches that can be applied to balance the load in clouds. The performance of the given algorithms can also be increased by varying different parameters.

AwatifRagmani et al. (2016) [2] Author tries to deal with performance issues in Cloud computing is the load balancing. In fact, an efficient load balancing contributes to the decrease of costs and maximizes availability of resources. Through this paper, author studied Ant Colony Optimization Technique which has treated the load balancing in the Cloud computing. In a second step, he proposes an improved load balancing architecture and algorithm for Cloud computing which aims to allow a better response time.

Gagandeep Kaur et al. (2016) [3] Author proposed the ACCLB to balance the load on the cloud and compared it with the existing load balancing methods such as Vector Dot, Join idle queue. The main objective will be to balance the load on cloud and to reduce the energy consumption as compared to previous, on the cloud by using proposed method. Also we have to prove that our proposed technique is more efficient for load balancing and energy consumption on cloud.

PalakShrivastava et al. (2016) [4] Author laid stress on the main unease on Load Balancing in Cloud Computing environment. This load could be CPU load, recollection volume, or system load. Load balancing safeguards all the systems, as every node in the system does roughly the equal quantity of exertion at any instantaneous of time. There are many approaches to determination the tricky of load balancing in cloud environment, hence by investigation of such procedures with their countless returns, restrictions and issues a new and competent practice for Load Balancing is instigated in future.

Rajeshkannan et al. (2016) [5] In this paper author discusses numerous load balancing algorithms so as to improve resource utilization and quality of services in cloud computing environment. The various load balancing algorithms are compared with quality of service parameters in a cloud network. This analysis helps to identify the effective load balancing algorithm for optimizes resource use, maximizes throughput, minimizes response time, and avoids overload. This paper reviewed on a methodological analysis of various load algorithms for services in a cloud network by concentrating on balancing the loads.

The authors identified a swarm-based algorithm to satisfy the user requirement for continuing the service by distributing the workload in balance manner to acquire maximum resource utilization and reduce system idle time.

III. Proposed Methodology

Load balancing in cloud is the process of distributing the work load among various nodes in a distributed system for better resource utilization and job response time. The load balancer calculates observed in the particular time-span and use this value to estimate the virtual machine availability for the next time span. Load balancing ensures that all the processor in the system or every node in the network does approximately perform the equal amount of work at any instant of time. It is a process of assigning 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. It can be observed CPU utilization, throughput etc. will be improvised while balancing the load to virtual machines on the basis of utilization of resources on an instant time. The aim of this research work is to save resources so that we can allocate the resources to more number of processes in order to increase productivity and make it environmental friendly.

Proposed Algorithm:

The steps of proposed algorithm are as follow:

1. Current Resource Load Balancer maintains a table that contains virtual machine id (VMid), states (BUSY/AVAILABLE), priority of VM and timestamp. Initially, all Virtual Machines are in available state and having timestamp 0.
2. Data Center Manager receives a new request.
3. Data Center Manager queries the Current Resource Load Balancer for next allocation and increase the timestamp of those requests which are arriving.
4. Current Resource Load Balancer parses the table from top to find the highest priority virtual machine, available state.

If found:

- The Current Resource Load Balancer returns the id of that machine to the Data Center Manager.
- The Center Controller inform the VM which is identified by VMid and at the same time Load Balancer will update the timestamp of arriving process and if timestamp increasing Load balancer will update the priority of process.
- Data Center Manager notifies the Load Balancer of new allocation of resources.
- Current Resource Load Balancer updates the table accordingly.

If not found:

- The Current Resource Load Balancer returns -1.
- The Data Center Manager queues the request and reduce its

timestamp.

5. When the VM finishes processing of requests, and the Data Center Manager receives the response cloudlet, it notifies the Current Load Balancer of the VM de-allocation.
6. The Data Center Manager checks if there are any waiting requests in the queue. If there are, it continues from step 3.
7. Continue from step 2.

