

Load Balancing of Resources Using Virtual Machines in a Cloud Computing Environment

Prof R. M. Wahul, Seema Kurawale, Anuja Joshi, Priyanka Langhe, Shital Aher
Computer Department, Savitribai Phule Pune University
Maharashtra, India

Abstract—

Currently Cloud computing is an emerging concept which efficiently combining many fields of computing. Cloud computing provides us various services and software and processing capacity over the internet. The biggest advantage of cloud computing is that it is capable of handling a huge amount of growing work in a predestined manner for the usage of the business customers. Here in cloud computing the main concept for cloud computing is virtualization which generalizes the physical infrastructure and makes it easy to use and manage. In our project virtualization is used to allocate resources based on their needs and also supports green computing concept. Another concept that is "Skewness" is introduced here in which the same is minimized to combine various workloads to improve the utilization of the server. And Overload avoidance is maintained in this project which leads to achieve good performance.

In order to achieve this dynamic resource allocation concept we have implemented Web hosting scenario.

Keywords—Cloud computing, dynamic resource allocation, Skewness, virtualization, Overload avoidance.

I. INTRODUCTION

This project mainly concentrates on two main concepts, i.e. overload avoidance and green computing. Here we also learn about how a cloud service provider is best multiplexing its virtual resources. From this we can understand a cloud model which is expected to have a scale up and down in order to manage the load variation. The main advantage is that it also reduces the hardware cost and saves electricity. Number of times the mapping of virtual machine to the physical are hidden from the cloud users. So it's the responsibility of the cloud providers to make the resources meet their needs. In cloud computing the VM live migration technology makes the VM and PM mapping possible when the execution is running. Here in our project the main goals that we achieved are:

- The capacity of PM should be able sufficient to satisfy the needs of the VM's running on it. Thus we should maintain the utilization of PMs low as possible.
- The number of PM's should be minimized. So in order to achieve this we have to maintain the utilization of PMs high. To measure the utilization of the server we introduce a concept of threshold value and by minimizing this we can find the utilization of the servers.

II. LITRETURE SURVEY

There are various work done in Dynamic resource allocation using various techniques which are described in this section V, Holy Angel Jenitha and R.Veeramani presented "Dynamic Memory allocation using Ballooning and virtualization in cloud computing". This paper present a dynamic memory allocation system that improves the performance of virtual machines by increasing/decreasing the memory usage based on the process running on it. The mina modules of the system are Memory Statistics Collector and Resource Management and Balloon list details. Here important module that is Memory Statistics Collector is a script that collects OS level details on memory usage of Guest OS and page fault.

Another research by Kaleeswari and Noble Mary Juliet presented "Dynamic Resource Allocation Using Elastic Cloud Computing Service". Elastic Compute Cloud (EC2) service is used for dynamic resource allocation for cloud computing environment this service is provided by Amazon Web Service (AWS) which is a collection of remote computing services that together make up a cloud computing platform that is offered over the Internet by Amazon. EC2 allows deployment of applications by providing a web a service through which a user can boot an Amazon Machine Image to create virtual machine.

Davide Tammaro [1] said that computing job requests that are characterized by their arrival and teardown times and also a predictive profile of their computing requirements during their activity period. Here they are assuming a prior knowledge of the predicted computing resources required by end-users Davide propose and investigate several algorithms with different optimization criteria. In this project prediction errors may occur resulting in some cases in the drop of one or several computing requests.

In [9] Chandrashekhar S. Pawar mention that services execute by priority based where application is pre-emptable. They have presented an algorithm which is divided into four steps where first work is distributed among working VM by load balancer then forming a task list based on priorities. And then on the third steps Cloud min-min scheduling (CMMS) used for scheduling and Priority Based Scheduling Algorithm (PBSA) used at the last.

III. IMPLEMENTATION DETAILS

We have implemented our project by using a web hosting scenario. We have developed our project using Java programming language and MySQL database.

The architecture of this system is to create a domain for a particular company. In web hosting scenario there are two users of the system first one is Customer or we can say a user and another is Cloud Service Provider or Admin. At first User registers with the system by adding his details, Username and Password. By using this username and password he can login to system. After creating a domain that has to be sent to the provider, the admin will check if some other domain exists with the same name or not. If not the provider will send the approval and if some other domain exists with the same name then an acknowledgement will be given by the provider. User then adds the details like space required to his site, Web Site duration and other required details. This request is then sent to admin to create a domain and allocate a space.

On the admin site the admin can view the various requests for web hosting. Admin approves the request for web hosting. Based on the required space a particular server is allocated to the web site. All servers have a specific threshold value. If the server gets overloaded then the resources get dynamically shifted to another server. If all server space is full then an alert is given to admin to add a server. Admin can then add the server and fulfill the user requests.

