

Importance of Virtual Machines in a Cloud Computing Environment

Suneetha Taduri

Lecturer in Computer Science, Department of Computer Science and Engineering, Loyola Academy Degree and PG College, Secunderabad, Alwal, T.S, India.

Abstract: Efficiently combining many fields of computing is an emerging concept of Cloud computing. Over the Internet Cloud computing is providing various services with software and its processing capacity for the utilization of various servers. For the usage of the business customers cloud computing is capable of handling a huge amount of growing work in a predestined manner, it is the main advantage of cloud computing. With virtualization cloud computing generalizes the physical infrastructure and makes this easy to use and manage the various servers. Based on the users' needs resources are allocated with the usage and implementation of virtualization and at the same time it supports the green computing concept. To improve the utilization of the main server "Skewness" is introduced with which the same is minimized to combine various workloads. And Overload avoidance is maintained which leads to achieve good performance. Web hosting scenario is used to achieve this dynamic resource allocation concept.

Keywords: Cloud computing, dynamic resource allocation, skewness, virtualization, overload avoidance.

1. INTRODUCTION

Overload avoidance and green computing are two main concepts in this paper. Along with these I can also learn about how all virtual resources are best multiplexed by a cloud service provider. In order to manage the load variation is having a scale up and down among the various virtual resources a cloud model is introduced. With this model the hardware cost is reduced and saves the electricity. With this main advantage of cloud model is the system can manage the interactions among the virtual machine and the physical machine. So that the virtual machine can be mapped number of times to physical are hidden from the cloud users. To meet the cloud users needs now it's the main responsibility of the cloud providers to make the resources availability at any time. To make the VM and PM mapping with VM live migration technology in cloud computing there is a possibility of executing and running the predicted computing resources required by end-users. The two main goals achieved here are:

When the VM's running on PM, the capacity of PM should be sufficient to satisfy the needs of the VM. This can be achieved when the utilization of PMs are maintained as low as possible.

To minimize the number of PM's is the second goal. So to achieve this, the maintenance and utilization of PMs must be high. Threshold value is used to measure the utilization

of the server and by minimizing this we can find the utilization of the servers.

II. LITERATURE SURVEY

Using various techniques on dynamic resource allocation there are various work done which are described in the "Dynamic Memory allocation using Ballooning and virtualization in cloud computing" presented by

V. Holy Angel Jenitha and R. Veeramani. "To improve the performance of virtual machines by increasing/decreasing the memory usage based on the process running on it, a dynamic memory allocation system is used". Memory Statistics Collector, Resource Management and Balloon list details are main modules of this system. OS level details on memory usage of Guest OS and page fault are collected by the script i.e., Memory Statistics Collector module.

"Dynamic Resource Allocation Using Elastic Cloud Computing Service" presented by Kaleeswari Noble Mary Juliet. For dynamic resource allocation for cloud computing environment Elastic Compute Cloud (EC2) service is used and which is provided by Amazon Web Service (AWS) which is a collection of remote computing services. EC2 and AWS make up a cloud computing platform that is offered over the Internet by Amazon. A web service is provided by EC2 which allows deployment

of applications through which an Amazon Machine Image can be booted by a user to create virtual machine.

“Computing job requests are characterized by their arrival and teardown times and also during their activity period a predictive profile of their computing requirements are to be listed out” is given by Davide Tammaro [1]. Predicted computing resources are required by end-users based on a prior knowledge that is resources are actively responding with in the time arrivals or not. With different optimization criteria several algorithms proposed and investigated by Davide. In the drop of one or several computing requests prediction errors may occur resulting in some cases.

“Application is preemptable when services executed by priority based” mentioned by Chandrashekhar, S. Pawar and Rajnikant B. Wagh[6]. “An algorithm is divided into 4 steps where first work is distributed among working VM by load balancer then based on priorities forming a task list” presented by them. Cloud min-min scheduling (CMMS) used on the third steps for scheduling at the last Priority Based Scheduling Algorithm (PBSA) used.

III. IMPLEMENTATION DETAILS

A web hosting scenario developed using Java programming language and MySQL database and is implemented.

A domain for a company is created by the architecture of this system. There are two users of the system in web hosting scenario. The customer or any user is the first user and the Cloud Service Provider or Admin is the second user. User registers with the system by adding his details like Username and Password at first. The registered user can login in to the system by using these username and password. To the provider a domain has been sent by the registered user, then the admin will check with the same name if some other domain exists or not. Provider will send the approval if that domain does not exists and an acknowledgement will be given by the provider if some other domain exists with the same name. After receiving the acknowledgement the user has to adds the required details like space required to site, web site duration. Admin will create a domain and allocate a space after receiving the re-request from the user.

For web hosting the various requests can be viewed by the admin on the admin site. Request approval done by the admin for web hosting. A particular server is allocated to the web site based on the required space. A specific

threshold value allotted to all servers so that if the server gets overloaded then the resources get dynamically shifted to another server. An alert is given to admin to add a server if all server space is full. Then the server can be added by the admin and the user requests will be fulfilled.

Green computing concept also implemented in which the sites get deleted if a web site period gets over and resources will gets revealed simultaneously.

The webpage will be hosted successfully if the approval is given by admin. The website status can be viewed by the user as well. Each server status can be viewed by the admin that is the load on each server. The threshold value of each server can also be set by the admin.

Based on the server’s threshold value coldspot and hotspot concepts used in order to achieve the goal of dynamic resource allocation.

The various screenshots of this system are shown below:



Fig1. User Home page



Fig 2. User Home showing the domain details added by user



Fig 3. Admin Home showing Server details added by Admin



Fig 4. Admin Home showing the threshold value updating

IV. CONCLUSION

The design and implementation of a resource management system for cloud computing services, I have presented in this paper. Based on the changing demand of various web sites this system multiplexes virtual to physical resources adaptively. Both overload avoidance and green computing for systems with multi resource constraints are achieved in this system. As future work about this system is to add more security with our existing work and also to integrate the mobile environment where it requires optimum resource utilization.

REFERENCES

[1] Davide Tammaro and Elias A. Doumith “ Dynamic Resource Allocation 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. Fraster, 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] 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).