THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Efficient Management of
Resource Provisioning Cost in Cloud Computing

Oasi A. Joseph
PG Student, M.Tech, Department of CSE, JCET, Lakkidi, Palakkad, Kerala, India
Sarath V. Sankaran
Assistant Professor, Department of CSE, JCET, Lakkidi, Palakkad, Kerala, India

Abstract:
Cloud computing is a large-scale distributed computing paradigm in which a pool of computing resources is available to users via the internet. Cloud computing allow to use remote servers to run your applications. Cloud computing provides various services to users and organizations. Basically the services are divided into three categories-SaaS, PaaS, IaaS. One has to login to the portal and submit the request. The services are given by the cloud provider and the provider validates the request. After validation it will provision the request. The cloud user then wants to pay for using cloud resources. The user can buy the resources either by on-demand or by reservation. Reservation plan is cheaper than on-demand plan. But due to the over provisioning and under provisioning problems, the pricing will be fluctuating. In order to obtain optimal resource provisioning cost the optimal resource provisioning algorithm is used based on the time parameter. i.e., the cloud user only wants to pay for the time spent on utilizing the cloud. Thus optimal cloud resource provisioning algorithm reduces the cloud utilization cost. Since the benefits of cloud computing many users using the cloud, it might create heavy network traffic. The proposed method will manages the network traffic effectively and reduce the delay in job completion as well as the cost for cloud services. Hence the resource provisioning can be made fast and efficient resource provisioning cost can be obtained.

Key words: Cloud computing, resource provisioning, OCRP, latency reduction, low processing time servers

1. Introduction

People are living in a tech-fast world. IE; day by day the technology is developing which makes people do things much faster and easier. Years before man lived in a world where there were no ways to communicate between long distances people and now the technologies advanced so far that distant communication became a nearby chat, it's all because of the invention of the Internet. The internet is a global network that provides various services like the World Wide Web, communication, data transfer etc. A communication service available on the internet is email. In olden days, the messages, documents, files etc that used in email were stored in-house. Later in 20th century companies like Google or Yahoo provide their servers to users, to store email information for free. Users only want to use their application like Gmail, Yahoo mail to access the data. Practically this is called cloud computing- using other's servers to run your application.

Cloud computing is a technology that uses internet to store, manage and process the data on remote servers or can say that cloud computing allow to use other's servers to run one's application. For example: Gmail-is a service provided by the Google. A user requires only an internet connection for using this service. After log in one get the data in his account, the software and storage for his account is available on remote servers, not on his system. User doesn't want to pay for using the Gmail service. Likewise there are many free services, also paid services provided by the cloud. Cloud computing provides various services to users and organizations. Basically there are three types of service models[1] called Software as Service (SaaS) which allows users to access the application software, Platform as a Service (PaaS) in which providers provide the platform for the developers to develop and run software applications, and Infrastructure as a Service (IaaS) providers offer virtual machines and other resources. These cloud services are provided by the various cloud providers like Google, Amazon, Azure, Microsoft etc. The pricing for the services provided by these providers may be different. The users want to pay according to their usage of cloud services.
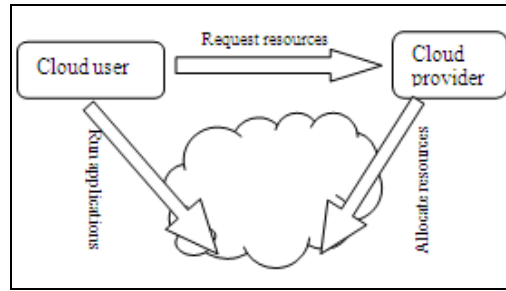
Figure 1: working of cloud computing

The main advantage of cloud computing is its cost efficiency and accessibility of data from anywhere in the world. If particular software is not in pc then it can be used from the cloud without actually installing, hence can reduce the cost for installing the software. When companies started to take the advantages of cloud computing it leads to the economical development of the company. Companies pay a subscription fee for the use of the software and users also save money for software updates, data storage cost. Cloud computing provides large storage than personal computer. When company's implements new system it is profitable to do using cloud computing since it saves the hardware-software cost, maintenance cost, storage cost etc. The other important advantages of cloud computing are its speed, sharing of resources, enterprises are allow to focus on innovation, location independence, optimal resource utilization and so on.

Large number of users and companies started to use cloud computing due to its benefits it offers. In order to request for resources like operating system, software or applications one has to login to the portal and submit the request. The services are given by the cloud provider and the provider validates the request. After validation it will provision the request. The resources are allocated to cloud consumers by two provisioning plans: Reservation plan and On-demand plan. On-demand plan is based on pay-per-use and reservation plan based on one-time (e.g.: 1 year) payment. But there is some problem with reservation plan - under provisioning and over provisioning problems. Under provisioning problem means the reserved resources are unable to meet the demand and over provisioning problem means the reserved resources are more than the demand. The under provisioning problem can be solved by provisioning more resources by the on-demand plan. An optimized cost of on-demand cost and over- subscribed cost of under provisioning and over provisioning problem can be obtained by optimal cloud resource provisioning algorithm.

