

A Review on Load Balancing Approach in Cloud Computing

Garima Joshi
CSE Dept. GBPEC,
Ghurdauri, U.K, India

S. K. Verma
CSE Dept. GBPEC,
Ghurdauri, U.K, India

ABSTRACT

In the simplest way, cloud computing means classifying and getting data and programs over the Internet instead of your computer's hard disk drive. It is a dissemination of computing resources—everything from applications to data centers over the Internet on a reward-for-use basis. Cloud computing is mostly defined as a type of computing that focuses on *sharing computing resources* rather of having the local host or personal devices to manage applications. The major issues related to the cloud are resource discovery, fault tolerance, load balancing, and security. Load balancing is the primary consideration in the cloud-computing environment. Load reconciliation is a mean for propagating workloads across various computing resources. Its main motive is to optimize the usage of resources, boost turnout, deprecation reaction time, and avoiding the needless burden of any such resources. It becomes a severe problem with the ceaseless increase in the list of users and type of applications on clouds. This paper presents a quick review of cloud computing. The main highlight of this document is on the load balancing approach in cloud computing.

Keywords:

Cloud Computing, Load Balancing, Dynamic Load Balancing, Performance analysis, Data Center

1. INTRODUCTION

Cloud computing gets its name as an analogy for the Internet. Typically, the Internet is symbolized in network diagrams as a cloud [1]. The term “cloud” [2] systematized from the world of telecommunications when supplier began using virtual private network (VPN) services for data communications. Cloud computing simply means Internet Computing, broadly the Internet is seen as an accumulation of clouds. Thus, cloud computing is outlined as employing the internet to provide technology-enabled services to the people and organizations [3]. According to NIST(National Institute of Standards and Technology), cloud computing is a prototype for enabling convenient, on-demand network gateway to a shared group of configurable computing resources [2].

The concept of Evolution of Cloud Computing came into existence in 1950 with the implementation of mainframe computers, accessible to thin/ static clients. Since then, Cloud computing has been originated from static client to dynamic ones from software to services. The following figure 1 explains the evolution of cloud computing.

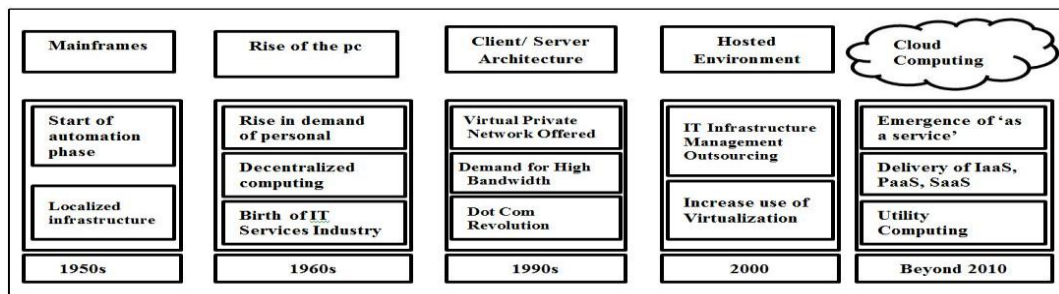


Figure 1 Evolution of Cloud Computing

Cloud computing is a recent drifting in IT that moves computing and data away from desktop and manageable PCs into large data centers. The primary gain of cloud computing is that customer do not have to pay for infrastructures, its installation and prescribed workforce to handle such support and maintenance [2]. Cloud computing is cheaper than other computing models; zero repair cost is involved since the service provider is

accountable for the accessibility of services and clients are devoid from maintenance and management issues of resource machines. Due to this attribute, cloud computing is also admitted as utility computing [3]. Cloud computing

is not application-program oriented instead service-oriented; it holds out on-demand virtualized resources as discoverable [4]. The possibility of cloud computing has significantly modified the field of parallel and distributed computing systems today [5].

The concern of cloud computing is still surfaced by many issues that will be discussed in this paper. Section 2 will foreground the fundamental Issues in the cloud platform. Section 3 will furnish a brief overview of load balancing. Section 4 gives a review of the past research work on load balancing on the cloud platform, and at last in Section 5 conclude the paper, followed by the references.

