Analysis of Heterogeneous Device Characteristics in Round Robin Based Load Balancing Algorithm with Closest Data Center as Service Broker Policy in Cloud

Kumar PJ, N.Sivakumar, J.Prabhu, P.S.Ramesh, Suganya P

Abstract: ***Load balancing is an important aspect in cloud to share load among different virtual machines running on various physical nodes. The user response time which is an important performance metric is being highly influenced by the efficient load balancing algorithm for cloud data centers. Virtual machines which are part of the cloud data centers consist of various types of physical devices. The user response time is affected significantly by the capacity of physical devices that exist as part of the data centers. Several load balancing algorithms exist in the literature to allocate task effectively on various virtual machines running in data centers. We investigate the performance of round robin based load balancing algorithm with closest data center as service broker policy in cloud data centers. We have performed a simulation with data centers that consist of devices with different physical characteristics such as memory, storage, bandwidth, processor speed and scheduling policy using Round Robin load balancing algorithm with closest data centers as service broker policy. We present the merits of heterogeneous device characteristics in reducing the user response time and the data center request service time. We used Cloud Analyst, an open source simulation tool for cloud computing environment.***

Index Terms: Data Centers, Virtual Machines, Load Balancing, Closest Data Center, Round Robin

I. INTRODUCTION

The evolution and quick popularity of cloud computing paradigm has gained more number of users who generates millions of request simultaneously to request the resources offered by the cloud. Obviously the response time for a certain request increases proportionately to the number of request generated by the users simultaneously. As the number of requests increases the load on the server also increases which leads to an increased response time for the user. The response time of a user is based on the availability server with high computing power. Since cloud services are offered in a transparent manner, the user may not be aware of the device on which the task will be executed. The hardware perspective of devices varies largely in cloud environment. Load balancing algorithm helps to improve the execution of task in a shorter response time by properly allocating or distributing the task on several hosts by measuring the existing load on a particular host. Round robin algorithm distributes the tasks to available hosts or machines on equal fashion so that all host have equal number of tasks to execute irrespective of their individual capacity. At the other side, certain load balancing algorithm exist which allocates the task to the host based on their individual capacity. i.e., allocating more tasks to hosts which has higher computing capacity and less number of tasks to hosts with moderate computing capacity. Load balancing algorithm should be designed by considering the underlying difference in the physical characteristics of hosts are servers. Significant number of load balancing algorithms are proposed in the literature to allocate task based on different policies such as Hybrid algorithms, Round Robin, Weighted Round Robin, Greedy, Number of Connections Based and Random [15]. We investigate the performance of Round Robin based load balancing algorithm with closest data center as service broker policy for different host configuration in cloud. We measure the average user response time and data center request service time for a given scenario that consist of homogeneous and heterogeneous types of host in data centers. Each and every aspect of computing such as software, platform and infrastructure are now provided as a service by the cloud to its users who can pay and avail the service as per their need [15]. The user can be specific about their need which they want to get as a service from the cloud such as CPU clock cycles, memory, storage, bandwidth etc. Generally the load balancing system receives a number of request from the users and distributes it to the host based on a certain policy. Balancing based on Round robin allocates fair amount of tasks to hosts irrespective of their individual capacities. This may lead to a poor response time for the user. Policies which allocates the task considering the capacity of the individual host can improve the response time marginally. We have analyzed the impact of Round Robin load balancing algorithm with closest data center as service broker policy on a variety of cloud environments with different device characteristics.

The remaining of the paper is arranged as below. Section II discusses about existing works in the literature. Section III presents the experimental setup. Section IV present the simulation results and interpretations. Section V

Revised Manuscript Received on July 05, 2019.

Dr Kumar PJ, Assistant Professor [SG], School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu.

Dr N.Sivakumar, Assistant Professor [SG], School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu.

Dr J.Prabhu, Associate Professor , School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu.

Prof. P.S.Ramesh, Assistant Professor [SG], School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu

Prof. Suganya P, Assistant Professor , School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu

concludes the paper with future extensions.