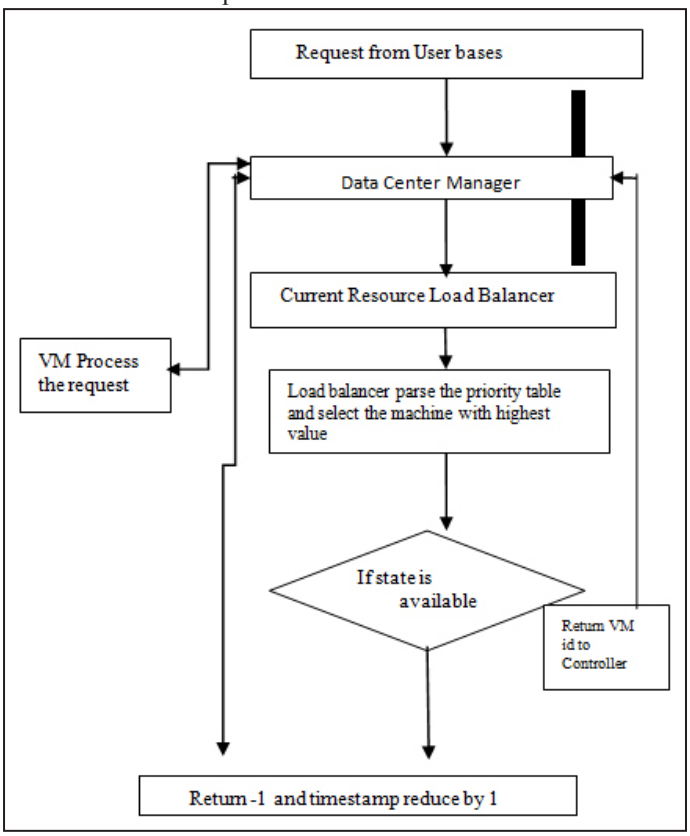

Fig. 3: Flowchart of Proposed Algorithm

IV. Results And Discussion

To implement the load balancing technique cloud analyst tool is being used. After GridSim, CloudSim provides novel support for modeling and simulation of virtualized Cloud-based data center environments such as interfaces for virtual machines (VMs), memory, storage, and bandwidth and model data centers, service brokers and allocation policies but it became apparent to have an easy to use tool with a user-friendly GUI that lead to Cloud Analyst tool. The load balancing technique is used by the VmLoadBalancer component of the cloud analyst

The response time for six user bases and three data centres is being simulated and executed by cloud analyst respectively, where average response time and data Centre processing time is calculated as:

$$Response\ time = (Response_{Recieved_{\Delta t}} - Request_{sent_t})$$

$$Response_{Recieved_{\Delta t}} = \frac{\sum data_{size}}{\sum_{c=1}^{c=k} request}$$

Where is the total number of requests received by the virtual machine at time is the job request sent to the virtual machine at time t. $\sum data\_size$ is the total data processed.

$$DC\_Processing\_Time = DC\_end\_time - DC\_start\_time$$

Where start time is the starting time when the data Centre is being started processing the task and is measured by gridsim_clock. End time is the ending time when data Centre is finished the serving of the allotted task and is measured by gridsim_clock.

Table 1: Overall Response Time Evaluation

Overall Response Time		
Scenarios	Optimized Load Balancer	ACO Load Balancer
6UB, 3DC	191.38	200.05
16 UB 2DC	48.71	54.15
10UB,5DC	168.89	184.14
15UB,5DC	211.91	217.84
12UB,4DC	69.32	71.61

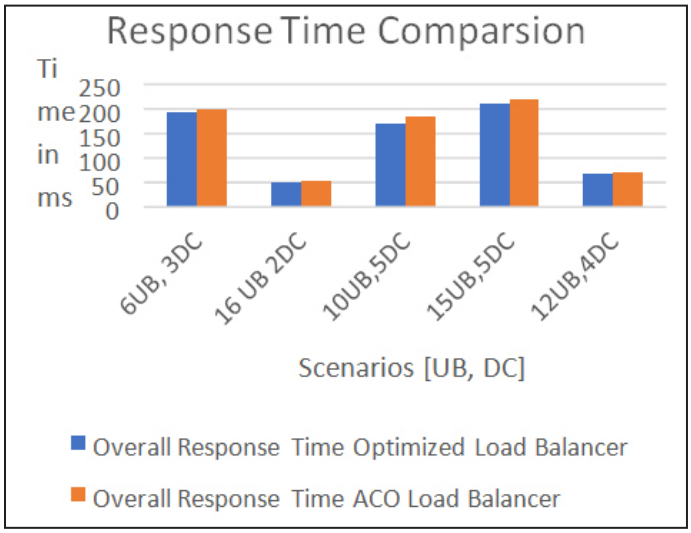

Fig. 4: Represent the number of virtual machines varies for response time.

Fig. 4 indicates about comparison of number of virtual machines varies for response time between existing and the proposed method wherever y-axis indicate Time in m/sec as well as x- axis indicates number of nodes Here, blue line indicates the proposed technique and orange line indicate the previous one. In our case the proposed Number of virtual machines for response time are comparatively lower than existing one.