2. ISSUE IN CLOUD PLATFORM

Distinct from the issues enumerated in the previous section, the other set of the technical load balancing; task scheduling, resource sensing, fault tolerance, protection, and lots more. Here is a brief introduction to most of the current issues regarding cloud computing:

Load balancing [6]: In computing, load balancing disseminates workloads throughout different computing devices, such as computers, network topology, a computer cluster, central processing units or magnetic disk. Load equilibrating intention is to optimize resource use, maximize turnout, minimize response time and avoid the burden of any single resource. Using numerous components with load balancing instead of a single component may amplify reliability through redundancy. Load balancing usually comprises dedicated software program or computer hardware, such as a multilayer flip-flop or a Domain Name System server process. Load balancing guarantee that all the processors in the system or every node in the network does approximately the adequate amount of work at any instant of time. The load balancing needs to be done correctly because failure in any one node can lead to unavailability of data.

Security [7]: It is one of the primary concerns with cloud computing. These issues fall into two broad division: security issues faced by cloud providers and security issues faced by their customers. The supplier must reassure that their base is secure and that their clients' information and applications program are protected. While the user must take view to payback, their application and use guarantee passwords and certification standards.

Reliability: Some people also worry about whether a cloud service provider is commercially stable and whether their data storage system is dependable. Most cloud providers seek to mitigate this fear by using excessive storage techniques. However, it is still potential that a

service could collapse, allowing users with bounded or no memory access.

Data Backup: Cloud backup, refers to backing up data to a faraway, cloud-based server. As a form of cloud storage, cloud backup data is stored in and approachable from multiple distributed and connected resources that constitute a cloud. Instead of storing data locally cloud backup enable individuals to store their information on the Internet, rather than storing the data locally on a physical records, such as a disc drive or tape support.

Data Portability and Conversion [8]: Porting and converting data is extremely dependent on the cloud provider's data format, which may survive, in some proprietary or deficient format. In all cases, when dealing with a cloud supplier, an organization should recognize how to displace and retrieve your data.

3. LOAD BALANCING IN CLOUD PLATFORM

It [9] is a process of reassigning the total capacity to the individual nodes of the corporate system to make resource utilization effectual. At the same time, getting rid of circumstances where some of the nodes are overloaded while some others are under loaded. Load balancers distribute the weight to different virtual machines in such a fashion that none of the nodes gets charged principally or gently. The load balancing demands to arrange properly because failure in any one node can lead to unavailability of data [10]. The load balancer accepts numerous requests from the client and distributing each of them across various computers or network devices irrespective of how busy the network device is. Load balancing assists to prevent a server or network device from getting vanquish with requests and helps in the dispersion of work. If the load balancer is unavailable, the client can wait for massive spam of time until the server is idle to accept their request [11]. The following figure explains load balancing in cloud computing.

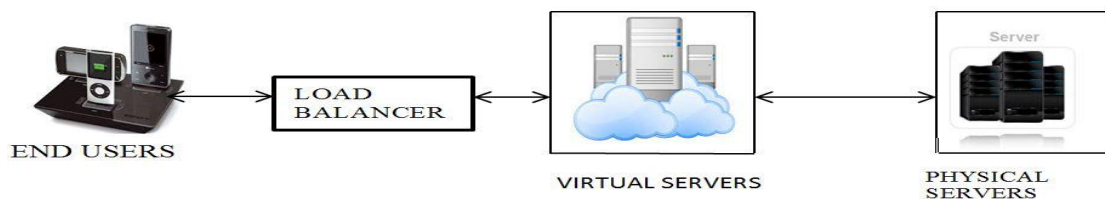


Figure 2. Load Balancer

Load balancers work in two ways: cooperative and non-cooperative. In cooperative, the nodes work at the same time in order to attain the common goal of optimizing the overall reaction time. In non-cooperative manner, the tasks run autonomously in order to improve the response time of local job. Load balancing algorithms is fragmented into two categories: static and dynamic load equilibrating algorithm. A static load-balancing algorithm does not take into consideration the previous state or behavior of a node while distributing the load. On the other hand, a dynamic load-balancing algorithm assures the previous state of a node while allotting the load. In dynamic load, balancing if any node breaks down, it will not block the system it will only degrade the system performance. In a dynamic load balanced system, the nodes can interact with each other rendering more messages when compared to a non-

distributed environment. However, picking out an appropriate server involves real-time communication with the other nodes of the network and, bring more number of messages in the web. The Dynamic load balancer uses policies for continuing track of modified information.

