IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 5, May 2016.
www.ijiset.com
ISSN 2348 – 7968

A Simulation of Load Balancing Policy Of Virtual Machines In a Cloud Computing Environment

B.Santhosh Kumar[1], Dr.Latha Parthiban[2]

[1] Dept. of CSE,G.Pulla Reddy Engineering College,
Kurnool,518002,India.
[2] Dept. of CSE, Community College, Pondicherry University,
Pondicherry,India.

Abstract

In the current IT world cloud computing is considered to be the most promising and one of the rapid growing technologies. The main strength of the cloud computing technology is to provide pay-per-usage model which drastically reduces the cost and complexity in developing and maintaining an application. The core portion of cloud computing is to provide virtualization process. One of the main aspects of virtualization is to create and maintain virtual machines effectively. This paper focuses on developing an algorithm to distribute the load among virtual machines based on their priorities. The algorithm has been developed and results have been simulated using the Cloud Analyst tool.

Keywords—*Cloud Computing, Virtualization, Cloud Analyst.*

1. Introduction

The services provided by the cloud can be broadly categorized as Software-as-a-service (SaaS), Platform-as-a-service (PaaS) and Infrastructure-as-a-service (IaaS).The applications which are a part of cloud and which are accessible to the users across the globe can be considered as a part of SaaS. Services like Gmail, face book fall under this category. The combination of hardware and software which are necessary to make the application run can be categorized as PaaS. The concept of Programming tools, web servers, database servers can be considered as PaaS. The Network services which are necessary to maintain the application effectively are considered as IaaS. Virtual machines, load balancers, routers are part of IaaS. The back bone of IaaS offerings is the process of virtualization. Virtualization can be considered as different configurations running as a part of single machine. This mainly reduces the cost of introducing numerous CPUs to maintain and run the application. In order to simulate the concept of virtualization and model the cloud platform a tool 'Cloud Analyst' has been developed by the CLOUDS laboratory established in university of Melbourne, Australia.

Table I. Region codes along with their names

Region	Cloud Analyst Region Id
North America	Ro
South America	R1
Europe	R2
Asia	R3
Africa	R4
Oceania	R5

The main aim of the Cloud Analyst is to model the cloud platform based on user requests which are arriving from different regions of the world. A Group of users requesting from a region is considered to be a user base and denoted by UB. The requests are sent to the data centers which can be spread across the globe and indicated by DC. The regions and their corresponding codes are depicted in Table I and the corresponding boundaries in the simulator are also indicated in Fig. 1.

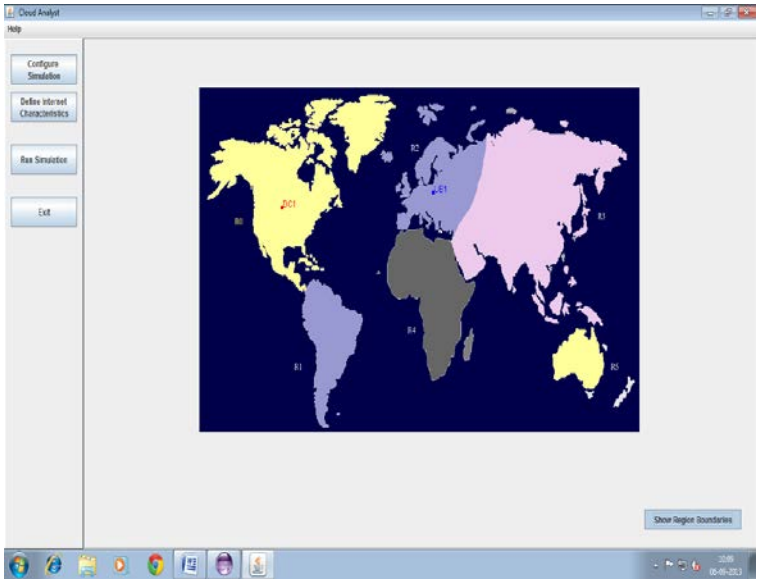

Fig.1. Region Boundaries in Cloud Analyst

A cloudlet which is considered to be a mini cloud is created and submitted to different data centers of the service providers across the globe. The simulator allows the user to specify the number of virtual machines and their corresponding images which should be

ISSN 2348 – 7968

created in each data centre. Depending on the requests from each user base the traffic is diverted to different data centers .The data centers create the virtual machine instances and submit the requests according to load balancing policies such as round robin, throttled requests. A comprehensive understanding of the cloud analyst tool can be obtained using [1] and every aspect of the tool can be explored in depth. In this paper an attempt is made to develop a new load balancing policy based on priority of virtual machines and results are simulated showing better performance than the already existing policies.