II. Literature Survey

The work proposed in [1] considers the computing capacity of the server and the current work load on the server to allocate task to it. CLB (Cloud Load Balancing) a dynamic load balancing algorithm proposed by [1] ensures equal assignment of tasks to servers. An in-depth survey has been made on load balancing algorithms that exist in the literature by the authors in [2]. A comprehensive study has been made by [2] to provide an insight into various load balancing techniques. Nature inspired load balancing, load balancing in general, application specific, network based, map reduce in Hadoop and Agent supported load balancing has been proposed by [3]. Increasing throughput, provision of QoS, utilization of resources and improved performance are considered as key objectives for load balancing. A particle swarm based optimization load balancing is proposed by [4]. Balance between energy and performance are considered as key factor to increase profit in the proposed approach. The proposed approach shows an increase by 14% in the energy consumed and simulated using cloudsim. Various QoS metrics has been proposed by the authors in [5] which will be a part of the user request that should be satisfied by the load balancing algorithm. The service provider will be penalized upon violating the QoS requirements. Cloud deals with large amount of data compared to the other conventional networking environment. The work in [6] proposed a load balancing approach to handle massive data in cloud which is called as LBMM (Load Balancing Method for Massive Data) The authors in [7] proposed load balancing algorithm for cloud which is based on Artificial Bee. The authors in [8][11][12] proposed algorithms to minimize the response time and access time in cloud based on the QoS parameters of the user. The source code of the simulator CloudAnalyst used in this paper can be obtained from the online sources as mentioned in [9] [10]. The authors in [13][14] discusses about various simulation environment available for cloud computing and also discusses the execution of load balancing algorithm in it.

III. Cloud Computing Architecture and Load Balancing

Software as a Service [SaaS]	Multimedia Applications and Web services	Google App and Facebook
Platform as a Service [PaaS]	Software Tools	Microsoft Azure
Infrastructure as a Service [IaaS]	Infrastructure and Hardware (CPU, Memory and Bandwidth)	Amazon EC2, Data centers

Table I: Various Layers of Services provided in Cloud [15]

The individual entities of computing such as software, hardware, platform and infrastructure are provided as service to the user by various cloud vendor. Some of the popular cloud vendors are listed as below. Elastic Cloud compute by amazon and Rackspace providers of infrastructure as a service. The users of the cloud need to pay for only what they use from cloud. Windows azure and App Engine examples for Platform as a service. Google App and Salesforce provide software as a service [15]. The host system contains a queue of jobs from which the load on a particular system can be measured.

IV. Analysis of Round Robin Based Load Balancing Algorithm with Closest Data Center as Service Broker Policy in a Data Center with Homogeneous and Heterogeneous Host

Name	Location	Processor Family	OS	VMM	Price/VM in $/Hour	cost of Memory in $/second	Cost of Storage $/s	Cost of Data Transfer $/Gb
DC1	3	X86	Linux	Xen	1	0.05	1	1
DC2	2	X86	Linux	Xen	1	0.05	1	1
DC3	0	X86	Linux	Xen	1	0.05	1	1
DC4	1	X86	Linux	Xen	1	0.05	1	1
DC5	4	X86	Linux	Xen	1	0.05	1	1

A. Experimental Setup

Table II shows the basic configuration of data centers with entities such as geographical location of data center, processor family, operating system running on the host, virtual machine monitor, price per VM, price of memory, price of storage and data transfer. Table III shows the configuration of host with homogeneous characteristics. Memory of all hosts is set to 2048 MB, storage of all host is set to 1000Mb, bandwidth support is set to 1000 Mb, number of processors is set to 3 and virtual machine policy is set as TIME_SHARED. Table IV shows the configuration of host in data centers with heterogeneous characteristics. The size of memory is varied as 2048MB, 4096 MB and 8192 MB. Storage is varied between 1000 MB to 4000 MB. Available bandwidth is varied between 1000 to 4000 Mb. Processor speed is varied between 1000 instructions to 3000 instruction per second. Round robin algorithm is used as load balancing algorithm across virtual machines. Closest data center is selected as service broker policy.

ID	Memory (Mb)	Storage (Mb)	Available BW	Number of Processors	Processor Speed	VM Policy
0	2048	1000	1000	3	1000	TIME_SHARED
1	2048	1000	1000	3	1000	TIME_SHARED

2	2048	1000	1000	3	1000	TIME_SHARED

Table III: Homogeneous Data Center Host Configuration

ID	Memory (Mb)	Storage (Mb)	Available BW	Number of Processors	Processor Speed	VM Policy
0	2048	1000	1000	3	1000	TIME_SHARED
1	4096	2000	2000	3	2000	TIME_SHARED
2	8192	4000	4000	3	3000	TIME_SHARED

