International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org

Hybrid Approach for Optimal Provisioning in Cloud

Ruta Ajgaonkar
Assistant Professor
Thakur College of Engineering and Technology

ABSTRACT

Cloud Computing is a paradigm where services are made available through the internet. It is providing services using many technologies. It has gained importance because of its pay-per-use structure which means the user has to pay only for his use. Because of the large number of cloud users, cloud Management has become of prime importance. Cloud Management involves many tasks such as provisioning, deployment, monitoring, metering, billing etc. To provide cloud computing services, provisioning of the cloud resources is a prior step.

Cloud resources include CPU, memory, disk, network bandwidth etc. Provisioning is simply allocation of resources. To provide services to the user or to deploy applications on the cloud, provisioning of resources has to be done. Due to the availability of limited resources in the cloud, a need for optimal use of resources is felt. Optimal allocation of resources, involves conservation of resources, minimizing the cost of resources, satisfying the need of the user etc.

Minimization of cost for renting virtual machines is considered, because cloud has
gained popularity due to its cost effective nature. The cost of renting a virtual machine is calculated by considering the cost of RAM utilized, CPU cores rented, the bandwidth utilized, also considering the time for which the virtual machine is rented. The focus of this project is on minimizing the cost of RAM utilized which in turn minimizes the cost of renting the virtual machine. Optimal allocation of resources can be done by cost minimization, load balancing, cost balancing etc. A Hybrid approach for optimal allocation has been developed to minimize the cost of the virtual machine that is rented by the user.

1. INTRODUCTION

1.1 What is Cloud Computing?

Cloud Computing is simply providing services over the Internet. It is referred to as the pay- per- use model because the user/ application owner pays only for the resources used by him. Provisioning is the process of allocating resources as per the need of the user/application. It is like fully configuring a computer and making it available to the user. For the application owner/user the cloud platform appears to consist of unlimited resources, however this is not the case. On a cloud platform a resource once allocated cannot be allocated again until the user releases the resource, even if the user is not using it. This leads to non-optimal use of resources. Also, to meet the demand of resources at peak times, over provisioning of resources may be done which leads to wastage of resources.

1.2 Issues faced in Provisioning of Resources

Over provisioning of resources is simply providing more number of resources compared to the actual need of the user/ application. Due to the large number of cloud users, another problem faced is the need for on-demand resource provisioning. A major challenge in on-demand resources provisioning is the partitioning of the CPU, memory, disk and network bandwidth among the resident virtual machines, and optimal configuration for virtual machines.

All these issues can be addressed by the optimal provisioning. Optimal provisioning is allocating the resources considering the need of the user/application. In simpler terms, optimal provisioning is providing adequate resources for satisfying the need of the application and the user. This can help reduce wastage of resources. Sufficient resource provisioning can help save money for users (e.g., cloud application developers or owners), which is the key to the success of the pay-per-use nature of cloud. If the requirements of the user are analyzed correctly, the allocation of resources can be done effectively.

1.3 Recent Work

There are many techniques that perform resource provisioning dynamically [1].Now, work is being done to perform optimal provisioning of resources. Optimal provisioning can be done by integrating the resource consumption and allocation management of a cloud application [2]. This technique allocates resources on- demand to improve the performance of an application on the cloud. A Vector Bin packing approach[3] can be used to minimize the number of resources provided to the user. As there are many sites associated with a service centre, it becomes difficult for the user to select the best service. An efficient QoS provisioning scheme is proposed[5], which helps the user to select the most appropriate service from the available service sets according to QoS metrics such as cost, response time etc. A flexible framework [4] is proposed which provides the user with different pricing quotes computed by different scheduling policies. The user presents the job, and the framework provides the user the flexibility to choose between different pricing models as per the need.

There are many approaches to allocate resources effectively.ie. Cost minimization, load balance, balancing cost etc. The cost minimization approach involves allocating the resource that has the minimum cost considering the availability of the resource. In the cost minimization approach, load of the hosts is not considered. The load balancing approach includes balancing the load on the hosts not considering the cost of the RAM, thus in this technique the cost may not be minimum. In cost balancing approach, the amount spent on both the resources has to be equal.

