

Adaptive Configuration Workload Queue in Cloud Computing

Rupinder Kaur¹, Sahil Vashist²

¹Student, Department of Computer Science, Chandigarh Engineering College, Landran, Mohali

²Assistant Professor, Department of Computer Science, Chandigarh Engineering College, Landran, Mohali,

E-mail- , rupinder225@gmail.com

Abstract - With the emergence of cloud computing, on-demand resources usage is made possible. This allows applications to scale out elastically according to the load. In this work we have tried to maximize the cloud profit and efficiency by efficient placement of virtual machines. The target of this algorithm is to maximize the service providers in the case of current resources are not enough to process all the requests in time. In this strategy, the requests are ranked according to the profits they can bring. Simulation results show the efficiency of our framework.

Keywords - Efficiency, migration, response time, Isolation messaging queue, Sharing message queue, CSP, SLA

1. INTRODUCTION

Cloud Computing is the evolving paradigm with changing their definitions but in this research project, it is defined in the term of a virtual infrastructure which can be provides the shared information and services of communication technology, via the internet “Cloud” for access of “external multiple users” through use of the Internet or the “large-scale private networks”. Cloud Computing is providing a computer user access to the Information Technology services i.e., data servers, storage, applications, without requiring understanding of a technology or even the ownership of infrastructure. Cloud computing is associated with a new paradigm for the provision of computing infrastructure. This paradigm shifts the location of this infrastructure to the network to reduce the costs associated with the management of hardware and software resources. Hence, businesses and users become able to access application services from anywhere in the world on demand. Therefore, it represents the long-held dream of envisioning computing as a utility where the economy of scale principles help to drive the cost of computing infrastructure effectively down. Big players such as Amazon, Google, IBM, Microsoft and Sun Microsystems have begun to establish new data centers for hosting Cloud computing applications in various locations around the world to provide redundancy and ensure reliability in case of site failures. The Cloud Computing is a subscription-based service where one can obtain networked storage space and the computer resources. Resources seem infinite, growing and shrinking supported the demand placed on the applying or service being delivered.

Cloud computing is fragmented into three segments which are storage, applications and connectivity. Each segment serves a different function and offers different products for businesses and individuals around the world. A study conducted by V1 in June 2011 [9], found that 91% of senior IT professionals actually don't know exact about cloud computing and two-thirds of senior finance professionals are clear by the concept, highlighting the upcoming nature of the technology.. An Aberdeen Group study found that disciplined companies achieved on average about 68% increase in their IT expense because cloud computing and only a 10% reduction in data center power costs. A simple example of cloud computing is Gmail Hotmail or Yahoo email etc. Just an internet connection is required and you can start sending and retrieving emails. The server and email information management software are available on the cloud and is totally managed by the cloud service provider of Yahoo, Google etc. The consumer gets to use the software alone and access the benefits.



Figure 1: Cloud Computing

Cloud computing provides opportunities for organizations reduce costs and speed time with eliminating the investment in IT capital and operational expenses traditionally related to running a business. Available solutions in the market nowadays including publicly hosted clouds, private internal clouds or a hybrid setting, make sure that the user—whether or not it's associate enterprise, service supplier or developer—can notice the best cloud computing answer to satisfy their desires and quickly gain the benefits that enable them to become additional agile. The success of Amazon, eBay and Google has led to the increase of cloud computing as a replacement, proven design for the way the standard datacenter is constructed and managed.

2. RELATED WORK

There have been several efforts in studying the deployment of scientific and high performance applications on various cloud computing platforms [1], [2]. Our work differs by showing the deployment of a high performance application on multiple cloud computing platforms through a cloud computing framework. We also establish guidelines for the design, implementation, and identification of cloud computing frameworks in this work. There have also been several efforts in building and migrating bio-molecular applications to distributed computing environments [4], [5]. The work in [4] present a framework that provides fault-tolerance and failure recovery for running replica-exchange simulations on distributed systems. This is achieved through check pointing and an external interface that monitors the execution of distributed applications. Work Queue differs by offering these functionalities inherently without overheads. The authors in [5] describe their experiences in running replica-exchange simulation software, NAMD, on the Condor grid. They add a dedicated set of resources to speedup slow replicas executing on the grid and notice improvement in the efficient usage of available resources. The authors go on to present database architecture for storing and retrieving bio-molecular simulation results. The work in [6] describes experiences in using Legion, an operating system that provides abstractions for managing and utilizing grid resources, to run replica-exchange simulations built using MPI. This work provides good insights on the effectiveness of abstractions in providing a seamless transition for users porting applications to run on grids. A work in [6] represents lightweighted, asynchronous high performance messaging queue for the cloud to calculate the capability of message queue with errors. The author in [7] describes the multitasking system under message queue in real time system. In [8] author illustrates the modeling and simulations of cloud computing environment.