Table IV: Heterogeneous Data Center Host Configuration

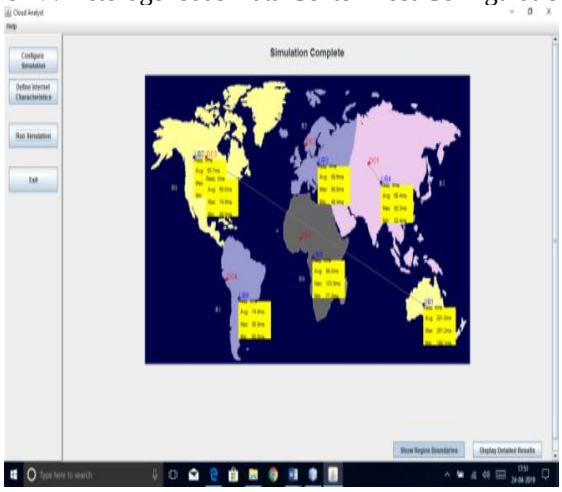

Fig.1: Assignment of user base and data center in Cloud Analyst Simulation Tool

Figure 1 shows the placement of data centers and user bases in different geographical location around the world.

V. RESULTS AND DISCUSSION

We run the simulation for a period of 12 hours in Cloud Analyst. The closest data center is selected as service broker policy. Round robin based load balancing is used to schedule virtual machines in data center. We created six user bases where each user base is located in different continents across the globe. The user grouping factor is set to 100. The number of request from each user is set to 100. Six different data centers are created with various virtual machine configurations. Basically we create two sets of host named homogeneous and heterogeneous. In homogeneous, the host machines are configured with same memory, storage, bandwidth and processor speed. In heterogeneous, the host machines are configured differently in terms of memory, storage, bandwidth and processor speed. During the simulation for both the cases, we used same bandwidth across different regions where data centers are deployed. The cost of virtual machine, memory, data transfer remains same for both the cases.

	Avg(ms)	**Min(ms)**	**Max(ms)**
Overall response time	211.28	44.37	659.44
Data Center Processing Time	11.09	0.10	76.68

Fig.2: Over all Response time (Homogeneous host in data center)

Userbase	Avg(ms)	Min(ms)	Max(ms)
UB1	340.00	168.49	528.26
UB2	122.60	47.44	207.52
UB3	109.89	44.37	185.70
UB4	144.67	49.38	254.66
UB5	326.22	61.15	659.44
UB6	224.17	54.14	421.14

Fig.3: Average Response Time per User base (Homogeneous host in data center)

Data Center	**Avg(ms)**	**Min(ms)**	**Max(ms)**
DC1	10.70	0.11	30.91
DC2	13.67	0.11	51.85
DC3	11.30	0.10	76.88
DC4	9.80	0.10	21.00
DC5	9.73	0.10	20.99

Fig.4: Data center request servicing time (Homogeneous host in data center)

	Avg(ms)	Min(ms)	Max(ms)
Overall response time	208.94	44.04	4118.81
Data Center processing time	5.42	0.03	4010.88

Fig.5: Over all Response time (Heterogeneous host in data center)

Userbase	**Avg(ms)**	**Min(ms)**	**Max(ms)**
UB1	338.09	168.26	523.26
UB2	120.02	47.34	206.30
UB3	106.50	44.04	4118.81
UB4	142.73	48.65	249.78
UB5	323.89	61.05	654.44
UB6	222.25	53.74	427.37

Fig.6: Average Response Time per User base

(Heterogeneous host in data center)

Data Center	Avg(ms)	Min(ms)	Max(ms)
DC1	5.39	0.04	27.58
DC 2	5.53	0.06	4010.88
DC 3	5.39	0.04	28.99
DC 4	5.41	0.03	20.25
DC 5	5.41	0.03	12.05

Fig.7: Data center request servicing time (Heterogeneous host in data center)

Figure 2 and 5 shows the overall user response time for homogeneous and heterogeneous configurations. The response time is low in case of heterogeneous configuration of host in data centers compared to the homogeneous. This is because of availability of high profile machines in the heterogeneous configurations that completes the task in significantly less amount of time. Figure 3 and 6 shows the response time for individual user bases for homogeneous and heterogeneous configuration. As expected the heterogeneous configuration produces less response time compared to the homogeneous configuration. Figure 4 and 7 shows the data center request servicing time for both cases. The heterogeneous configuration takes only 50% of the time taken in homogeneous configuration. Hence we suggest the data centers to be equipped with heterogeneous systems in order to produce faster response to the user which would improve the overall performance of cloud vendors also.

