Volume 5 Issue 2 December 2017
ISSN: 2321-6840

Load Balancing in Cloud Computing

Ankita Gupta
Department of Computer Science and Engineering, Govt. Engineering College, Bikaner, Rajasthan, India.
Email: 4ankitagupta@gmail.com

Abstract: Cloud computing becomes an important technology for distributed computing and parallel computing. Cloud computing provides various facilities; like to share resources, software packages, information, storage and many other applications depending on user demand at desired time place. It provides an extensive measure for computing and storage. A service provided by it to user follows pay-as-you-go model. Although it provides many facilities but still some problem attached to it are resource discovery, fault tolerance, load balancing, and security. Out of these, Load balancing is the main challenges. There are many techniques which used to distribute workload or task, equally across the servers. This paper includes cloud computing, cloud computing architecture, virtualization and MS load balancing technique which provide enhanced load balancing.

Keywords: Architecture, Cloud computing, Virtualization, Load balancing

I. Introduction

Cloud computing has become one of the important technology in IT world. Cloud computing provides online computing to satisfy demands of the users [15]. It encounters advancement in both the academe and industry [17]. Its main facility which attracts developers is that in this case of computing, computing Depends on sharing resources rather than having own servers or personal devices. With the help of cloud computing resource of software and hardware are shared reasonably to avoid shortcomings occurred in the early distributed network [2]. The cloud computing deployment models are essentially split into four groups, i.e., public, private, Hybrid, Community [4]. Cloud computing services are separated into three cases, i.e., platform as a service (Paas), software as a service (Saas), and infrastructure as a service (Iaas). On that point are several issues in cloud computing paradigm, as it is developing technology, simply load balancing is one of the major issues in a cloud computing environment. Load balancing is a methodology which offers methods to maximise throughput, optimized utilization of resources and better execution of system [3]. It likewise offers an gentle and flexible procedure to hold data or files and make them available for users [4]. To realize the efficient utilization of resources in the cloud system, there are various load balancing algorithms. The primary design of load balancing is to broadcast the local workload to the entire cloud. Load balancing can either be centralized or decentralized.

II. Cloud Computing

Cloud computing is an advanced technique which gives various computing resources and also provides storage. All users from all round the globe can access any of these resources on demand basis through the internet [1]. Essentially, cloud computing involves allocating tasks to several nodes efficiently in cloud system so that the request processing is performed in a well-organized manner [5]. Cloud computing allows a great figure of users around the globe to access virtualized sources and platform, scalable, distributed hardware and software resources and services through the net.

III. Cloud Computing Architecture

Cloud computing is fastest developing technology. Amazon, Google, Microsoft all is providing cloud computing services and working towards implementing and developing powerful, reliable and logical platforms for their users [7]. Cloud computing architecture as shown below consists of three service models, five essential characteristics and four cloud computing deployment models. Fig. 1 illustrates the architecture of cloud computing. A lower layer represents four deployment models, then the middle layer represents three basic service models and finally, the upper layer represents the underlying characteristics [1].

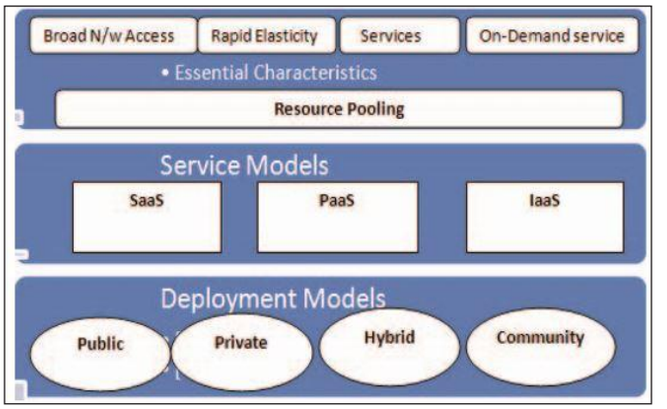

Fig. 1: Cloud Computing Architecture

A. Type of Cloud

Cloud computing is a avail which is based on the internet that offers services to the users on demand at whatever time. There are four types of clouds that provide computing services. These are:

- Public Cloud – This type of cloud infrastructure is openly used by the general public but available in a "pay per usage" manner [9]. Examples are Amazon or Google clouds which are usable for all users [10].