4. EXISTING LOAD BALANCING TECHNIQUES IN CLOUD COMPUTING

In this segment, some substantial contributions to load balancing in cloud computing, as mentioned in the literature is considered. A. Khiyaita et al. [12], gave an outline of load balancing in cloud computing, categorization based on system weight and system form. Although N.J. Kansas and I. Chana in [13] consider most

of the running process, which are directed to bringing down the associated overhead, service response time and bettering performance of the technique. The paper also furnishes details about several parameters, used to equate the existing methods. Further, D.A. Menasce et al. [14] discussed various existing cloud-computing podium, its pros and cons. He carried out experiments using Planet Lab, a cloud computing platform and capacity planning methods for cloud users and cloud service supplier. To maintain the load equilibrating in the cloud computing system, Kuo-Qin Yan et al. [15] proposed a scheduling algorithm. It combine the capabilities of both OLB (Opportunistic Load Balancing) [16] and LBMM (Load Balance Min-Min) [17] scheduling algorithms and is relatively more capable. Further, T. Anand Rajan in [18] give an estimate to find the most beneficial cloud resource while regarding Co-operative Power- aware Scheduled Load Balancing, a solution to the Cloud load balancing challenge.

J. M. Galloway et al. [19] gives an idea of PALB approach, where usage percentages of each compute node are figured. This algorithm aimed to decrease the overall power expenditure while maintaining the availability of resources as equated to other load balancing algorithms. In addition, Raul Alonso-Calvo et al. [20] gives an idea of cloud service for storage tremendous images and the data operations are adapted for bringing in a distributed mode. This work can be considered as another way of load balancing in cloud computing. Aside from the accessibility of resources, other factors like grading of resources and power consumption are also crucial concerns in load balancing [21]. Alexandru Iosup et al. [22] analyzed the operation of cloud computing services for scientific computing workloads. He evaluated the presence in real technological computing workloads of Many-Task Computing users, who utilize loosely linked applications comprise many tasks to achieve their scientific goals. Srinivas Sethi et al. [23] use idea of fuzzy logic to propose a load-balancing algorithm in a cloud-computing environment. In this scheme, the load is distributed across the different virtual machine, to counter balance the overall load through fuzzy logic. Although, A. Kalam in [24], presented a new fuzzy logic based dynamic load balancing algorithm. It considered usage of memory, bandwidth, disk space and virtual machine status and named it as Fuzzy Active Monitoring Load Balancer (FAMLB). Milan E. Sokile [25], have discoursed different load balancing techniques in a distributed surroundings, namely diffusive capacity, static, round robin and shortest queue in different client environments. Analysis have been done showing static and round robin approach are less efficient than diffusive load balancing dynamic environment. Ankush Deshmukh and Prof. Kumaraswamy Pamu [26], presents a literature review of various existing load-balancing algorithms. The analysis shows that dynamic load distribution is more effective than other static load balancing techniques. J. Yao in [27] gives an idea of a Network Processor that comprises of an amount of on-chip processors to carry through packet level multi-processing function, assuring proper load balancing between the processors. This approach maximizes the throughput of the system. In this paper, the authors first suggest an Ordered Round Robin (ORR) strategy to schedule packets in a heterogeneous network processor, acquiring that the workload is divisible. The processed loads from the processors are ordered. This paper examines the

throughput and reduces expressions for the batch size, scheduling a time and maximum number of schedulable processors. The research has further extended, and Jaspreet Kaur [28] has discussed an active VM load-balancer algorithm to detect a desirable virtual machine in less time. A simulation describing the relative depth of round robin and equal spread current execution policies of load balancing with varying service broker contracts for the data center in a cloud environment and matched their reaction time and cost. Zhang Bo et al. [29], suggested an algorithm, which adds subject to the dynamic equilibrium mechanism for the cloud. The experiments demonstrate that the algorithm has obtained a better load balancing degree, uses less time in loading all tasks. They have also review various algorithms of load balancing like Round robin algorithm, Central queuing algorithm, and Randomized algorithm, their analysis is accomplished on MIPS vs. VM and MIPS vs. HOST basis. Analysis of the simulation shows that the change of MIPS will affect the response time. K. Nishant et al. in [30], have suggested an algorithm for distribution of workloads within nodes in a cloud, by the use of Ant Colony Optimization (ACO). Shridhar G. Domanal and G. Ram Mohana Reddy [31] try to distribute incoming job request uniformly between servers or virtual machine in a cloud environment to obtain load balancing. The performance of their algorithm is analyzed using the Cloud Analyst simulation setup. They also compare their approach with Round Robin [32] and Throttled algorithm [33]. S. Mohapatra in [34], have analyzed various policies in combination with different load balancing algorithms using a tool called Cloud Analyst. They presented different variants of Round Robin load balancing algorithm, demonstrating the pros and cons of each. This paper examined the Dynamic Round Robin algorithm with altering arguments of host bandwidth, cloudlet length, VM image size, and VM bandwidth. In [35], Himanshu goel et al. gives an overview of various loads balancing for scheduling job in cloud environment.