3. SYSTEM MODEL

In this Figure 2 we can see that broker has managed to get all the required information of other CSPs. By getting information from other CSPs we mean the free VMs which the broker can lend from the other DC within or outside CSPs so that more optimized results can be generated.

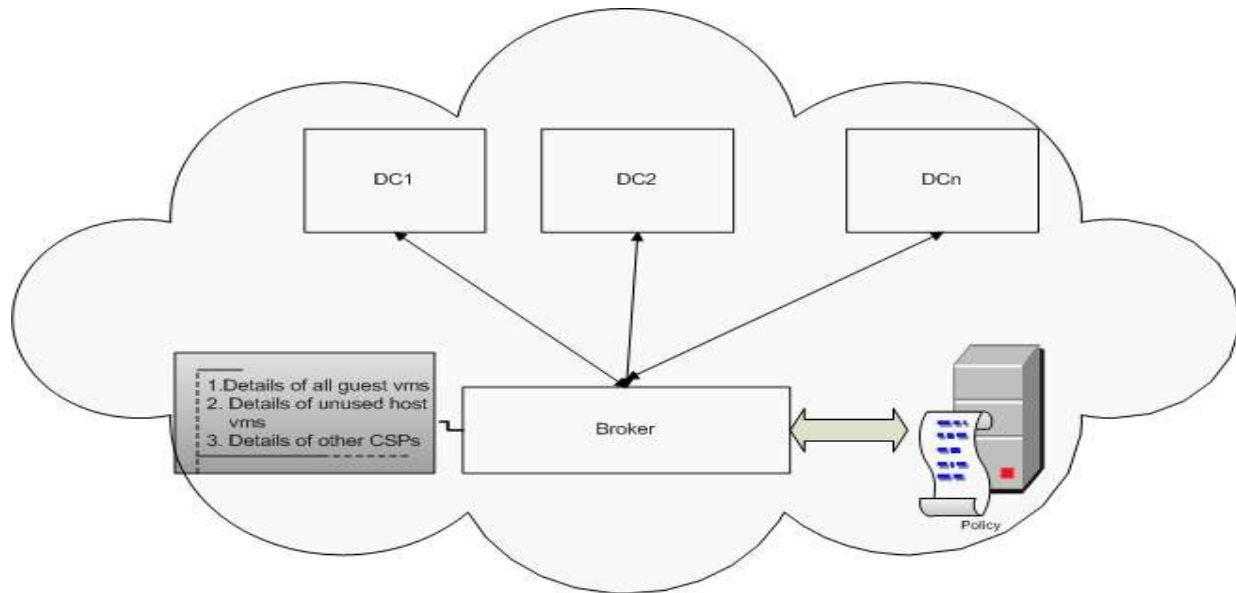


Figure 2: Schematic Model

The VMs details are being offered by the other CSPs on regular basis and each broker is capable to view and compare all the VMs from its and other sources. If we look at the algorithm as shown below broker is responsible for maintaining all the VMs. Each VM is tagged by the health card which checks the capability of VMs for a particular load. Only after this checking the VMs are allocated to the algorithm for efficient VM migration.

4. SIMULATION RESULTS

To setup our experiments, proposed algorithm is implemented is done using CloudSim as a framework in the simulator environment. This work considers various Datacenter, Virtual Machines (VM), host and cloudlet components from CloudSim for execution analysis of the algorithms. To evaluate the results we have set up the mechanism of base paper and then recorded the results, after that we have used the same experimental setup and implemented our framework on same parameters as of base paper. In the base paper the authors implemented same experimental setup for different resource allocation policy and compare it with its own. In the authors algorithm main focus is on to increase revenue of each resource which further accumulate to make a big profit for the CSPs. However in order to perform this work they focused to reduce SLA violations to keep profit increasing. But in our case we focused on the configuration of the queue as shown in Algorithm below:

Algorithm for Adaptive Configuration Workload Message Queue

1. let $N :=$ total number of VM in Cloud Service Provides (CSPs)
2. let $M :=$ total number of VM currently under use
3. let $x := N - M$ (total number of unused/guests VMs)
4. let $page[N] :=$ set of all guest VM pages
5. let $pivot := 0$; $bubble := 0$
6. ActivePush (Guest VM)
7. while $bubble < \max(pivot, N - pivot)$ do

8. get_health_card(vmID);
9. Vm_in_Queue for transmission
10. bubble++
11. PageFault (Guest-page X)
12. discard pending queue

In our approach we track the total number of VMs available in a CSP and put them in a queue. We use two queues; in this case the other queue is used for the reserved virtual machines which may or may not be collected from other CSP. The use of other CSPs VM is totally depending upon the unavailability of the VM. The experimental setup for comparative study of our technique with base paper is as under:-

DCs	10
Hosts	2
VMs/Host	4
Hypervisor	Xen
Requests Frequency	1~1000

The VM is of same size i.e. each VM constitutes of Mips = 1000, RAM size = 512 MB, B/W = 1000 kbps and no. of CPUs cores = 1. In the figure 3 we can see that the number of served requests for VM isolated vs. Adaptive configuration.