Fig. 2: Public Cloud

- Private cloud (it is a type of cloud infrastructure that operated solely for a single system, whether done internally or by a third-party. It can be hosted either internally or externally [10]. The private cloud project requires significant engagement to the virtualization of the business environment, and also calls for the organization to reevaluate existing resources. It improves business, but every step in the project requires high security that must be spoken to prevent serious threats [11].

Fig. 3: Private Cloud

- Hybrid Cloud – It is the type of cloud infrastructure which combines the public cloud and the private cloud. This type of cloud is basically used for commercial usage [10].

Fig. 4: Hybrid Cloud

B. Cloud Computing Service Models

Cloud computing provides various types of services to its users:

- Infrastructure as a Service (IaaS) – It consists services that allow its consumers to request storage and computational resources as per requirement on demand. It also enables "pay-per-use" paradigm. The most common examples of IaaS are Amazon EC2 [9].

Fig. 5: Infrastructure as a Service

- Platform as a Service (PaaS) – It contains high levels of services. It offers a platform for acquiring and bringing off the software base. In PaaS, the developer can either build or deploy various types of applications using libraries, voice communications, and tools offered by the cloud service providers. Google App Engine is one of the common examples of PaaS [9].

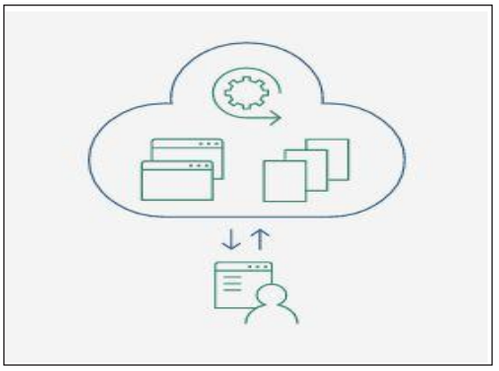

Fig. 6: Platform as a Service

- Software as a Service (SaaS) – It consists of end users, applications which are delivered to consumers as network services. Thus, this rules out the demand for installation and it runs different applications on consumers' computers. A usual case of SaaS is Google mail [9].

Fig. 7: Software as a Service

IV. Cloud Virtualization

In cloud computing, virtualization is very important. Virtualization, as the epithet connotes, is not a real but virtual system which offers all the facilities that actually exist. Virtualization includes software implementation on different computers on which a bit of distinct programs can be executed as a real machine [17]. For exemplar, in Amazon EC2 in which IT infrastructure is deployed in the cloud and providers' data centres lying in the concept of virtual machines [15]. Virtual infrastructure management methods and information centre management tools have been around since before the evolution of cloud computing, and nowadays it became the industry's latest emerging IT system [18]. Universal range of users can access multiple services of cloud computing simultaneously and at whatever time. Thus the entire services are available to the many users by remote data centres, which are founded on the concept of virtualization [15]. Virtualization is divided into two types as follows:

- Full Virtualization – In a full virtualization, software and resources that are usable in the real server are also useable in the virtual system and hence for the broad installation, one system has to be fully installed on some other organization. Henceforth in this virtualization, computer system shares among many users and the hardware located on multiple systems [14].
- Para Virtualization – In this case of virtualization, all the system resources such as computer storage and the central processing unit that allows multiple operating systems to run along a single operating system [13]. As the epithet entails, only partial services are afforded by this type of virtualization, but complete services are not fully accessible by the unrivalled organisation. Hence migration, disaster recovery, and capacity management are central features of para virtualization [15].