2.Related Work

A considerable amount of work has been done in framing the load balancing policies in a cloud environment. The work done in [2] mainly focused on how the virtual machines can be load balanced so that they can be efficiently used. A comparative study in to distributed load balancing algorithms in a cloud computing environment has been made in [3]. The work presented in [4] introduced a load balancing algorithm based on cloud partitioning and game theory. GangSim which was mainly used for grid related studies is developed in [5]. The work in [6] introduced a SimGrid simulation framework which also supported simulation of scheduling in distributed applications. The authors in [7] emphasized OptorSim and gave the idea of simulation of dynamic grid replication strategies. The research done in [8] mainly focuses on toolkit for modeling and simulation of grid computing known as GridSim.

As the technology kept on changing and a rapid advancement has been made from grid computing to a large scale environment and pay-per-usage models the cloud computing gained importance and eventually this led to the researchers focus on developing the simulators in that area. A significant one among them was the CloudSim developed in [9], a toolkit mainly developed for the modeling and simulation of cloud computing environment. It was developed by the CLOUDS laboratory, university of Melbourne which gave a lot of scope to the researchers to simulate the algorithms applicable to the cloud. This type of simulator was much necessary because to simulate the cloud environment a lot of resources have to be purchased which will incur more money and the setup environment will also be cumbersome. But this tool gives an opportunity to explore the things on the PCs assuming as if the traffic is from the entire globe. Based on CloudSim a number of cloud related tools have been developed .The visual tool which makes many users understand the application easily and which simulates the traffic of rich online applications such as face book graphically is Cloud Analyst developed by the authors in [10]. It allows the users to specify the number of data centers for processing the requests, the requests from each of the regions in the world and number of virtual machines, physical resources at each data centre. It graphically displays the results and gives the response times based on the chosen strategies.

3. Proposed Work

Currently the load balancing policies followed for the virtual machines are Round Robin, Equally Spread Current Execution Load and Throttled request. A new algorithm which is based on the priority of the virtual machines is given and it is simulated using Cloud Analyst.

Algorithm **PriorityLoadBalancer**

Input: Task ‘T’ to be allotted to a virtual machine for processing.

1. Initialize vmCount ← number of virtual machines created in a data centre.
2. Create a hash map VmPriorityList<vmid, vmpriority> which stores the virtual machine number and its priority.
3. Based on the factors such as previous response times, regions populate the hash map with the number of virtual machine and its priority.
4. Store it in VmPriorityList.
5. If task ‘T’ is in waiting state and has to be assigned to any virtual machine DO

```
For i ← 0 to vmCount-1
    For j ← i+1 to vmCount-1
        If (VmPriorityList.Get (i). vmpriority >= VmPriorityList.Get (j).vmpriority)
            Assign ‘T’ to virtual machine with vmid ‘i’
        End If
    End For
End For
```

6. The virtual machine with highest priority is assigned first and moved to a circular queue VmAllottedQ.
7. The assignment of the tasks to virtual machines is continued in an ascending order of priorities and placing the already allotted ones in to VmAllotedQ.
8. The procedure is repeated until all the tasks are fulfilled.

END.

ISSN 2348 – 7968

To demonstrate the algorithm using Cloud Analyst and to display the results accordingly first the configuration screen has been modified as shown in Fig. 2.

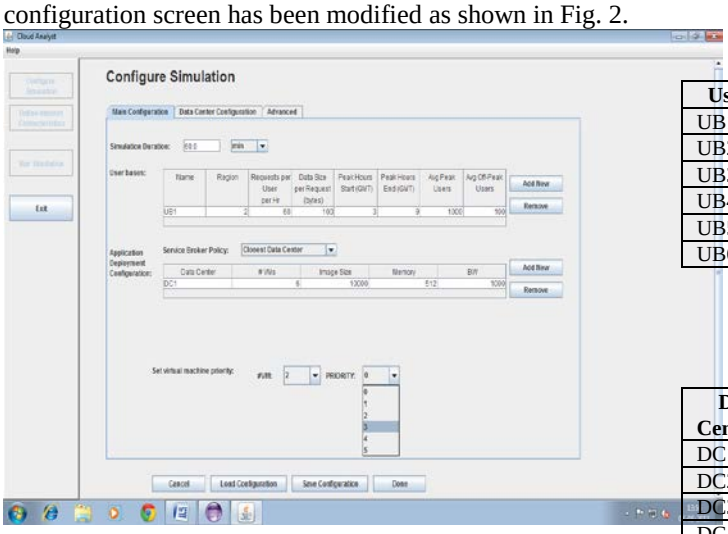

Fig. 2. Configuration Screen To Set Priorities Of Virtual Machines

The proposed algorithm is added as part of the Cloud Analyst tool in a separate file called vmPriorityLoadBalancer which extends the abstract class vmLoadBalancer. Two methods GetVmPriorityList () as well as AssignVmPriorityList () have been added to achieve the purpose of assigning and retrieving the priorities assigned to the virtual machines. The total simulation process has been done by creating total of 6 data centers and 6 user bases with virtual machines in all the data centers. Fig.3. demonstrates the output of simulation of the priority based algorithm.

