www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 1 January 2015, Page No. 9874-9881

An Overview of Load balancing Techniques in Cloud Computing Environments

Bhatt Hirenkumar .H, Prof. Hitesh A. Bheda
RK University, Rajkot
RK University, Rajkot
bhiren48@yahoo.com
hitesh.bheda@rku.ac.in

Abstract- **The present era has witnessed tremendous growth of the internet and various applications that are running over it. Cloud computing is the internet based technology, emphasizing its utility and follows pay-as-you-go model, hence became so popular with high demanding features. Load balancing is one of the interesting and prominent research topics in cloud computing, which has gained a large attention recently. Users are demanding more services with better results. Many algorithms and approaches are proposed by various researchers throughout the world, with the aim of balancing the overall workload among given nodes, while attaining the maximum throughput and minimum time. In this paper, various proposed algorithms addressing the issue of load balancing in Cloud Computing are analyzed and compared to provide a gist of the latest approaches in this research area.**

***Index Terms*—** Cloud Computing; Green Computing; Load Balancing; Virtualization

INTRODUCTION

Cloud Computing is the most recent emerging paradigm promising to turn the vision of “computing utilities” into reality, it provides a flexible and easy way to store and retrieve huge data without worrying about the hardware needed.

As the number of users on cloud increases, the existing resources decreases automatically which leads to the problem of delay between the users and the cloud service providers.

Thus, the load balancing comes into the picture. The traffic over the network must be dealt smartly such that the situation in which some nodes are overloaded and some other are under loaded should never arise.

To overcome this situation, many load balancing algorithms are proposed by researchers, with their own pros and cons.

In this paper, an overall review of the current load balancing algorithms in the Cloud Computing environment is presented. The ideas of each algorithm are discussed and finally summarized as an overview.

The issues related to load balancing in cloud computing environmentare discussed in Section 2.

In Section 3, the current literature and the load balancing algorithmin cloud proposed by various researchers were discussed.

In section 4, the proposed approaches were discussed and compared. Finally, in Section 5 we concluded the paper and also presented some ideas which can be implemented in the future.

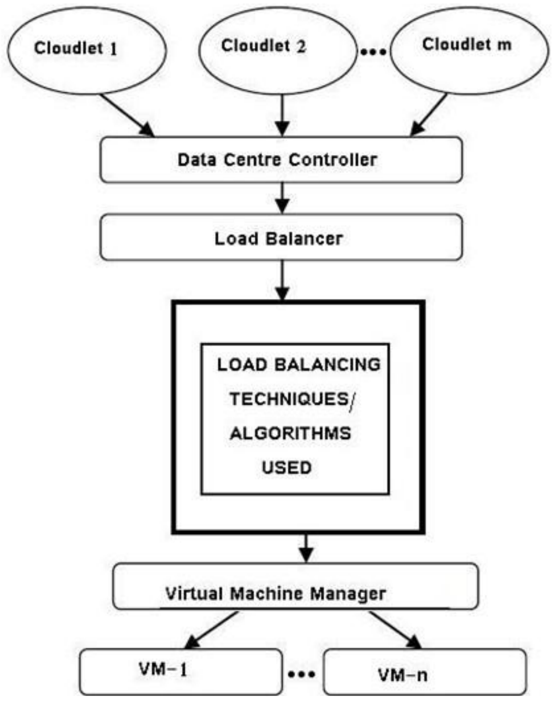

Fig.1 Load Balancing Algorithms Execution

2. ISSUE EFFECTING LOAD BALANCING IN CLOUD COMPUTING

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. Figure 1 presentsa framework under which various load balancing algorithms work in a cloud computing environment. Some major issues which must be considered while designing any load balancing algorithm are as Follows:

2.1 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 Face book, Twitter,etc.

A well-distributed system of nodes in cloud environment is helpful in handling fault tolerance and maintaining the efficiency of the system.

2.2 Dynamic Vs Static behaviour of Algorithms

Any algorithm concerning load balancing is designed, based on the state or behaviour of the system; this may be static or dynamic.

2.2.1 Static Algorithms

These algorithms do not depend upon the current state of the system and have prior knowledge regarding system resources and details of all tasks in an application. These kinds of algorithms face a major drawback in case of sudden failure of system resource and tasks.

2.2.2 Dynamic Algorithms

These algorithms take decisions concerning load balancing based upon the current state of the system and don’t need any prior knowledge about the system. This approach is an improvement over the static

approach. The algorithms in this category are considered complex, but have better fault tolerance and overall performance.