For allocation of resources, all these factors play an important role. A hybrid optimization policy is developed to provide a combined optimization considering all these factors. The user is given a choice to either select cost minimization, load balancing and balancing cost by assigning weights. The user either selects one of the three or the user can select two of the three or the user may give equal weights to all the three.

The rest of the paper is organized as follows. Section 2 gives the literature survey. Section 3 gives the proposed technique. Section 4 discusses the evaluation. Section 5 gives the conclusion.

2. Problem Definition

From the recent techniques, it is observed that the following issues have to be addressed in provisioning:

- Conservation of resources.
- Minimizing user cost.
- Satisfying the need of the user.
- User has to compute the number of resources to be rented.

Computing the number of resources to be rented imposes an unnecessary burden on the user. In case the user cannot correctly compute the number of resources required, the user may rent excessive or insufficient number of resources. If the user rents excessive resources and does not use them, the resource cannot be allotted to any other user until it is released. So, the user has to pay for the resource even if it is not used.

The focus is on minimization of the user cost. The user has to rent virtual machines as per requirement of the application. The cost of the user should be as minimum possible. The focus is on reducing the cost of the RAM used which in turn reduces the cost of the virtual machine. The cost of the time for which the virtual machine is rented is not considered.

2.1 Why optimization in Cloud?

On a cloud platform, provisioning of virtual machines is done at two levels 1) allocation of virtual machine to a host, 2) allocation of a virtual machine to an application. A virtual machine is allocated to a host only if the requirements of the virtual machines (for ex. RAM, bandwidth, processing power etc.) are satisfied considering the constraints applicable. Hence, the allocation of virtual machines at both the levels should be optimal thereby leading to conservation of resources and also reducing the cost associated with the virtual machines.

Also, on a cloud platform the bill of the user is calculated according to the usage of the user. The bill of the user is displayed in the account of the user. So, the user has to calculate the amount that is to be spent on the resources that he hires, so that he hires resources in an affordable range. The user has to keep track of the billing amount and has to use resources accordingly.

For the above reasons the need for optimal allocation of resources is essential. Allocating resources optimally helps the user save money and actually pay for the resources that he uses as per the need of the application. Also, this leads to the best utilization of the limited available resources on the cloud.

3. Proposed Technique

3.1 Motivation for the Proposed Solution

There are many approaches to allocate resources effectively.ie. Cost minimization, load balance, balancing cost etc. The cost minimization approach involves allocating the resource that has the minimum cost considering the availability of the resource. In the cost minimization approach, load of the hosts is not considered. The load balancing approach includes balancing the load on the hosts not considering the cost of the RAM, thus in this technique the cost may not be minimum. In cost balancing approach, the amount spent on both the resources has to be equal.

Cost minimization approach focuses on minimizing cost; the load balancing approach focuses on balancing the load on the cloud, and does not provide the minimum cost and the cost balancing approach only deals with balancing the cost among the resources and not with the balancing of load. However, a need for a combination of policies is felt. The user may want the minimum cost but not compromising on the performance. Also the user may want to rent a machine considering all these factors. The priority of the approaches is based on the requirement of the user. A provision has to be made for the user to decide the approach to be taken. These factors led to the need for the development of the Hybrid approach.

3.2 Hybrid Optimization approach for Cloud

In the Hybrid Optimization approach for cloud (i.e. HOC), the concepts of Cost minimization, load balancing, costs balancing are combined together. The priority of the approaches is decided by the user. The user may give equal priority to all the approaches, or may give first priority to one of the approaches and give zero priority to the others. A model of the hybrid approach is shown below (figure [3]). The cost of the virtual machine is calculated considering the cost of the no of CPU cores utilized, the amount of RAM utilized, the bandwidth utilized etc. In this project the focus is on minimizing the cost of the RAM utilized which in turn reduces the total cost of the virtual machine. The cost of the Ram depends on the type of RAM, the amount of RAM, speed of the RAM. More the speed of the RAM more is the cost of the RAM.