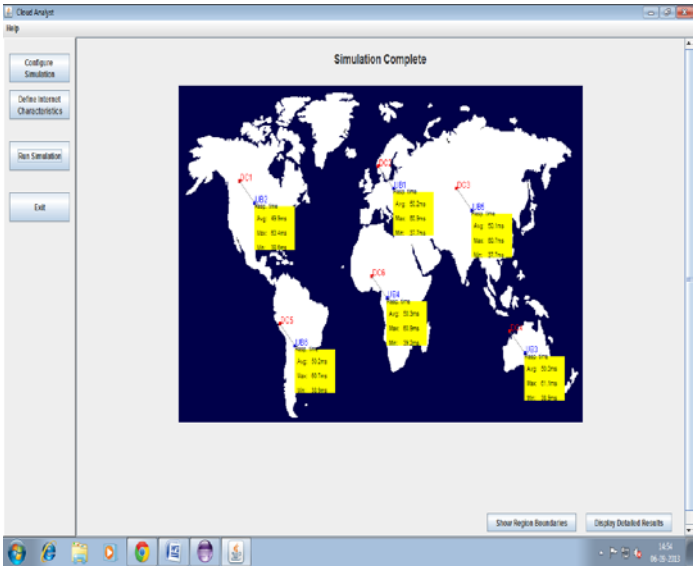

Fig. 3. Complete Simulation Of The Priority Based Algorithm.

The results are demonstrated according to the response time at each user base and at each region in Table II and Table III respectively.

Table II. Data Illustrating Response Time By Region

User Base	Avg(ms)	Min(ms)	Max(ms)
UB1	50.20	37.65	60.90
UB2	49.93	38.61	60.36
UB3	50.03	38.91	61.07
UB4	50.31	39.18	60.94
UB5	50.14	37.66	60.66
UB6	50.17	38.91	60.67

Table III. Data Indicating Data Centre Request Servicing Times

Data Centre	Avg(ms)	Min(ms)	Max(ms)
DC1	0.48	0.02	0.86
DC2	0.51	0.02	0.90
DC3	0.51	0.02	0.91
DC4	0.51	0.02	0.91
DC5	0.52	0.03	0.92
DC6	0.53	0.03	0.92

.

4. Performance Analysis

The proposed algorithm has been compared with other existing algorithms like Round Robin, Throttled and the results have proved that Priority Based algorithm behaves much better when compared to both throttled and round robin as illustrated in Fig. 4. There are some situations where this algorithm might lead to starvation. When there are more virtual machines with the same priority then a suitable policy has to be accommodated so as to continue the assignment of tasks to virtual machines so that they can serve as much as possible. When the results are observed closely there are situations where both throttled as well as round robin have resulted in the same performance.

The six data centers have been taken for our simulation and six user base regions have also been considered. The same data is repeated for all the algorithms except choosing the load balancing policy option where we can choose either Round Robin or Priority Based Allocation or Throttled. One important thing to remember while using cloud analyst is after performing one time simulation it doesn't allow to repeat the procedure immediately. So the entire configuration has to be loaded in to a file and then the process has to be repeated for all the algorithms. The X-axis represents data centers and the Y-axis represents data centre request servicing time in milliseconds.