2.3 Algorithm complexity

The complexity of any load balancing algorithm affects the overall performance of the system. Sometimes the algorithm is complex, but is better in terms of throughput and resource utilization. On the other hand, the algorithms which are simpler in terms of complexity may give poor performance in terms of fault tolerance, migration time and response time. Therefore, based on the system requirements, care should be taken to decide a better or suitable load balancing algorithm. A trade-off between all the Parameters must be set wisely.

3. Exiting Load Balancing Techniques In Cloud Computing

In this section, some significant contributions on load balancing in cloud computing, as mentioned in the literature are discussed. A. Khiyaita et al. [1], in their paper gave an overview of load balancing in cloud computing, classification of load balancing algorithms based upon system load and system topology, examples of load balancing and different research challenges in load balancing. While in [2] authors discussed most of the existing techniques which are aimed at reducing the associated overhead, service response time and improving performance of the technique. The paper also provides details about various parameters, used to compare the existing techniques. D. A. Menasce et al. [3] discussed the concept of cloud computing, its advantages, and disadvantages and described several existing cloud computing platforms.

He also discussed the results of quantitative experiments carried out using Planet Lab, a cloud computing platform and capacity planning methods for cloud users and cloud service providers. To maintain the load balancing in the cloud computing system, Kuo-Qin Yan et al.[4] proposed a scheduling algorithm. Their algorithm combined the capabilities of both OLB (Opportunistic Load Balancing) [5] and LBMM (Load Balance Min-Min) [6] scheduling algorithms, and is comparatively more efficient.

In [7], idea is to find the best cloud resource while considering Co-operative Power aware Scheduled Load Balancing, solution to the Cloud load balancing challenge. The algorithm proposed in this paper, utilizes the benefits of both distributed and centralized approach of computing. Inherent efficiency of centralized approach as well as the energy efficiency and fault-tolerance nature of distributed approach is well used in this algorithm. In PALB [8] approach, the utilization percentages of each compute node is estimated. This helpsin deciding the number of compute nodes which must keep operating while other nodes completely shut down. The algorithm has three sections: Balancing section, upscale section and downscale section.

Balancing section is responsible for determining where virtual machines will be instantiated based on utilization percentages. The upscale section power-on's the additional compute nodes. And the downscale section shut-downs the idle compute nodes. This algorithm is guaranteed to decrease the overall power consumption while maintaining the availability of resources as compared to other load balancing algorithms. Raul Alonso-Calvo et al. [9] have discussed on managing large image collections in companies and institutions.

A cloud computing service and its application for storage and analysis of very large images have been created and the data operations are adapted for working in a distributed mode by using different sub-images that can be stored and processed separately by different agents in the system, facilitating processing very-large images in a parallel manner. This work can be viewed as another way of load balancing in cloud computing. Apart from availability of resources, other factors like scaling of resources and power consumption are also important concerns in load balancing that cannot be ignored. Load balancing techniques should be such as to obtain measurable improvements in resource utilization and availability of resources in the cloud computing environment [10]. Alexandru Iosup etal. [11] Analyzed the performance of cloud computing services for scientific computing workloads and quantified the presence in real scientific computing workloads of Many-Task Computing (MTC) users, that is, of users who employ loosely coupled applications comprises many tasks to achieve their scientific goals. They also perform an empirical evaluation of the performance of four commercial cloud computing services.

Srinivas Seth et al. [12] proposed a load balancing algorithm using fuzzy logic in a cloud computing environment. This algorithm uses two parameters processor speed and assigned a load of virtual machine, to balance the overall load through fuzzy logic, although in [13], the authors have introduced a new fuzzy logic based dynamic load balancing algorithm with additional parameters-memory usage, bandwidth usage, and disk space usage and virtual machine status and named it as Fuzzy Active Monitoring Load Balancer (FAMLB). Milan E. Sokile [14], have discussed different load balancing techniques in adistributed environment, namely diffusive load, static, round robin and shortest queue in different clientenvironm1ents. Experimental analysis have been done showing diffusive load balancing is more efficient than static and round robin load balancing in a dynamic environment. Ankush P. Deshmukh and Prof Kumarswamy Pamu [15], discussed on different load balancing strategies, algorithms and methods. The research also shows that the dynamic load balancing is more efficient than other static load balancing techniques. Efficient load balancing can clearly give major performance benefits [16]. A Network Processor consists of a number of on-chip processors to carry out packet level parallel processing operations, ensuring good load balancing among the processors. This process increases the throughput of the system. However, such type of multiprocessing increases out-of-order departure of processed packets.