V. Load Balancing

Load Balancing is a mechanism in which the workload is distributed on the resources of a node to respective resources on the other node of the network in such a way that it doesn't get rid of any of the currently running task [10]. So balancing of the cargo between several nodes of the cloud system became an important chore in a cloud computing environment. The load can be of any type basically it may be network load, memory load, CPU load and delay shipment. Hence it is important to efficiently share work load across the multiple nodes of a network for improved operation and resource utilization [16]. Major goals of load balancing are:

- Establishing fault tolerance system.
- Maintaining system stability.
- It will improve performance and overall efficiency.
- Minimizing the job execution time and waiting time in the waiting line.
- To increase user satisfaction.
- To improve the resource utilization.

To poise the total load, there are two cases of load balancing algorithm. These two components will be introduced as follows:

- Static Load Balancing – This algorithm requires prior knowledge of system resources. Thus, the determination of load distribution doesn't depend on the present state of the system [5]. In this type of environmental performance of processors is defined as the starting of the execution and it doesn't alter the executing process at run time while one making changes in the system workload [5]. This doesn't provide flexibility and therefore this case of algorithm is merely suitable for homogenous system environment.
- Dynamic Load Balancing – This algorithm doesn't require any prior knowledge of the system resources because the load distribution decision is directly founded

on the present state of the system [5]. As this arrangement provides flexibility hence this is suited for heterogeneous system environment. It causes changes in load at run time. This algorithm provides better performance than static algorithm [6].

VI. Challenges & Issues of Load Balancing

Load balancing challenges are:

- Throughput: Its value should be eminent for safe performance and it is assessed by calculating the execution time of processes by the CPU.
- Overhead: Its value should be less for safe performance and it measured by the involved overhead at the time of execution.
- Fault tolerance: Load balancing should have less number of faults for getting better performance.
- Migration Time: Time Taken By processor to transfer one process from one system to another, its value should be less.
- Response Time: Its value should be less. It is set as the time is consumed by the organization to the reaction of the process.
- Resource Utilization: It is the ability to efficiently utilize the resources of the machine. It should be in an optimized fashion.
- Scalability: It is the ability to do load balancing on a virtual machine with multiple clients.
- Performance: It can be utilised to evaluate the carrying out of the processor and it should be high.

VII. MS Load Balancing

Load balancing is the major issue in cloud computing. The objective of load balancing is to satisfy users' need by distributing the load on multiple nodes in a defined system and maximize resource utilization and also help in improving the overall system performance. So efficient load balancing is very important for system operation, resource utilization, system stability, maximizing throughput and minimizing response time that are the principal targets of the cloud system. In this paper for load distribution, a load balancer needed, which received tasks from a different location and then distributed to the data centre. If load balancing used in the right manner, then it achieves optimal resource usage which will minimize the resource consumption to balance the load among multiple nodes in the organisation, there are several load balancing algorithms could be inserted. This report represents an algorithm in load balancing is done by managing three factors. In this, a watchdog is set for collecting the full request from all the users' virtual machine for resources. When the watchdog is set to zero, all request start to swear out. Now for processing, it uses the shortest job first and then processes through the mechanism of the round robin algorithm. Essentially, when watchdog set to zero data centre considers the full request and sets them in society for the chore which has small time and granting to the request have less number of the resource request. As the order sets, then round robin is applied. So that no petition will have to undergo the starvation process. And the operation time of the whole arrangement is shortened.

This system employs dynamic and stable approach to treat the entire petition and for load balancing. It is an effective means to cut down the throughput time and increase the efficiency of the data centre to balance the burden. It is also effective in managing energy. As in this data centre are distributive in nature, then it helps in store data management. It Minimizes total Makespan, Better processing time, better processing cost, improved execution and resource usage. It too possesses the advantage of Virtual machine migration when Datacenter is overloaded. Load balancing is an important critical issue which involves the usage of resources and performance of the system run on the swarm. In this paper, MSLB techniques are studied and discussed for load balancing.

VIII. Result and Conclusion

Cloud computing is a system in which different resources are accessed by worldwide users over the internet according to their demand. These resources are produced quickly. Only in that respect there are several issues and challenges in cloud computing. Load balancing is main issue in cloud computing. The primary job of load balancing is to satisfy users' by fulfilling their demands and request while maintaining load among multiple nodes in the organisation and also maximise the resource usage and improves system performance. In this department, the result of MSLB is discussed and demonstrated. There are two data centres and 4 VM is taken for the research and for presenting result paradigm.