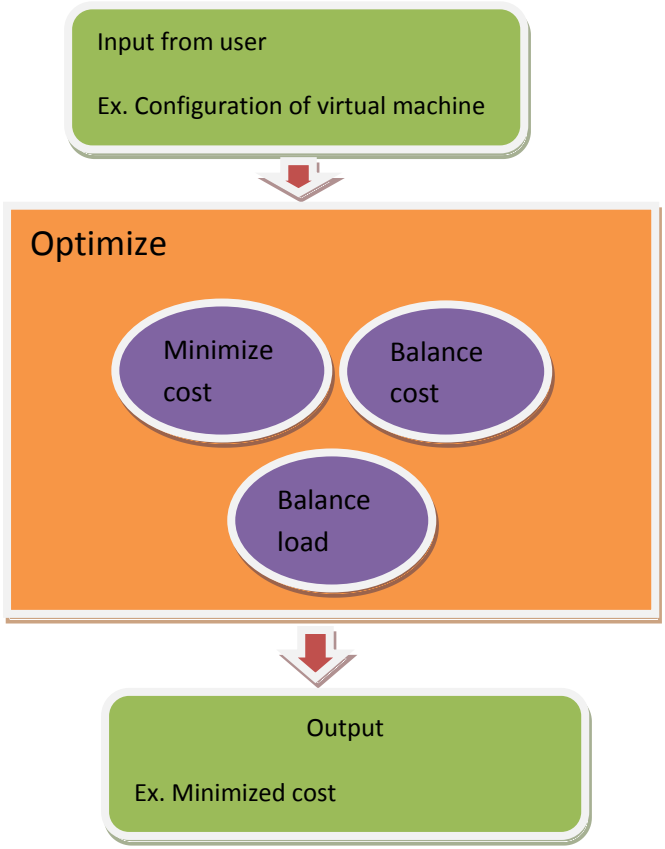

Figure 1: Model of the Hybrid Optimization approach for Cloud

The user gives as input the configuration of the virtual machine to be rented, the requested RAM is attained from the user input. The user is given the flexibility to assign weights to the three approaches .i.e Cost minimization, load balancing, Cost balancing. According to the weights assigned by the user, the priority of the approaches is decided. If the user gives weight as 1 for cost minimization and 0 for the other two approaches then the other two approaches are not considered and the focus is only on Cost minimization. Similarly, if the user gives priority to load balancing and has no issues with the cost of RAM, then the load balancing approach is only given importance. The requested RAM is split among the hosts to reduce the cost and improve the performance.

In the Cost minimization approach [8], if the cost of the RAM of host1 is less than the cost of RAM of host2, then the RAM of host1 is provided to the user until the requested RAM is less than the RAM available with the host1. If the requested RAM is greater than the RAM available with the host1, then the remaining amount of RAM is provided from the RAM available with the host2. The same procedure is repeated in case the cost of RAM of host2 is greater than the cost of RAM of host1.

In the load balancing approach [8], RAM of both the hosts is provided. A calculated amount of RAM from both the hosts is provided to the users. In load balancing approach, the most optimal solution can be attained when the load is equal on both the hosts. The amount of RAM to be provided is calculated by equating the load on both the machines.

In cost balancing approach [8], an equal amount is spent on the RAM from both the hosts. Thus, the minimum cost is determined by equating the amount spent on both the hosts.

3.3 Steps of the HOC

a) The user gives the configuration of the virtual machine as input.
b) The user assigns weights to the three approaches i.e. Cost minimization, load balancing, Cost balancing as per requirement.
c) The RAM is split into RAM1 and RAM2.
d) RAM1 and RAM2 are calculated by cost minimization approach, followed by load balancing and then by cost balancing.
e) Then a weighted average is taken of the RAM calculated in all three approaches considering the weights assigned by the user.
f) Then the cost of the requested RAM is computed and provided to the user.

3.4 Formulations

a) Cost minimization policy

If c1 > c2

Where c1= cost of RAM on host1, c2 = cost of RAM on host2

Then allocate all ram from c2 until RAM is available on host. i.e.RAM2 = requested RAM and RAM1=0

If RAM2 > RAM available on host2

Where

RAM2 = RAM to be allocated from host2

RAM1 = RAM to be allocated from host1

Then RAM2 = RAM available on host2

And

RAM1 = requested RAM – RAM available on host2.