first propose an Ordered Round Robin (ORR) scheme to schedule packets in a heterogeneous network processor, assuming that the workload is perfectly divisible. The processed loads from the processors are perfectly ordered. This paper analyzes the throughput and derives expressions for the batch size, scheduling time and maximum number of schedulable processors. Jaspreet Kaur [17] has discussed active vm load balancer algorithm to find a suitable virtual machine in less time period. She has done simulation showing comparative analysis of round robin and equal spread current execution policies of load balancing with varying service broker policies for data center in a cloud computing environment and compared their response time and cost. hang Bo et al. [18], proposed an algorithm which adds capacity to the dynamic balance mechanism for the cloud. The experiments demonstrate that the algorithm has obtained a better load balancing degree.

and used less time in loading all tasks. Soumya Ray and Ajanta De Sarkar [19] have discussed various algorithms of load balancing like Round robin algorithm, Central queuing algorithm and Randomized algorithm, their analysis is carried out on MIPS vs. VM and MIPS vs. HOST basis. Their results demonstrate that these algorithms can possibly improve the response time in order of magnitude, with respect to number of VMs in datacenter. Execution analysis of the simulation shows that the change of MIPS will affect the response time. Increasing MIPS vs. VM decreases the response time. In order to handle the random selection based load distribution problem, dynamic load balancing algorithm can be implemented as the future course of work to evaluate various parameters. In [20], the authors have proposed an algorithm on load distribution of workloads among nodes of a cloud, by the use of Ant Colony Optimization (ACO). This algorithm uses the concept of the ant colony optimization method. Shridhar G. Domanal and G. Ram Mohana Reddy [21] have proposed a local optimized load balancing approach for distributing incoming job request uniformly between the servers or virtual machines. They analyzed the performance of their algorithm using the Cloud Analyst simulator. Further, they also compared their approach with Round Robin [22] and Throttled algorithm [23]. A similar work was doneby Shridhar G. Domanal and G. Ram Mohana [21] using CloudSim and Virtual Cloud simulators. In [24], the authors have analyzed various policies in combination with different load balancing algorithms usinga tool called Cloud Analyst. They presented various variants of Round Robin load balancing algorithm, demonstrating the pros and cons of each. The Dynamic Round Robin algorithm is an improvement over static Round Robin algorithm [25], this paper analyzed the Dynamic Round Robin algorithm with varying parameters of host bandwidth, cloudlet length, VM image size and VM bandwidth. Results have been analyzed based upon the simulation carried in CloudSim simulator. In [26], the authors Ching-Chi Lin, Pangfeng Liu, Jan-Jan Wu, have proposed a new Dynamic Round Robin (DRR) algorithm for energy-aware virtual machine scheduling and consolidation. This algorithm is compared with other existing algorithms, namely

Greedy, Round Robin and Power save strategies used in Eucalyptus.

4. DISCUSSION AND COMPARISION

In the previous section different load balancing techniques proposed by various researchers have been discussed. Table1 gives a comparative analysis of different load balancing techniques with respect to different performance parameters.

In Table1 various metrics have been considered to compare different techniques. The metrics on which the existing load balancing techniques have been measured are discussed below:

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

Overhead 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.

Table1.Comparison of existing Load Balancing Techniques

Metrics/ Techniques	Throughput	Overhead	Fault tolerance	Migration time	Response time	Resource utilization	scalability	Performance
Round Robin [22]	YES	YES	NO	NO	YES	YES	YES	YES
Dynamic Round Robin	YES	YES	YES	YES	NO	YES	NO	NO

[26]								
PALB [8]	YES	YES	YES	YES	YES	YES	NO	NO
Active Monitoring [23]	YES	YES	NO	YES	YES	YES	YES	NO
FAMLB [13]	YES	YES	YES	YES	NO	YES	YES	YES
Min-Min [6]	YES	YES	NO	NO	YES	YES	NO	YES
Max-Min [27]	YES	YES	NO	NO	YES	YES	NO	YES
OLB+LBMM [4]	NO	NO	NO	NO	NO	YES	NO	YES
Throttled [23]	NO	NO	YES	YES	YES	YES	YES	YES
Honeybee Foraging [28]	NO	NO	NO	NO	NO	YES	NO	NO
Active Clustering [28]	NO	YES	NO	YES	NO	YES	NO	NO
Biased Random Sampling[28]	NO	YES	NO	NO	NO	YES	NO	YES