Here we have also implemented the concept of green computing in which if a web site period gets over the sites get deleted and resources allocated to that site gets revealed.

If the approval is given by admin then the webpage will be hosted successfully. User can view the status of his website. Admin can view the status of each server that is the load on each server. He can also set the threshold value of each server.

Here in order to achieve the goal of dynamic resource allocation we used the concept of coldspot and hotspot based on the server's threshold value.

The various screenshots of our application are show below

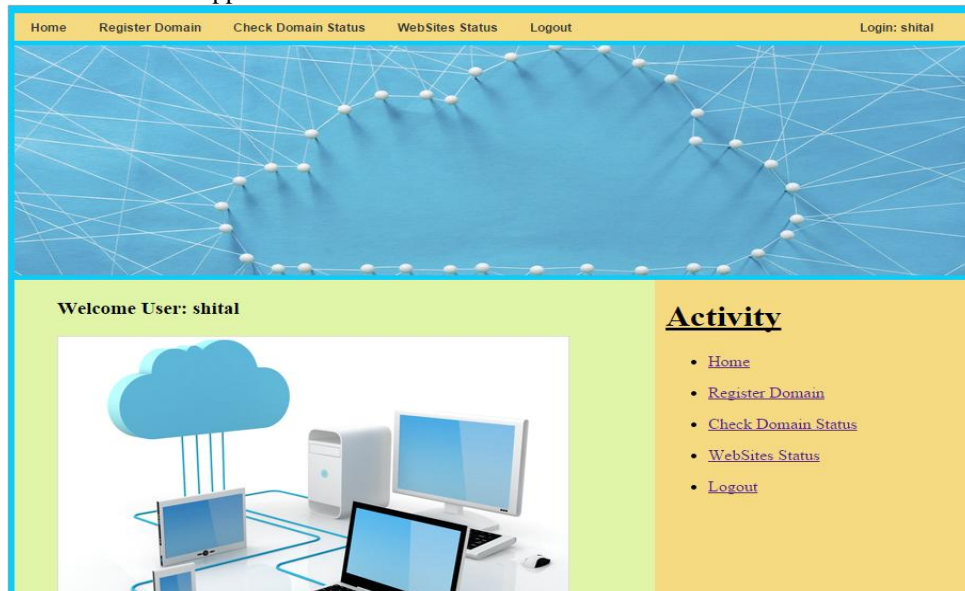


Fig1. User Home page

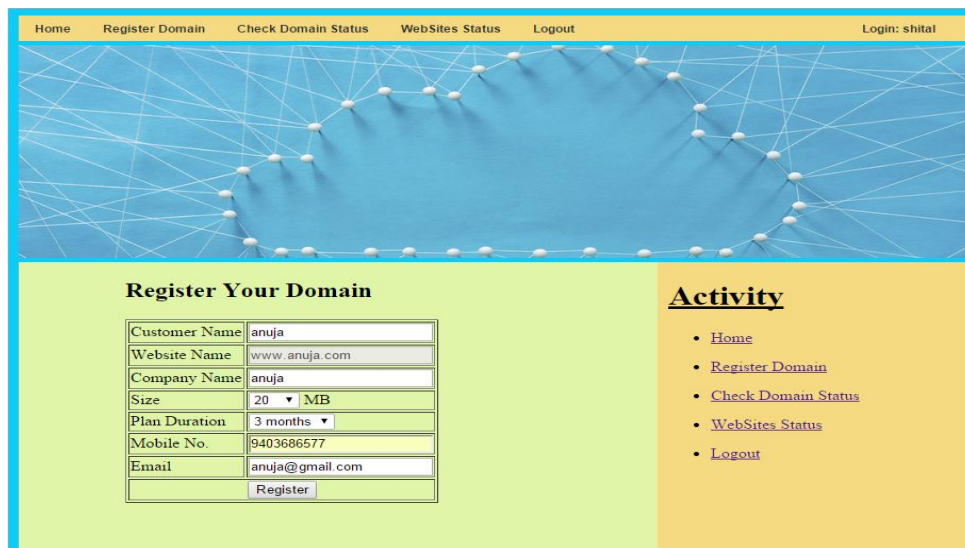


Fig2. User Home showing the domain details added by user

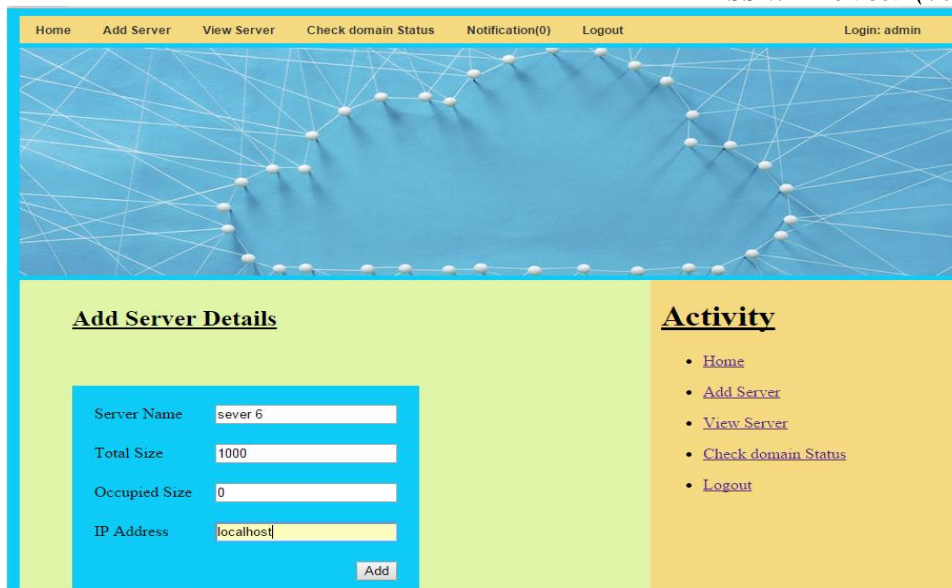


Fig3. Admin Home showing Server details added by Admin

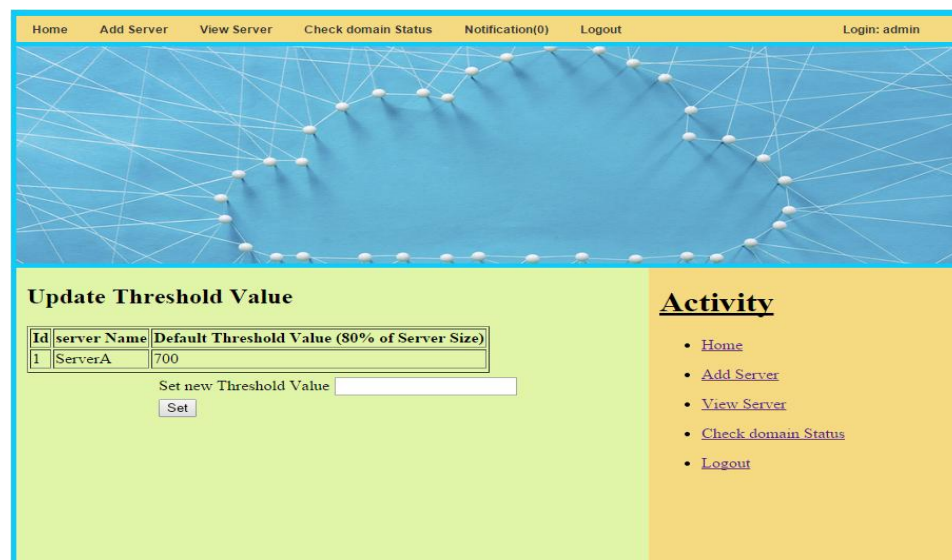


Fig4. Admin Home showing the threshold value updating

IV. CONCLUSION

Here, in this paper we have presented the design and implementation of a resource management system for cloud computing services. Our system multiplexes virtual physical resources adaptively based on the changing demand of various web sites. Our System achieves both overload avoidance and green computing for systems with multi resource constraints. As future work of this project we plan to add more security with our existing work and also to integrate the mobile environment where it requires optimum resource utilization.