In the Figure 3 we have noticed that with the increase in the availability rate the requests served has also being increased. However in our algorithm the virtual machine shortage is overcome by other CSPs and all the available VMs are kept in queue. Therefore the number of requests are processed are 10% faster in our algorithm.

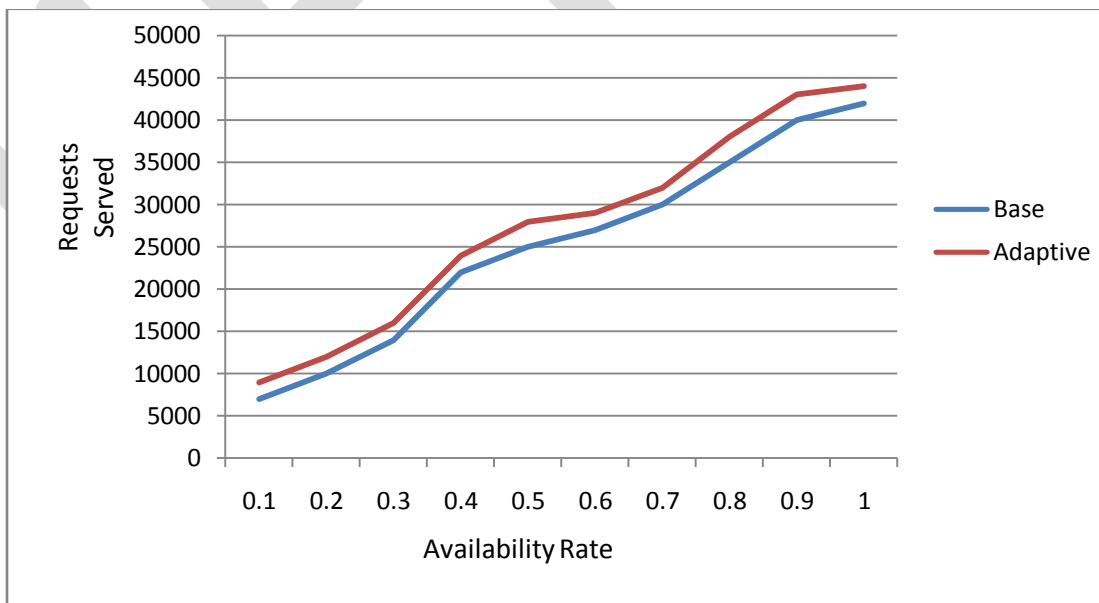


Figure 3: Number of Requests Served

In Figure 4 we can see that Adaptive configuration do not allow service level violation due to reserved VM in the queue. In this case we have supposed the SLA threshold is 50ms and with the increase in the number of requests our algorithm do not let the processing to cross this threshold and get the VM from the reserved queue if the required VM is not available.