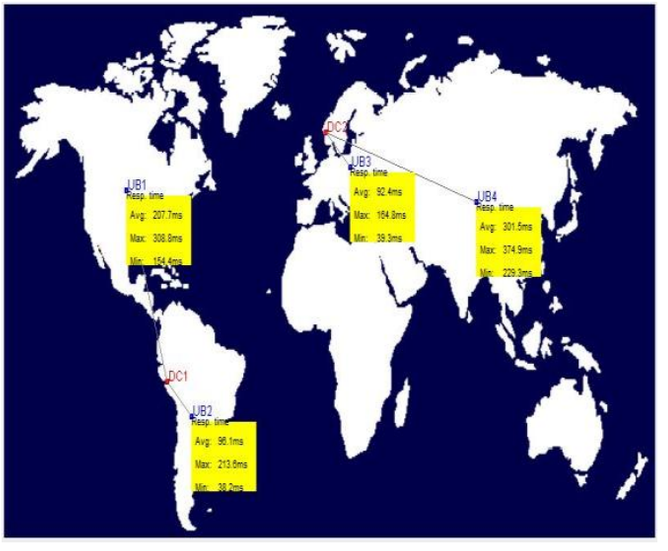

Fig. 8: Representation of Data Centers and VMS

Fig (9) interpret the overall response time of the MLB and the response time by various region or by various virtual machines.

Overall Response Time Summary

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	174.40	38.18	374.88
Data Center processing time:	24.67	0.02	167.80

Response Time by Region

Userbase	Avg (ms)	Min (ms)	Max (ms)
UB1	207.72	154.38	308.84
UB2	95.94	38.18	222.58
UB3	92.41	39.35	164.80
UB4	301.43	229.29	374.88

Fig. 9: Response Time by Various VMS

Fig (10,11,12,13) represent a graphical representation of the over response time of each virtual machine to process request and data respectively.

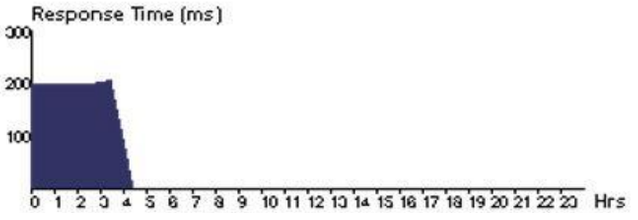

Fig. 10: UB1 Response Time

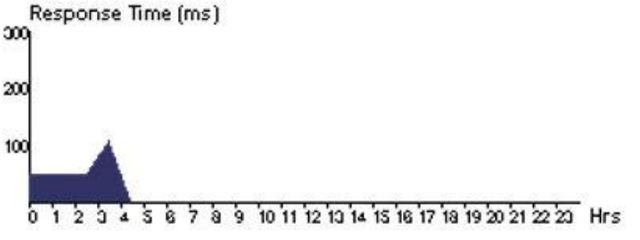

Fig. 11: UB2 Response Time

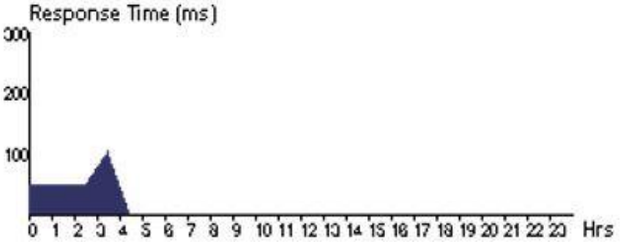

Fig. 12. UB3 Response Time

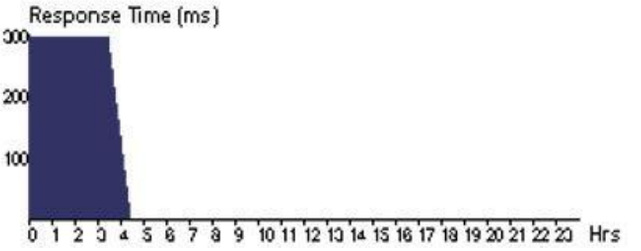

Fig. 13: UB4 Response Time

Fig. (14) represent data center Request Servicing Times

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	27.32	0.02	167.80
DC2	21.99	0.02	109.27