Table 2: DC Processing Time Evaluation

DC Processing Time		
Scenarios	Optimized Load Balancer	ACO Load Balancer
6UB, 3DC	0.39	4.07
16 UB 2DC	0.64	0.52
10UB,5DC	0.53	11.4
15UB,5DC	0.42	0.53
12UB,4DC	0.5	0.96

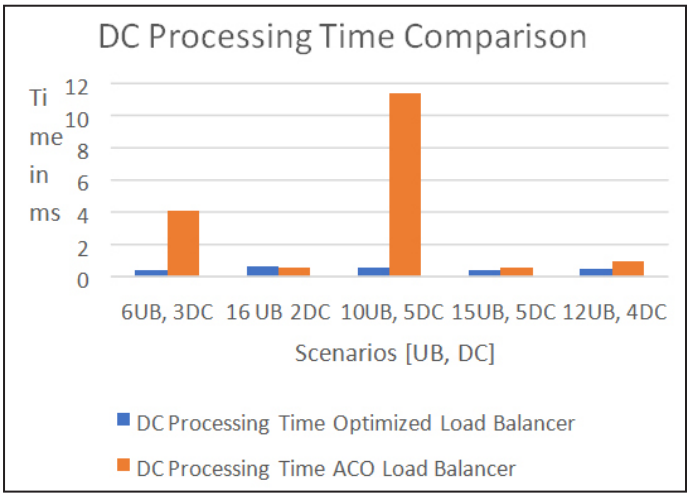

Fig. 5: Represent the number for Data Centre Processing Time Comparison

Figure 5 indicates about comparison of number of virtual machines varies for response time between existing and the proposed method wherever y-axis indicate Time in m/sec as well as x- axis indicates number of nodes Here, blue line indicates the proposed technique and orange line indicate the previous one. In our case the proposednumber for Data Centre Processing Time Comparison are comparatively lesser than existing one.

Table 3: Total Cost Evaluation

Total Cost		
Scenarios	Optimized Load Balancer	ACO Load Balancer
6UB, 3DC	2.35	3.99
16 UB 2DC	3.07	3.7
10UB,5DC	5.81	5.89
15UB,5DC	4.07	6.51
12UB,4DC	4.66	4.84

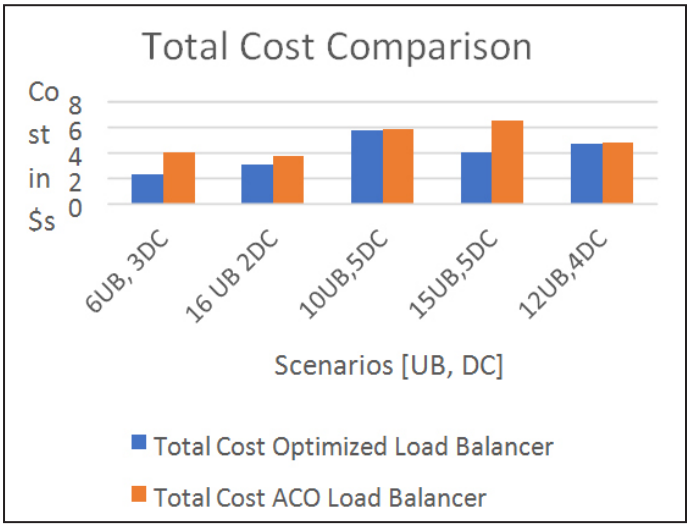

Fig. 6: Represent the figure of virtual machines differ for cost.

Fig. 6 indicates about comparison of the figure of virtual machines differ for throughput between existing and the proposed method wherever y-axis indicate Time in m/sec as well as x- axis indicates number of nodes Here, blue line indicates the proposed technique and orange line indicate the previous one. In our case the proposed figure of virtual machines differ for cost Comparison are comparatively decreasing than existing one

Table 4: Throughput Evaluation

Throughput		
Scenarios	Optimized Load Balancer	ACO Load Balancer
6UB, 3DC	2.56	0.31
16 UB 2DC	2.52	2.03
10UB,5DC	3.19	0.09
15UB,5DC	2.4	2.39
12UB,4DC	2.2	2.06

Fig. 7 indicates about comparison of the figure of virtual machines differ for throughput between existing and the proposed method wherever y-axis indicate Time in m/sec as well as x- axis indicates number of nodes. In our case the proposed figure of virtual machines differ for throughput Comparison are comparatively increasing than existing one

V. Conclusion and Future Scope

The proposed load balancing algorithm optimized Load Balancer will be able to perform quick and reliable load balancing in cloud computing environment through utilization of all virtual machines according to their computing capacities. In the proposed technique, every request from user bases arrive at Data Center Controller. Data Center Controller queries the Central Load Balancer for allocation of requests. Central Load Balancer maintain a table that consist of id, states and priority of virtual machines. Central Load balancer parses the table and find out highest priority virtual machine, then check its states and if its states available then return that virtual machine id (VMid) to Data Center Controller. Finally, Data Center Controller assigns the request to that VMid that is provided by proposed. The Optimized Load Balancer is connected to all users and virtual machines present in cloud data center through Data center Controller. The Central Load Balancer calculates the priorities of virtual machines based on their CPU speed (MIPS) and memory. This research work has proposed a novel technique to analyse the performance of optimized load balancer. The offered technique is based on condition which will provide high availability to clients, and estimating the required measures by varying the interval time. In proposed Optimized Load Balancer technique, we tried to avoid the situation of over loading and under loading of virtual machines. The Optimized Load Balancer manages load distribution among various virtual machines and assigns load corresponding to their priority and states. In this way this technique efficiently shares the load of user requests among various virtual machines.

