International Journal of Engineering Technology and Applied Science
(ISSN: 2395 3853), Vol. 1 Issue 6 December 2015

A Contemplate on various Load Balancing schemes in Cloud Computing Environments

Vipin Verma[1], Praveen Kataria[2]

M-tech scholar ASCT, Bhopal[1], Asst. Prof. Dept. of CSE, ASCT Bhopal[2]
vipinvermacs@gmail.com

ABSTRACT:- Cloud computing, which provides online resources as a service to users, brings a technology revolution in IT world. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Load Balancing is the one of the most important parts of the current virtual environment. In the case of cloud computing environments there were various challenges are there in the load balancing techniques like security, fault tolerance etc. Many algorithms had been proposed for finding the solution of load balancing problem in these fields. But very few algorithms are proposed for cloud computing environment. Since cloud computing is significantly different from these other types of environments, separate load balancing algorithm need to be proposed to cater its requirements. This paper describes a survey on load balancing schemes in cloud environments and their corresponding advantages, disadvantages and performance metrics are studied in detail.

KEYWORDS:- Cloud computing, Load balancing, Virtualization.

I. INTRODUCTION

In Cloud Computing the main concerns involve efficiently assigning tasks to the nodes such that the effort and request processing is done as efficiently as possible [3], while being able to tolerate the various affecting constraints such as heterogeneity and high communication delays.

The load balancing is triggered when host is over loaded then some virtual machine need to be migrated from over loaded host to another host to balance the load among the hosts. In cloud computing the load balancing is very important task because in cloud environment number of physical machines are used to serve the users requests simultaneously. Load balancing is the process of equally distributing the total load on all running physical machines in cloud data center. Due to improper scheduling, a situation may occur where some of the nodes are heavily loaded while other nodes are idle or doing very little work which reduces system performance. So load among the various physical machines must be equally distributed in other words, at any instant of time every node should do approximately the equal amount of work.

II. LOAD BALANCING ALGORITHMS

A. Liable on who initiated the process, load balancing algorithms can be of three categories as given in [5].

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

B. Liable on the current state of the system, load balancing algorithms can be divided into two categories as given in [4].

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 this paper we present a survey of the current load balancing algorithms developed specifically to suit the Cloud Computing environments. We provide an overview of these algorithms and discuss their properties. In addition, we compare these algorithms based on various properties.

Thus cloud computing environment requires a load balancing algorithm which could cater to dynamic service demands of users while providing optimized load balancing. Following parameters

are available in literature for measuring efficiency of a load balancing algorithm in CC environment [6].

A general load balancing algorithm works as follows shown in the flow chart in fig:1

Step-1 Load Evaluation:

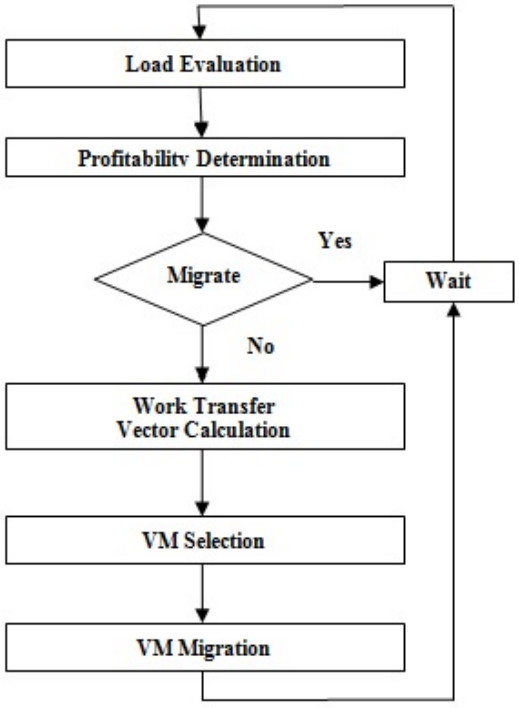

Fig:1 Flow Chart of Load Balancing Algorithm

III. FACTORS EFFECTING LOAD BALANCING

There are various issues while dealing with load balancing in a cloud computing environment. Each load balancing algorithm must be such as to achieve the desired goal. Some algorithms aims at achieving higher throughput, some aims at achieving minimum response time, some other aims to achieve maximum resource utilization while some aims at achieving a trade-off between all these metrics [5].

1. Geographical distribution of nodes:- The geographical distribution of the nodes matters a lot in the overall performance of any real time cloud computing systems, especially in case of the large scaled applications like Facebook, Twitter, etc.

2. Behavior of algorithms:- Any algorithm concerning load balancing is designed, based on the state or behavior of the system, which may be static or dynamic.

3. Algorithm complexity:- The complexity of any load balancing algorithm affects the overall performance of the system. Some-times the algorithm is complex, but is better in terms of throughput and resource utilization.

4. Traffic analysis:- For any load balancing algorithm, it is very important to analyze the traffic flow in real-time scenarios over different geographic regions, and then balance the overall workload accordingly.

IV. DISCUSSION AND COMPARISION

In the previous section categories of load balancing techniques proposed by various researchers have been discussed. Table.1 gives a comparative analysis of different load balancing techniques with respect to different performance parameters [7].

1. Throughput:- This metric is used to estimate the total number of tasks, whose execution has been completed successfully. High throughput is necessary for overall system performance.

2. Overhead:- It associated with any load balancing algorithm indicates the extra cost involved in implementing the algorithm. It should be as low as possible.