If the requested RAM is greater than the availability, then the remaining amount of RAM has to be allocated from host1.

If c2 > c1

then reverse of the above procedure.

b) Load balancing approach

If RAM on host 1 = R1

And RAM on host 2 = R2

RAM1 / R1 = RAM2/R2

Where

RAM1 = RAM to be allocated from host1

RAM2 = RAM to be allocated from host2

R = requested RAM

i.e. RAM1 = RAM2*R1/R2

RAM1+RAM2 = R

RAM1 = R - RAM2

i.e. R - RAM2 = RAM2*R1/R2

RAM2 (1+R1/R2) = R

RAM2 = R / (1+R1/R2)

RAM1 = R - RAM2

c) Cost balancing approach

If c1= cost of RAM on host1

c2= cost of RAM on host2

It is evident that cost will be minimum when

RAM1*c1 = RAM2*c2

Where

RAM1 = RAM to be allocated from host1

RAM2 = RAM to be allocated from host2

R = requested RAM

RAM1+RAM2 = R

RAM1 = R-RAM2

R - RAM2 = RAM2*c2/c1

RAM2 * (c2/c1+1) = R

RAM2 = R / (c2/c1+1)

RAM1 = R – RAM2

3.5 Assumptions

a) The cost of RAM on host1 is not equal to the cost of RAM on host2.

b) The datacenter is considered as the service provider.

c) The total cost of renting a virtual machine is calculated by keeping into consideration the cost of the RAM and the CPU cores.

4. Experiments and Simulations

Simulator used: CloudSim-2.1 version
CloudSim[6] is a simulation toolkit that helps modelling the Cloud computing systems and environments. The toolkit supports system and behaviour modeling of Cloud system components (for example, data centers, virtual machines (VMs) and resource provisioning policies).

The experimental evaluation is done by considering 9 datasets. In each dataset the values for RAM on host1, RAM on host2, cost of RAM1, cost of RAM2 are varied over an assumed range. Also, the results for all the three approaches are computed and graphs are plotted for the same.

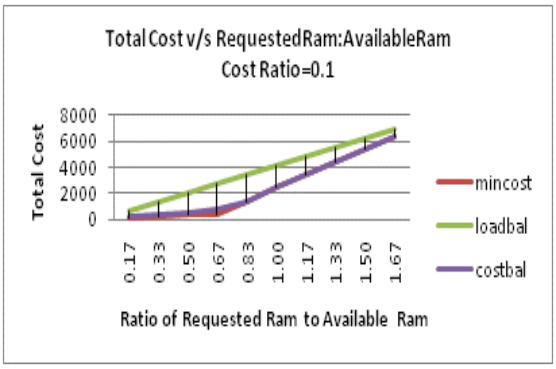

Figure 2: Graph of seventh dataset

When cost ratio is high and the RAM ratio is low, the three approaches (i.e. Cost Minimization, Load balancing, and Cost balancing) tend to give almost same results. However as the cost ratio decreases and the RAM ratio increases, the results of the three approaches (i.e. Cost Minimization, Load balancing, and Cost balancing) tend to differ and in such a case HOC becomes relevant. The user assigns weights to the approaches mentioned above as per requirement. If the user assigns different weights (i.e. 1 or more than 1) to the approaches, the cost may be more than minimum cost.

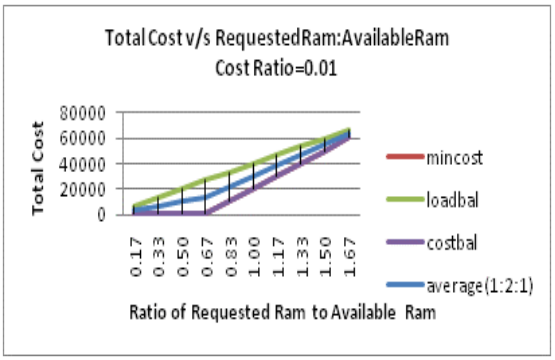

Figure 3: Graph of ninth dataset

In the eighth and the ninth dataset the cost for HOC was computed.