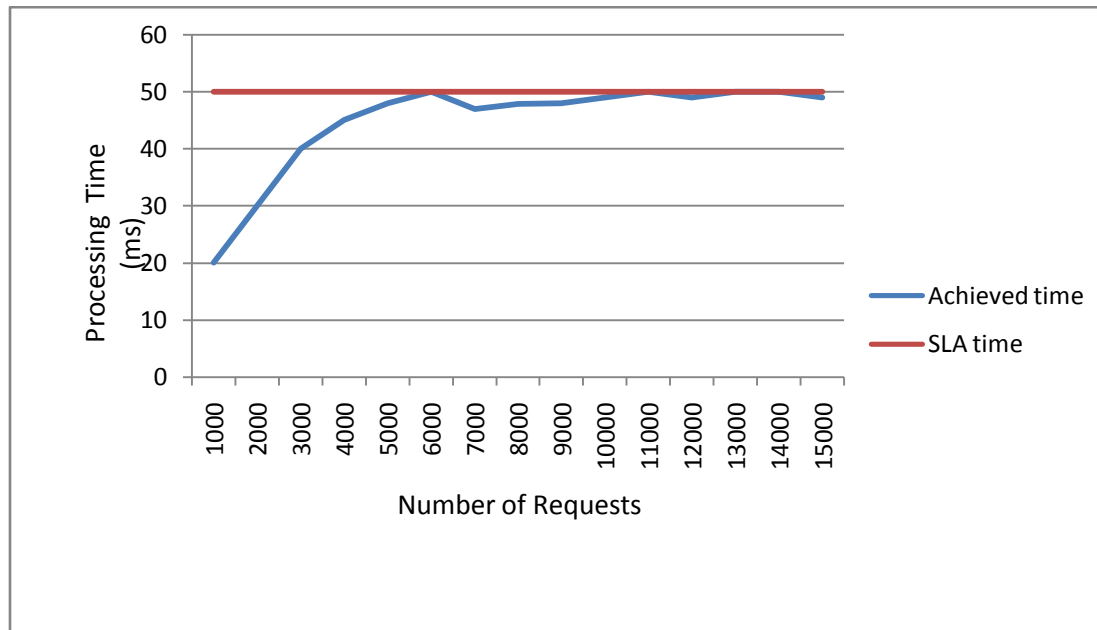


Figure 4: Restrictions for SLA Violations

5. CONCLUSION AND FUTURE SCOPE

In this paper we have evaluated the response time and cost of the proposed frame work and found that the efficiency of the proposed framework is far superior to base paper. Further we can add the parameter of SLA violation and will try to minimize it. The effect of minimization could be very crucial in overall performance increasing and managing virtual machines.

We evaluate the comparison between VM isolated and adaptive configuration message queue. VM isolated method is better way to handle with the massive requirement of resources at a particular time and resources get hire it from other CSPs. Under such condition, when load got raised on the server, sharing can also be the possible way to tackle with this but if sharing effects the efficiency of the system then adaptive message queue should is preferred so that it will work according to the on-demand requirements. Additionally, in previous research hiring CSPs approach is considered as better approach and we continued it by maintaining message queue according to the on-demand desire.

For future scope, as we consider adaptive configuration by hiring CSPs (whenever required) to generate efficient result by which we can enhance massive number of available CSPs so we can handle more number of requests (in case of shortage of CSPs) at a time along quality of service and better efficiency.

References

- [1] K. Keahey, R. Figueiredo, J. Fortes, T. Freeman, and M. Tsugawa, "Science clouds: Early experiences in cloud computing for scientific applications," 2008.
- [2] G. Juve, E. Deelman, K. Vahi, G. Mehta, B. Berriman, B. Berman, and P. Maechling, "Scientific workflow applications on amazon ec2," in *E-Science Workshops, 2009 5th IEEE International Conference on*, December 2009, pp. 59–66.

- [3] A. Luckow and et al., “Distributed replica-exchange simulations on production environments using saga and migol,” in *IEEE Fourth International Conference on eScience, 2008*, December 2008, pp. 253–260.
- [4] C. J. Woods and et al., “Grid computing and biomolecular simulation,” *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, vol. 363, pp. 2017–2035, 2009.
- [5] A. Natrajan, M. Crowley, N. Wilkins-Diehr, M. Humphrey, A. Fox, A. Grimshaw, and I. Brooks, C.L., “Studying protein folding on the grid: Experiences using charmm on npaci resources under legion,” in *Proceedings of the 10th IEEE International Symposium on High Performance Distributed Computing*, 2001, pp. 14–21.
- [6] Joerg Fritsch and et.al. represents “A lightweighted asynchronous high-performance message queue in cloud computing “ in *Journal of cloud computing: advances, system and applications*,2012.
- [7] M. Joseph and P.Pandya, “ Finding Response time of message queue in a real-time system, *BCS Computer Journal*. 29(J):390-395,2002.
- [8] Rodrigo N.Calheiros and et. al., “A novel framework for modeling and simulations of cloud computing infrastructure and services”, published in *Journal of Computing and InformationTechnology*, CIT-16-2008, 4-235-246.
- [9] [http:// asynchronous message queues in cloud computing.org/](http://asynchronousmessagequeuesincloudcomputing.org/) accessed in June,14
- [10] D. Ardagna, M. Trubian, and L. Zhang, “SLA based profit optimization in multi-tier systems”, *Proceedings of the 4th IEEE International Symposium on Network Computing and Applications, Cambridge, Massachusetts, USA*, July 27-29, 2005.
- [11] Mohammed Alhamad, Tharam Dillon and Elizabeth Chang, “Conceptual SLA Framework for Cloud Computing”, *4th IEEE International Conference on Digital Ecosystem and Technologies, 2010*.
- [12] Hyun Jin Moon, Yun Chi and Hakan Hacigumus, “SLA-Aware Profit Optimization in Cloud Services via Resource Scheduling”, *IEEE 6th World Congress on Services, 2010*