5. CONCLUSION

Cloud computing is a rising field of information technology (IT). It modifies a wide range of users to access distributed, scalable, virtualized, hardware or software base over the Internet. Load balancing is one of the leading issues of cloud computing. Therefore, there is a demand for a regular load-balancing algorithm for efficient utilization of resources. We gave an overview of already proposed algorithms by several researchers in the literature.

This paper is an earnest effort to unveil the concept of load balancing and its type's, especially dynamic load balancing. Nevertheless, even there are miles to go. As cloud computing is yet in its early childhood and there are many more exposed issues like security measure, resource usage, etc. that need to be researched. The future work is to provide an organized or efficient load balancing and fault tolerant technique in order to remove these issues permanently and to design the approach more cost-effective and successful.

6. REFERENCES

- [1] Velte, A.T., Veltey, T.J., and Elsenpeter, R., 2010, "Cloud Computing: A Practical Approach," Tata McGraw-Hill Education Private Limited, New Delhi, Edition.
- [2] Jadeja, Y., and Modi, K., 2012, "Cloud Computing- Concepts, Architecture and Challenges," International Conference on Computing, Electronics and Electrical Technologies, pages 877-880.
- [3] Shaikh, F.B., and Haider, S., 2011, "Security Threats in Cloud Computing," Internet Technology and Secured Transactions, pages 214-219.
- [4] Srinivas, J., Reddy, K.V.S., and Qyser, A.M., July 2012, "Cloud Computing Basics," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 5.
- [5] Ray, S., and Sarkar, A.D., October 2012, "Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment," International Journal of Cloud Computing Services and Architecture, Vol. 2, Issue 5.
- [6] Kaur, J., May-Jun 2012, "Comparison of load balancing algorithms in a Cloud," International Journal of Engineering Research and Applications, Vol. 2, Issue 3, pages 1169-1173.
- [7] Kuyoro, S.O., Ibikunle, F., Awodele, O., 2011, "Cloud Computing Security Issues and Challenges," International Journal of Computer Networks (IJCN), Vol. 3, Issue 5.
- [8] Carpenter, R.H., 2010, "Walking from Cloud to cloud: the portability issue on cloud computing," Washington Journal of Law, Technology & Arts, Vol. 6, Issue 1.
- [9] Ramana, K., Subramanyam, A., and Rao, A.A., "Comparative Analysis of Distributed Web Server System Load-Balancing Algorithms Using Qualitative Parameters," 2011, VSRD-IJCSIT, Vol. 1, Issue 8, pages 592-600.
- [10] Moharana, S.S., Ramesh R.D., and Powar, D., May 2013, "Analysis of Load Balancers in Cloud Computing," International Journal of Computer Science and Engineering, Vol.2, Issue2.
- [11] Roy, A., and Dutta, D., April 2013, "Dynamic Load Balancing: Improve Efficiency in Cloud Computing," International Journal of Emerging Research in Management & Technology, Vol. 2, Issue 4.
- [12] Khiyaita, A., Zbakh, M., Bakkali, H. El., and Kettani, D.El, 2012, "Load balancing cloud computing: state of art," In National Days of Network Security and Systems (JNS2), IEEE, pages 106-109.
- [13] Kansal, N. J., and Chana, I., 2012, "Existing load balancing techniques in cloud computing: A systematic review," Journal of Information Systems & Communication, Vol. 3, Issue 1,
- [14] Menasce, D.A., and Ngo, P., "Understanding cloud computing: Experimentation and capacity planning," 2009, In Computer Measurement Group Conference.
- [15] Wang, S.C., Yan, K.Q., Liao, W.P., and Wang, S.S., 2010, "Towards a load balancing in a three-level cloud computing network," In Computer Science and Information Technology (ICCSIT) 3rd IEEE International Conference, IEEE, Vol. 1, pages 108-113.
- [16] Braun, T. D., Siegel, H. J., Beck, N., Boloni, L. L., Maheswaran, M., Reuther, A. I., J. P. Robertson, M. D. Theys, B. Yao, D. Hensgen, et al., 2001, "A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems," Journal of Parallel and Distributed Computing, Vol. 61, Issue 6, pages 810-837.
- [17] Kokilavani, T., and Amalarethinam, D., 2011, "Load balanced min-min algorithm for static meta-task scheduling in grid computing," International Journal of Computer Applications, Vol. 20, Issue 2.
- [18] Rajan, T.A., and Bhagyaveni, M., 2011, "Co-operative scheduled energy aware load-balancing technique for an efficient computational cloud," International Journal of Computer Science Issues (IJCSI), Vol. 8, Issue 2.
- [19] Galloway, J.M., Smith, K.L., and Vrbsky, S.S., 2011, "Power aware load balancing for cloud computing," In Proceedings of the World Congress on Engineering and Computer Science, Vol. 1, pages 19-21.
- [20] Alonso-Calvo, R., Crespo, J., Garcia-Remesal, M., Anguita, A., and Maojo, V., 2010, "On distributing load in cloud computing: A real application for very-large image datasets," Proceeded in Computer Science, Vol. 1, Issue 1, pages 2669-2677.
- [21] Chaczko, Z., Mahadevan, V., Aslanzadeh, S., and Mcdermid, C., 2011, "Availability and load balancing in cloud computing," International Conference on Computer and Software Modeling, Singapore, Vol. 14.
- [22] Iosup, A., Ostermann, S., Yigitbasi, M. N., Prodan, R., Fahringer, T., and Epema, D. H., 2011, "Performance analysis of cloud computing services for many tasks scientific computing," Parallel and Distributed Systems, IEEE Transactions, Vol. 22, Issue 6, pages 931-945.
- [23] Sethi, S., Sahu, A., and Jena, S. K., 2012, "Efficient load balancing in cloud computing using fuzzy logic," IOSR Journal of Engineering, Vol. 2, Issue 7, pages 65-71.
- [24] Nine, Z., SQ, M., Azad, M., Kalam, A., Abdullah, S., and Rahman, R. M., 2013, "Fuzzy logic based dynamic load balancing in virtualized data centers," IEEE International Conference, pages 1-7.
- [25] Soklic, M.E., 2002, "Simulation of load balancing algorithms: a comparative study," ACM SIGCSE Bulletin, Vol. 34, Issue 4, pages 138-141.
- [26] Deshmukh, A.P., and Pamu, K., 2012, "Applying load balancing: A dynamic approach," International Journal, Vol. 2, Issue 6.
- [27]