References

[1] S.C. Wang, K.Q. Yan, W.P. Liao and S.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, 2010, pp. 108-113.

[2] HadiGoudarzi, Mohammad Ghasemazar, and Massoud Pedram, "SLAbased Optimization of Power and Migration Cost in Cloud Computing" in 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2012.

[3] Pradeep K.Sinha, "Distributed operating Systems Concepts and Design" IEEE Computer Society Press.
[4] E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good, "The cost of doing science on the Cloud: the Montage example," Proc.of the 2008 ACM/IEEE Conference on Supercomputing, IEEE, Nov. 2008.
[5] Rimal, B. Prasad, E. Choi and I. Lumb, "A taxonomy and survey of cloud computing systems." In proc. 5th International Joint Conference on INC, IMS and IDC, IEEE, 2009.
[6] Wickremasinghe, Bhathiya, Rodrigo N. Calheiros, and RajkumarBuyya. "Cloud analyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications." In Advanced Information Networking and Applications (AINA), Volume.5, No. 2, pp.446-452, 2010, Chicago.
[7] Hu, Jinhua, JianhuaGu, Guofei Sun, and Tianhai Zhao. "A scheduling strategy on load balancing of virtual machine resources in cloud computing environment." In Parallel Architectures, Algorithms and Programming (PAAP), Volume.30, No.7, pp. 89-96, IEEE, 2010.
[8] Padhy, Ram Prasad. "Load balancing in cloud computing systems." PhD dissertation., National Institute of Technology, Rourkela, 2011.
[9] Mishra, Ratan, and AnantJaiswal. "Ant colony optimization: A solution of load balancing in cloud." International Journal of Web & Semantic Technology,Volume.3, No. 2, pp.33-50, 2012.
[10] Xu, Gaochao, Junjie Pang, and Xiaodong Fu. "A load balancing model based on cloud partitioning for the public cloud." Tsinghua Science and Technology , Volume.18, No.1, pp.34-39, 2013.
[11] Begum, Suriya, and C. S. R. Prashanth. "Investigational Study of 7 Effective Schemes of Load Balancing in Cloud Computing." International Journal of Computer Science Issues, Volume.10, No.6, pp.343-353, 2013.
[12] Gao, Yongqiang, Haibing Guan, Zhengwei Qi, Yang Hou, and Liang Liu. "A multi-objective ant colony system algorithm for virtual machine placement in cloud computing." Journal of Computer and System Sciences ,Volume.79, No.8, pp. 1230-1242 , 2013.
[13] Khan, Rafiqul Z., and Md F. Ali. "An Efficient Diffusion Load Balancing Algorithm in Distributed System." International Journal of Information Technology and Computer Science (IJITCS),Volume.6, No.8, pp.65-71, 2014.
[14] Mohanty, Suneeta, Prasant Kumar Pattnaik, and Ganga BishnuMund. "A Comparative Approach to Reduce the Waiting Time Using Queuing Theory in Cloud Computing Environment." International Journal of Information and Computation Technology, Volume.4, No.5, pp.469-474, 2014.
[15] Singh Aarti, Dimple Juneja and Manisha Malhotra. "Autonomous agent based load balancing algorithm in cloud computing." International Conference on Advanced Computing Technologies and Applications, Volume.45, No.4, pp.832841, 2015.
[16] Tripathi, Akanksha and Sandeep Raghuwanshi. "Cloud Task Scheduling in Hybrid Environment using Ant Colony Optimization." International Journal of Advanced Engineering and Global Technology, Volume.3, No.7, pp.934-940, July 2015.
[17] D. Saranya and L. SankaraMaheswari "Load Balancing Algorithms in Cloud Computing: A Review" International Journal of Advanced Research in Computer Science and Software Engineering, Volume.5, No.7, pp.1107-1111, July 2015.
[18] P. P. G. Gopinath and S. K Vasudevan, "An In-depth Analysis and Study of Load Balancing Techniques in the Cloud Computing Environment," 2nd International Symposium on Big Data and Cloud Computing, Volume.50, No.6, pp.427-432, 2015
[19] M. Berwal and C. Kant,"Load Balancing in Cloud Computing," International Journal of Computer Science ,Volume.6, No.2, pp.52-58, 2015