3. Fault Tolerance:- It measures the capability of an algorithm to perform uniform load balancing in case of any failure. A good load balancing algorithm must be highly fault tolerable.

4. Migration Time:- It is defined as, the total time required in migrating the jobs or resources from one node to another. It should be minimized.

5. 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.

6. Resource Utilization:- It is used to ensure the proper utilization of all those resources, which comprised the whole system. This factor must be optimized to have an efficient load balancing algorithm.

7. Scalability:- It is the ability of an algorithm to perform uniform load balancing in a system with the increase in the number of nodes,

according to the requirements. Algorithm with higher scalability is preferred.

8. Performance:- It is used to check, how efficient the system is. This has to be improved at a reasonable cost, e.g., reducing the response time though keeping the acceptable delays.

Table.1 Comparison of existing Load Balancing Techniques

Metrics/ Techniques	Throughput	Overhead	Fault tolerance	Migration time	Response time	Resource Utilization	Scalability	Performance
Round Robin [8]	YES	YES	NO	NO	YES	YES	YES	YES
Dynamic Round Robin [9]	YES	YES	YES	YES	NO	YES	NO	NO
PALB [10]	YES	YES	YES	YES	YES	YES	NO	NO
Active Monitoring [11]	YES	YES	NO	YES	YES	YES	YES	NO
FAMLB [12]	YES	YES	YES	YES	NO	YES	YES	YES
Min-Min [13]	YES	YES	NO	NO	YES	YES	NO	YES
Max-Min [14]	YES	YES	NO	NO	YES	YES	NO	YES
OLB+LBMM [11]	NO	NO	NO	NO	NO	YES	NO	YES
Throttled [16]	NO	NO	YES	YES	YES	YES	YES	YES
Honeybee Foraging [3]	NO	NO	NO	NO	NO	YES	NO	NO
Active Clustering [3]	NO	YES	NO	YES	NO	YES	NO	NO
Biased Random Sampling [3]	NO	YES	NO	NO	NO	YES	NO	YES

V. CONCLUSION

In this paper, we have observed various load balancing schemes in the Cloud Computing environment. We focused on major issues which must be taken into consideration while designing any load balancing algorithm. We have discussed the already proposed algorithms by various researchers in literature, their advantages and disadvantages. A comparison has been done on the basis of different criteria like scalability, network overhead, resource utilization, algorithm complexity, fault tolerance, response time, etc. In future we will focus on designing algorithms which will maintain a better trade-off among all performance parameters.

REFERENCES

[1]. Geethu Gopinath P P, Shriram K Vasudevan, "An in-depth analysis and study of Load balancing techniques in the cloud computing environment", in Proc. ELSEVIER 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), April 2015.

[2]. Aarti Singh, Dimple Juneja, Manisha Malhotra, "Autonomous Agent Based Load Balancing Algorithm in Cloud Computing", in Proc. ELSEVIER International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), March 2015.

[3]. Randles, M., D. Lamb and A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA), Perth, Australia, April 2010.

[4]. Rimal, B. Prasad, E. Choi and I. Lumb, "A taxonomy and survey of cloud computing systems", in Proc IEEE 5th International Joint Conference on INC, IMS and IDC, 2009.

[5]. Ali M. Alake el, "A Guide to Dynamic Load Balancing in Distributed Computer Systems", IJCSNS International Journal of Computer Science and Net work Security, VOL.10 No.6, Jun e 2010.

[6]. J. Hu, J. Gu, G. Sun, T. Zhao. A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment in Proc. PAAP, 2010, pp. 89-96.

[7]. N. S. Raghava, Deepti Singh. "Comparative Study on Load Balancing Techniques in Cloud Computing", Open Journal Of Mobile Computing And Cloud Computing, VOL.1 No.1, August 2014.

[8]. S. Subramanian, G. Nitish Krishna, M. Kiran Kumar, P. Sreesh, and G. Karpagam, "An adaptive algorithm for dynamic priority based virtual machine scheduling in cloud.," International Journal of Computer Science Issues (IJCSI), vol. 9, no. 6, 2012.

[9]. C.-C. Lin, P. Liu, and J.-J. Wu, "Energy-aware virtual machine dynamic provision and scheduling for cloud computing," in Cloud Computing (CLOUD), 2011 IEEE

International Conference on, pp. 736–737, IEEE, 2011.

[10]. J. M. Galloway, K. L. Smith, and S. S. Vrbsky, "Power aware load balancing for cloud computing," in Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp. 19–21, 2011.

[11]. B. Wickremasinghe, "Cloudanalyst: A cloudsim-based tool for modelling and analysis of large scale cloud computing environments," MEDC Project Report, vol. 22, no. 6, pp. 433–659, 2009.

[12]. Z. Nine, M. SQ, M. Azad, A. Kalam, S. Abdullah, and R. M. Rahman, "Fuzzy logic based dynamic load balancing in virtualized data centers," in Fuzzy Systems (FUZZ), 2013 IEEE International Conference on, pp. 1–7, IEEE, 2013.

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

[14]. Y. Mao, X. Chen, and X. Li, "Max–min task scheduling algorithm for load balance in cloud computing," in Proceedings of International Conference on Computer Science and Information Technology, pp. 457–465, Springer, 2014.

[15]. 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 IEEE International Conference on, vol. 1, pp. 108–113, IEEE, 2010.