ISSN 2348 – 7968

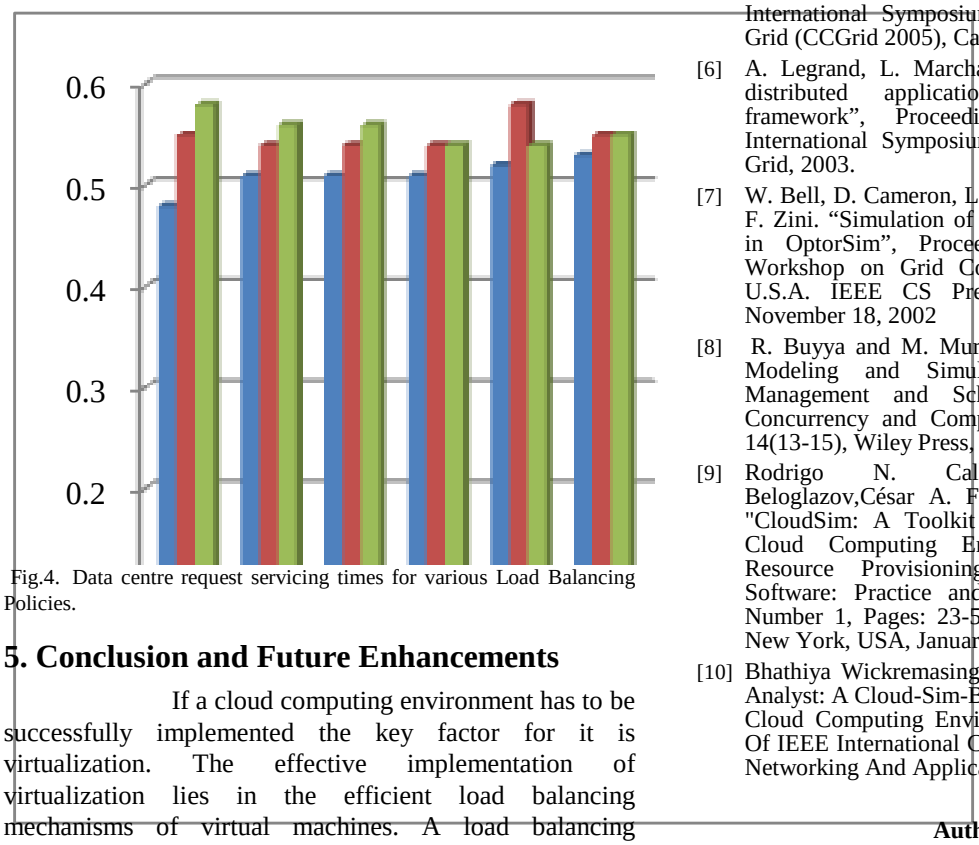

Fig.4. Data centre request servicing times for various Load Balancing Policies.

5. Conclusion and Future Enhancements

If a cloud computing environment has to be successfully implemented the key factor for it is virtualization. The effective implementation of virtualization lies in the efficient load balancing mechanisms of virtual machines. A load balancing algorithm based on the priority of the virtual machines has been proposed and the results are simulated using the cloud analyst tool. The comparison shows that this algorithm yields some best results when compared to the other existing algorithms. Further enhancements can be made by further combining this algorithm with most efficient factors based on response times, number of times virtual machines have served the requests and many other factors.

References

[1] BhathiyaWickremasinghe,"http://www.cloudbus.org/students/MEDC_Project_Report_Bhathiya_318282.pdf", 2009.

[2] Jianhua Gu,Guofei Sun,Tianhai Zhao, "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment" Parallel Architectures, Algorithms and Programming (PAAP), 2010 Third International Symposium on cloud, 18-20 Dec 2010.

[3] Randles,M.,Lamb,D.,Taleb-Bendiab,A, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing" proceedings of Advanced Information Networking and Applications Workshops (WAINA), 2010.

[4] Gaochao Xu, Junjie Pang, and Xiaodong Fu."A Load Balancing Model Based on Cloud Partitioning for the Public Cloud",proceedings of Advanced Information Networking and Applications Workshops (WAINA), 2010.

[5] C. L. Dumitrescu and I. Foster. "GangSim: a simulator for grid scheduling studies", Proceedings of the IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2005), Cardiff, UK, 2005.

[6] A. Legrand, L. Marchal, and H. Casanova," Scheduling distributed applications: the SimGrid simulation framework", Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid, 2003.

[7] W. Bell, D. Cameron, L. Capozza, P. Millar, K. Stockinger, F. Zini. "Simulation of dynamic Grid replication strategies in OptorSim", Proceedings of the 3rd International Workshop on Grid Computing (Grid 2002), Baltimore, U.S.A. IEEE CS Press: Los Alamitos, CA, U.S.A., November 18, 2002

[8] R. Buyya and M. Murshed," GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing." Concurrency and Computation: Practice and Experience, 14(13-15), Wiley Press, Nov.-Dec., 2002.

[9] Rodrigo N. Calheiros,Rajiv Ranjan, Anton Beloglazov,César A. F. De Rose and Rajkumar Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms." Proceedings of Software: Practice and Experience (SPE), Volume 41, Number 1, Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.

[10] Bhathiya Wickremasinghe ,Roderigo N. Calherios, "Cloud Analyst: A Cloud-Sim-Based Visual Modeler For Analyzing Cloud Computing Environments And Applications". Proc Of IEEE International Conference On Advance Information Networking And Applications ,2010

Authors Information

B.Santhosh Kumar did his B.Tech from Rajeev Gandhi Memorial College Of Engineering and Technology and M.S from Western Kentucky University, USA. He is currently working as an assistant professor in CSE department in G.Pulla Reddy Engineering College, Kurnool. His areas of interests include Cloud Computing, Cryptography and Web Technologies. He is currently pursuing his Ph.D. from Pondicherry University. He has published papers in six international journals.

Dr.Latha Parthiban has obtained her B.E in Electronics and Communication Engineering from University of Madras.Her experience spans over 16 years in various Engineering colleges and her research interest includes Soft computing, Expert systems, Image Processing and cloud computing. She has published papers in 40 international journals and presented papers in 45 international and national conferences. She has also published a book in the area of computer aided diagnosis.