Fig. 14: Request Servicing Times

Fig. (15,16) represent Data Center Hourly Average Processing Times

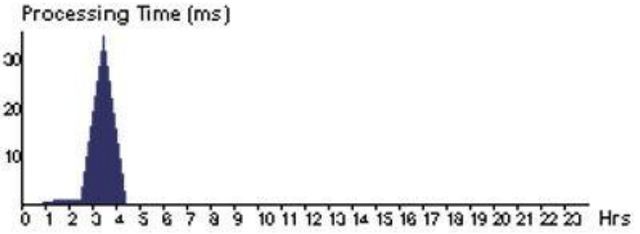

Fig. 15: DC1 Hourly Avg. Processing Time

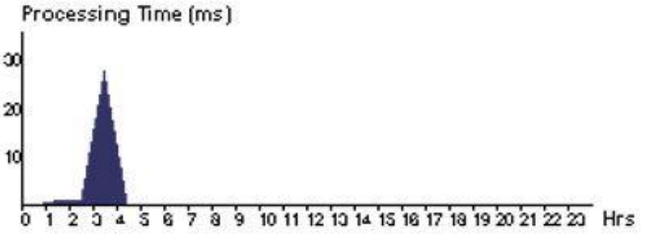

Fig. 16: DC2 Hourly avg. Processing Time

Fig (17,18): Represent Data Center Hourly Loading.

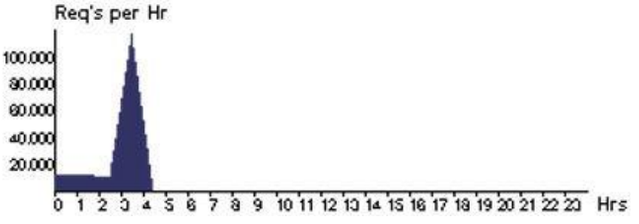

Fig. 17: DC1 Hourly Loading

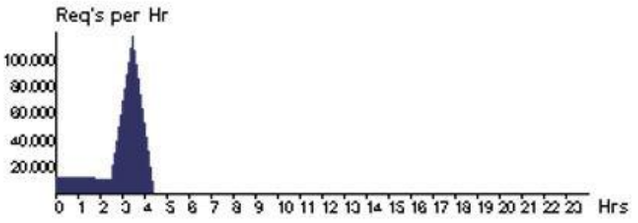

Fig. 18: DC2 Hourly Loading

Fig. (19) represent the cost efficiency of the proposed system.

Total Virtual Machine Cost ($):	64.58
Total Data Transfer Cost ($):	3.25
Grand Total: ($)	67.83

Data Center	**VM Cost $**	**Data Transfer Cost $**	**Total $**
DC2	32.25	1.61	33.87
DC1	32.32	1.64	33.96

Fig. 19: Cost Efficiency of System

Acknowledgement

I would wish to convey my earnest thanks to Asst. Prof. Ranu Lal Chouhan for his advice during my research work. As my supervisor, he has always encouraged me to stay concentrated on reaching my end. I must acknowledge the academic resources that I have got from Govt. Engineering College, Bikaner. I would like to thank administrative and technical staff members of the Department who have been kind enough to propose and help in their respective offices.

References

[1] T. Desai, and Prajapati, J. "A survey of various load balancing techniques and challenges in cloud computing," *International Journal of Scientific & Technology Research*, vol. 2, no. 11, pp. 158-161, 2013.

[2] L. Wang, J. Tao, and M. Kunze, "Scientific cloud computing: Early definition and experience," The *10th IEEE International Conference Computing and Communications,* 2008.

[3] M. Rana, S. Bilgaiyan, and U. Kar, "A study on load balancing in cloud computing environment using evolutionary and swarm based algorithms," *Control, Instrumentation, Communication and Computational Technologies (ICCICCT), International Conference on. IEEE,* 2014.