5. Conclusion and Observations

There are many approaches to allocate resources effectively.ie. Cost minimization, load balance, balancing cost etc. Cost minimization approach focuses on minimizing cost, the load balancing approach focuses on balancing the load on the cloud, and does not provide the minimum cost and the cost balancing approach only deals with balancing the cost among the resources and not with the balancing of load. The user may want the minimum cost but not compromising on the performance. Also the user may want to rent a machine considering all these factors. Hence, there is a need for a combination of all the approaches. This led to the development of the Hybrid approach for optimal allocation of resources.

The following conclusion was made after the experimental and theoretical evaluation:

- Observation:

1) When the ratio of cost of RAM1 (RAM provided from host1) to the RAM2 (RAM provided from host2) is high and the RAM ratio (ratio of RAM1 to RAM2) is 1, the results of the three approaches (i.e. Cost Minimization, Load balancing, Cost balancing) tend to overlap.

2) When the cost ratio increases, all the three schemes tend to give different results, and in this case (i.e. Cost Minimization, Load balancing, Cost balancing) HOC becomes relevant.
3) When the ratio of the requested RAM to the available RAM becomes very high (more than or equal to 1.67) or very small (or less than or equal to 0.17), results of the three schemes (i.e. Cost Minimization, Load balancing, Cost balancing) again tend to give more or less the same results, and hence HOC is not that important.
4) The cost minimization and cost balancing approach give more or less the same results, when the RAM ratio increases and the cost ratio decreases.

- Conclusion:
  When cost ratio is high and the RAM ratio is low, the three approaches (i.e. Cost Minimization, Load balancing, and Cost balancing) tend to give almost same results. However as the cost ratio decreases and the RAM ratio increases, the results of the three approaches (i.e. Cost Minimization, Load balancing, and Cost balancing) tend to differ HOC becomes relevant. The user assigns weights to the approaches mentioned above as per requirement. If the user assigns different weights (i.e. 1 or more than 1) to the approaches, the cost may be more than minimum cost.

- Future Work
  - Monitoring of the execution time of the technique.
  - Comparison of the execution time with the execution time of CloudSim-2.1
  - Calculation of the number of virtual machines on behalf of the user.

REFERENCES

[1] Qi Zhang ,LudmilaCherkasova"A Regression-Based Analytic Model for Dynamic Resource Provisioning of Multi-Tier Applications" in Fourth International Conference on Autonomic Computing ,Jacksonville, FL,June 2007.

[2] Ying Zhang, Gang Huang∗ , Xuanzhe Liu, and Hong Mei "Integrating Resource Consumption and Allocation for Infrastructure Resources on-Demand" IEEE 3rd International Conference on Cloud Computing

[3] J. Shahabuddin, A. Chrungoo, V. Gupta, S. Juneja, S. Kapoor, and A. Kumar, "Stream-Packing: Resource Allocation in Web Server Farms with a QoS Guarantee," Lecture Notes in Computer Science, 2001.

[4] Henzinger, T.A. Singh, A.V. Singh, V. Wies, T. Zufferey, D. "FlexPRICE : Flexible Provisioning of Resources in a Cloud Environment" IEEE 3rd International Conference on Cloud Computing, Miami, FL ,July 2010.

[5]Yanping Xiao, Chuang Lin, Yixin Jiang "Reputation-based QoS Provisioning in Cloud Computing via Dirichlet Multinomial Model" inIEEE International Conference onCommunications Cape Town, May 2010.

[6] R. Buyya, R. Ranjan, and R. N. Calheiros."Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities" Proceedings of the Conference on High Performance Computing and Simulation (HPCS 2009), IEEE Press, New York, USA, Leipzig, Germany, June 21 - 24, 2009.

[7]Book on "Operation Research – AnIntroduction " eighth edition, by author Hamdy. A. Taha

[8] Book on "NETWORK ROUTING ALOGRITHMS,PROTOCOLS,ARCHITECTURES", by DeepankarMedhi, KarthikeyanRamasamy