- [28] Yao, J., Guo, J., and Bhuyan, L. N., 2008, "Ordered round-robin: An efficient sequence preserving packet scheduler," *IEEE Transactions*, Vol. 57, Issue 12, page 1690–1703.
- [29] Kaur, J., 2012, "Comparison of load balancing algorithms in a cloud," *International Journal of Engineering Research and Applications*, Vol. 2, Issue 3, pages 1169–173.
- [30] Bo, Z., Ji, G., and Jieqing, A., "Cloud loading balance algorithm," 2010, In *Information Science and Engineering (ICISE)*, 2nd International Conference on IEEE, page 5001–5004.
- [31] Ray, S., and Sarkar, A. De., 2012, "Execution analysis of load balancing algorithms in cloud computing environment," *International Journal on Cloud Computing: Services & Architecture*, Vol. 2, No. 5.
- [32] Nishant, K., Sharma, P., Krishna, V., Gupta, C., Singh, K.P., Nitin, N., and Rastogi, R., 2012, "Load balancing of nodes in cloud using ant colony optimization," in *Computer Modelling and Simulation (UKSim) 14th International Conference on IEEE*, Vol. 4, Issue 4, pp. 3–8.
- [33] Subramanian, S., Krishna, G.N., Kumar, M.K., Suresh, P., and Karpagam, G., 2012, "An adaptive algorithm for dynamic priority based virtual machine scheduling in cloud," *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, No. 6.
- [34] Wickremasinghe, B., 2009, "Cloudanalyst: A cloudsim-based tool for modeling and analysis of large scale cloud computing environments," *MEDC Project Report*, Vol. 22, Issue 6, pages 433–659.
- [35] Mohapatra, S., Mohanty, S., and Rekha, K.S., 2013, "Analysis of different variants in round robin algorithms for load balancing in cloud computing," *International Journal of Computer Applications*, Vol. 69, Issue 22.
- [36] Goel, H., Chamoli, N., June 2014, "Job Scheduling Algorithms for Cloud Computing: A Survey," *International Journal of Computer Applications*, Volume 95, Issue 23.