[4] I. Foster, Y. Zhao, I. Raicu, and S. Lu. "Cloud computing and grid computing 360-degree compared, IEEE Grid Computing Environments (GCE08) 2008, co-located with IEEE/ACM Supercomputing 2008," *2012 ACM/IEEE 13th International Conference on Grid Computing,* 2008.

[5] R. Buyya, R. Ranjan, and R. N. Calheiros. "Intercloud: A utilityoriented federation of cloud computing environments for scaling of application services," *International Conference on Algorithms and Architectures for Parallel Processing,* Springer Berlin Heidelberg, 2010.

[6] S. Ray, and A. D. Sarkar, "Execution analysis of load balancing algorithms in cloud computing environment," *International Journal on Cloud Computing: Services and Architecture (IJCCSA),* vol. 2, no. 5, p. 113, 2012.

[7] S. Katoch, and J. Thakur, "Load balancing algorithms in cloud computing environment: A review," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 2, Aug 2014.

[8] El-Gazzar, R. Fahim. "An Overview of Cloud Computing Adoption Challenges in the Norwegian Context." Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on. IEEE, 2014.

[9] S. S. Rajput, and V. S. Kushwah. "A review on various load balancing algorithms in cloud computing," *International Journal,* vol. 6, no. 4, 2016.

[10] More, S. R. Hiray. et al., "Load balancing and resource monitoring in the cloud," *Proceedings of the CUBE International Information Technology Conference,* ACM, 2012.

[11] M. Almubaddel, and A. M. Elmogy, "Cloud computing antecedents, challenges, and directions," *Proceedings of the International Conference on Internet of things and Cloud Computing,* ACM, 2016.

[12] A. Lenk, "What's inside the Cloud? An architectural map of the Cloud landscape," *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing,* IEEE Computer Society, 2009.

[13] B. Sotomayor, R. Santiago Montero, I. Martín Llorente, and I. Foster, "Virtual infrastructure management in private and hybrid clouds," *IEEE Internet Computing,* vol. 13, no. 5, pp. 14-22, 2009.

[14] S. Rajkumar, and J. Ojha, "A hybrid approach for vm load balancing in cloud using CloudSim*," International Journal of Science, Engineering, and Technology Research (IJSETR),* vol. 3, no. 6, pp. 1734-1739, 2014.

[15] A. A. Rajguru, and S. S. Apte, "Various strategies of load balancing techniques and challenges in distributed systems," *International Journal of New Innovations in Engineering and Technology,* vol. 5, no. 3, pp. 1-6.

[16] E. Gupta, V. A. Deshpande, "Technique based on ant colony optimization for load balancing in cloud data center*," In Information Technology (ICIT), 2014 International Conference,* pp. 12-17, IEEE, Dec. 2014.

[17] K. Li, G. Xu, G. Zhao, Y. Dong, and D. Wang, "Cloud task scheduling based on load balancing ant colony optimization," In *2011 Sixth Annual China Grid Conference*, pp. 3-9, IEEE, Aug. 2011.

[18] R. Kaur, and N. Ghumman, "Hybrid improved max min ant algorithm for load balancing in cloud," In *International Conference On Communication, Computing & Systems,* IEEE, 2014.

[19] K. Nuaimi, N. Mohamed, M. Nuaimi, and J. Al-Jaroodi, "A survey of load balancing in cloud computing: Challenges and algorithm," *2012 Second Symposium Network Cloud Computing and Application (NCCA),* pp. 137-142, IEEE, 2012.

[20] S. Kumar, and D. Singh, "Various dynamic load balancing algorithm in cloud environment: A survey," *International Journal of Computer Applications,* vol. 129, no. 6, pp. 4-9, Nov. 2015.

[21] A. Hans, and S. Kalra, "A comprehensive study of various load balancing techniques used in cloud based biomedical services*," International Journal of Grid and Distributed Computing,* vol. 8, no. 2, pp. 127-32, April 2015.