5. CONCLUSION AND FUTUREWORK

In this paper, we have surveyed various load balancing algorithms in the Cloud Computing environment. We discussed 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] A. Khiyaita, M. Zbakh, H. El Bakkali, and D. El Kettani, “Load balancing cloud computing: state of art,” in Network Security and Systems (JNS2), 2012 National Days of, pp. 106–109, IEEE, 2012.[2] N. J. Kansal and I. Chana, “Existing load balancing techniques in cloud computing: A systematic review.,” Journal of Information Systems & Communication, vol. 3, no. 1, 2012.[3] D. A. Menasc´e and P. Ngo, “Understanding cloud computing: Experimentation and capacity planning,” in Computer Measurement Group Conference, 2009.[4] 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.[5] T. D. Braun, H. J. Siegel, N. Beck, L. L. B¨ol¨oni, M. Maheswaran, A. I. Reuther, J. P. Robertson,M. D. Theys, B. Yao, D. Hensgen, et al., “A comparison of eleven static heuristics for mapping aclass of independent tasks onto heterogeneous distributed computing systems,” Journal of Parallel and Distributed computing, vol. 61, no. 6, pp. 810–837, 2001.[6] 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.[7] T. Anandharajan and M. Bhagyaveni, “Co-operative scheduled energy aware load-balancing technique for an efficient computational cloud.,” International Journal of Computer Science Issues
(IJCSI), vol. 8, no. 2, 2011.[8] 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.[9] R. Alonso-Calvo, J. Crespo, M. Garcia-Remesal, A. Anguita, and V. Maojo, “On distributing load in cloud computing: A real application for very-large image datasets,” Procedia Computer Science,vol. 1, no. 1, pp. 2669–2677, 2010.[10] Z. Chaczko, V. Mahadevan, S. Aslanzadeh, and C. Mcdermid, “Availability and load balancing in cloud computing,” in International Conference on Computer and Software Modeling, Singapore, vol. 14, 2011.[11] A. Iosup, S. Ostermann, M. N. Yigitbasi, R. Prodan,T. Fahringer, and D. H. Epema, “Performance analysis of cloud computing services for many-tasks scientific computing,” Parallel and Distributed Systems, IEEE Transactions on, vol. 22, no. 6, pp. 931–945, 2011.[12] S. Sethi, A. Sahu, and S. K. Jena, “Efficient load balancing in cloud computing using fuzzy logic,”IOSR Journal of Engineering, vol. 2, no. 7, pp. 65–71, 2012.[13] 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.[14] M. E. Soklic, “Simulation of load balancing algorithms: a comparative study,” ACM SIGCSE Bulletin,vol. 34, no. 4, pp. 138–141, 2002.[15] A. P. Deshmukh and K. Pamu, “Applying load balancing: A dynamic approach,” International Journal, vol. 2, no. 6, 2012.[16] J. Yao, J. Guo, and L. N. Bhuyan, “Ordered round-robin: An efficient sequence preserving packet scheduler,” Computers, IEEE Transactions on, vol. 57, no. 12, pp. 1690–1703, 2008.[17] J. Kaur, “Comparison of load balancing algorithms in a cloud,” International Journal of Engineering Research and Applications, vol. 2, no. 3, pp. 1169–173, 2012.[18] Z. Bo, G. Ji, and A. Jieqing, “Cloud loading balance algorithm,” in Information Science and Engineering (ICISE), 2010 2nd International Conference on, pp. 5001–5004, IEEE, 2010.
[19] S. Ray and A. De Sarkar, “Execution analysis of load balancing algorithms in cloud computing environment.,” International Journal on Cloud Computing: Services & Architecture, vol. 2, no. 5,2012.[20] K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Nitin, and R. Rastogi, “Load balancing of nodes in cloud using ant colony optimization,” in Computer Modelling and Simulation (UKSim),2012 UKSim 14th International Conference on, pp. 3–8, IEEE, 2012.[21] S. G. Domanal and G. R. M. Reddy, “Load balancing in cloud computing using modified throttled algorithm,” 2013.[22] S. Subramanian, G. Nitish Krishna, M. Kiran Kumar, P. Sreesh, and G. Karpagam, “An adaptivealgorithm for dynamic priority based virtual machine scheduling in cloud.,”

International Journal of Computer Science Issues(IJCSI), vol. 9, no. 6, 2012.[23] B. Wickremasinghe, “Cloud analyst: 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.[24] S. Mohapatra, S. Mohanty, and K. S. Rekha, “Analysis of different variants in round robin algorithms for load balancing in cloud computing,” International Journal of Computer Applications, vol. 69,no. 22, 2013.[25] A. Gulati and R. K. Chopra, “Dynamic round robin for load balancing in a cloud computing,” 2013.[26] C.-C. Lin, P. Liu, and sJ.-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.[27] 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.[28] M. Randles, D. Lamb, and A. Taleb-Bendiab, “A comparative study into distributed load balancing algorithms for cloud computing,” in Advanced Information Networking and Applications Workshops(WAINA), 2010 IEEE 24th International Conference on, pp. 551–556, IEEE, 2010.