ACKNOWLEDGMENT

We would like to take this opportunity to express our profound gratitude and deep regards to the faculty members of the Computer Department of MES College Of Engineering, Pune, for their exemplary guidance, monitoring and constant encouragement throughout the research work, which helped us in completing this paper through various stages. We are obliged for the valuable information provided by them in their respective fields. .

REFERENCES

- [1] Davide Tammaro and Elias A. Doumith "Dynamic Resource Allocation in Cloud Environment Under Time-variant Job Requests" at 2011 Third IEEE International Conference on Cloud Computing Technology and Science.
- [2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," in Proc. of the ACM Symposium on Operating Systems Principles (SOSP'03), Oct. 2003.
- [3] "Amazon elastic compute cloud (Amazon EC2), <http://aws.amazon.com/ec2/>."

- [4] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," in Proc.of the Symposium on Networked Systems Design and Implementation(NSDI'05), May 2005.
- [5] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in Proc. of the USENIX Annual Technical Conference, 2005.
- [6] Lim, and G. Hutchins, "Fast transparent migration for virtual machines," in Proc. of the USENIX Annual Technical Conference, 2005.
- [7] Chandrashekhar S. Pawar and Rajnikant B. Wagh "Priority Based Dynamic Resource Allocation in Cloud Computing with Modified Waiting Queue" at 2013 International Conference on Intelligent Systems and Signal Processing (ISSP)