V. CONCLUSION

Measures to improve the performance of the cloud services are key importance to the vendors of cloud. Load balancing algorithm aims at improving the performance of the system by distributing the task to all hosts in the cloud based on certain policy. In addition, the characteristics of physical devices on which the virtual machine runs plays a crucial role in determining the performance of the system. In this paper, we investigate the performance of Round Robin load balancing algorithm with different data centers configuration. We use closest data center as service broker policy along with load balancing. The hosts in the data centers are configured as homogeneous and heterogeneous and the scenario is simulated using an open source simulation tool called Cloud Analyst. The result shows a significant difference in response time and data center request servicing time when the hosts in data centers are heterogeneous compared to the homogeneous hosts. Hence we suggest the configuration of data centers with more heterogeneous systems that adds power to data centers so that response time of the user and data center request servicing time will be reduced significantly.

REFERENCES

1. Shang-Liang ChenYun-Yao Chen Suang-HongKuo, “CLB: A novel load balancing architecture and algorithm for cloud services, Computers & Electrical Engineering, Volume 58, Pages 154-160, February 2017.
2. 2. Minxian Xu Wenhong Tian Rajkumar Buyya, “A survey on load balancing algorithms for virtual machines placement in cloud computing”, Concurrency and computation Practice and Experience, Wiley Online Library, Volume 29, Issue 12, march 2017.
3. 3.Einollah Jafarnejad Ghomia Amir Masoud RahmaniaNooruldeenNasih Qaderb, “Load-balancing algorithms in cloud computing: A survey”, Journal of Network and Computer Applications, Volume 88, , Pages 50-71, June 2017
4. Seyed Ebrahim Dashti & Amir Masoud Rahmani , “Dynamic VMs placement for energy efficiency by PSO in cloud computing”, Journal of Experimental & Theoretical Artificial Intelligence, Volume 28, Issue 1-2: Advances and Applications of Swarm Intelligence, 2016
5. M. H. Ghahramani, MengChu Zhou,Chi Tin Hon, “Toward cloud computing QoS architecture: analysis of cloud systems and cloud services”, EEE/CAA Journal of Automatica Sinica, Volume 4, Issue 1, PP:6-18, Jan 2017
6. Jianhua Peng, Ming Tang, Ming Li & Zhiqin Zha,, “A Load Balancing Method For Massive Data Processing Under Cloud Computing Environment”, Intelligent Automation & Soft Computing, Issue 4, PP: 547-553, 2017.
7. Yao, J.H., Ju-hou, Load Balancing Strategy of Cloud Computing Based On Artificial Bee Algorithm in Computing Technology and Information Management (ICCM), IEEE: Seoul. p. 185 - 189. 2012
8. Kumar P.J., Ilango P., “MQRC: QoS aware multimedia data replication in cloud”, International Journal of Biomedical Engineering and Technology, Vol.No:25, Issue.No:2/3/4, PP:250-266, 2017
9. cloudsim.cloudbus;Available from: http://www.cloudbus.org/cloudsim/.
10. 10.https://sourceforge.net/projects/ cloudanalystnetbeans/
11. Kumar P.J., Ilango P., “BMAQR: Balanced multi attribute QoS aware replication in HDFS”, International Journal of Internet Technology and Secured Transactions, Vol.No:8, Issue.No:2, PP:195-208,2018.
12. P.J.Kumar, P.Ilango, “An Optimized Replica Allocation Algorithm Amidst of Selfish Nodes in MANET”, Wireless Personal Communications,Vol.No:94, Issue.No:4, PP:2719-2738,2017.
13. Pakize, S.R., S.M. Khademi, and A. Gandomi, Comparison Of CloudSim, CloudAnalyst And CloudReports Simulator in Cloud Computing. International Journal of Computer Science And Network Solutions, 2: p. 19-27, 2014
14. Ray, S. and A. De Sarkar, Execution Analysis of Load Balancing Algorithms In Cloud Computing Environment. International Journal on Cloud Computing: Services and Architecture (IJCCSA), 2(5): p. 1-13,2012..
15. Hafiz Jabr Younis, Alaa Al Halees, Mohammed Radi, “Hybrid Load Balancing Algorithm in Heterogeneous Cloud Environment”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-3, July 2015 PP: 61-65

AUTHORS PROFILE

Dr Kumar PJ is currently working as Assistant Professor [SG} in the School of Information Technology and Engineering, VIT, Vellore, Tamil Nadu. He pursued his under graduation in Engineering with Electrical and Electronics as specialization, Master in Engineering with Computer Science as specialization and Ph.D in Cloud and Mobile Adhoc Networks. His area of interest in research includes Operating systems, Distributed systems, Mobile Adhoc Networks and Internet of Things. He has published numerous papers in reputed journals with impact factor and filed an Indian patent recently for his innovative work in Routing protocols for MANET.

.