2. Related Works

Cloud provider has to provision the resources requested by the cloud provider. Various resource provisioning methods were discussed in [2]. An optimization framework for resource provisioning was developed in [3]. Resource provisioning is done based on some allocation polices. In [4] various provisioning and allocation policies were discussed. When request for resources, the cloud provider creates virtual machines. In [5] assignment of VMs to physical servers was discussed. Resource provisioning can be done by dynamic provisioning or advance provisioning. Reservation plan is cheaper than on-demand provisioning. In reservation plan user meet the over provisioning and under provisioning problems. User needs to buy more resource on-demand. That makes the user to spent more money. Due to this variation in demand and price, an optimal pricing scheme cannot be obtained. Optimal solution is obtained by optimal cloud resource provisioning algorithm. In [6] discussed about the algorithm based on various methods like stochastic programming model, bender's decomposition model. Motivated by this model, we introduce optimal solution based on time parameter. Solution for the problem related to multiple provisioning stages is proposed.

3. Optimal Cloud Resource Provisioning

Generally in cloud computing a user logs in to the portal and submits the service request. Cloud provider will provide the services to the users (called as cloud consumers). These services are developed by some other users of cloud (called as developers) and deploy their application to the cloud, so that other users can use those applications. In IaaS model virtualization [7] technology is used to provide the resources to the users (i.e., for cloud consumers). The resource has to be provisioned from cloud providers for executing the jobs requested by the cloud consumers. For that, first create virtual machine that includes the software to complete the jobs.
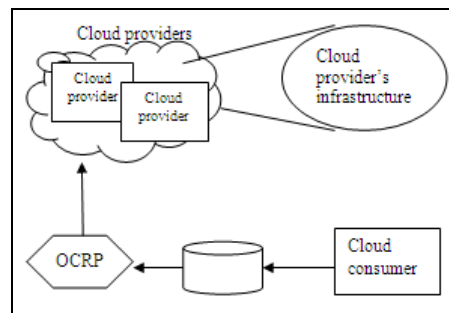

Figure 2: System Model

These VMs can be hosted on cloud provider's infrastructures whose resource can be utilized by the VMs. The cloud broker acts as an intermediary between the cloud consumers and the cloud providers. Cloud brokers help the consumers to get the required services from the appropriate cloud provider. The broker implements the OCRP algorithm to obtain an optimal solution for resource provisioning.

3.1. Resource Provisioning
The cloud providers provision the resources to the cloud users on demand. This can be done by advance provisioning or dynamic provisioning. In advanced provisioning method, the provider prepares the appropriate resources in advance before the start of service. In this method cloud user pays the bill contract (monthly or yearly) wise. In dynamic provisioning the pay-per-use method is used. Resource provisioning occurs in various time intervals called provisioning phases. The provisioning phases include reservation phase, expending phase, and on-demand phase. In the reservation phase the resources can be reserved in advance to execute the task. The expending phase contains the details about the cost of resources utilized by the consumer. Sometimes the reserved resources may not fulfil the user demand, then user need to buy more resource. Then the consumer request more from cloud provider and it leads more cost.
The aim of OCRP algorithm is to reduce the total resource provisioning cost and the optimization is performed based on the time parameter. The cost reduction is done based on the following equation:

$$Z = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} \, x_{ijk} + \Omega \quad (1)$$

where I denote the set of virtual machines, J is the set of cloud providers; K is the set of reservation contracts. In (1) the variable represents the reservation cost for provisioning all resource types and the variable c_{ijk} denotes the number of virtual machines provisioned. Ω denote the expected cost under uncertainty and is given by

$$\Omega = \prod_{t \in T} \Omega t$$

where   represent the time parameter. The optimal solution can be obtained by calculating the cost under the usage of time.

3.2. Reducing the Cost Based on Time
There are various cloud providers who provides various types of services and each service provider assign cost for each resource type per hour or half-an-hour. The cloud users have to pay for one hour (or half-an-hour) even if they are using the cloud for five minutes. This unwanted wastage of cost can be reduced by paying only for the time spent on utilizing the cloud. The step for reducing the cost is explained here. First, the cloud provider reserves the resources on cloud. When user request for resources the cloud provider should allocate the resources to the user. While registering on cloud, user can specify the cloud size required. When the user starts to using the cloud resources, calculate the time the user spent on the cloud. If extra resources are needed it should be made avail through on-demand plan. Then calculate the time when the user utilizes the resources. Time is calculated only during the utilization of cloud services and the user only wants to pay for the time usage. The cloud size will be reduced according to the usage and when the cloud size gets near to very low value, the cloud provider informs the user. Then the user demand for the resources and pay only for the time the user spent on utilizing the cloud resources. Thus observed that OCRP algorithm reduces the total cost when the cost is calculated based on time parameter.

4. Multiple Provisioning
The resource provisioning cost can be minimized by OCRP algorithm based on time parameter. When multiple clients request for the resources there is chance for latency in job completion. This delay will make the cloud user to spend more time on cloud i.e., the cost will be increasing. So this delay can be reduced by the servers whose processing time is low. Lot of jobs given by the users will be in queue until it gets the required resources to complete the task. Primary priority is given to the servers which got low processing time (time required to complete a job). When the request comes, they will be assigned to those servers according to the priority. Thus the latency can be reduced. The priority server is calculated based on the following parameters:
W= the time user wait before getting the requested resources allocated.
I= after allocating the VMs to cloud provider, the cloud consumers has to wait.
C= cost of requested resources that user has to pay.
S= profit that a provider gets when allocating resources to cloud users.
By optimization, job can be allocated to the high priority server. The priority is calculated using the following equation
P=W+I+C+S
After calculating this, best server can be selected and job is allocated to that node. Thus efficient resource provisioning cost can be calculated.

5. Results and Discussion
By using the OCRP with time parameter the cost can be reduced compared to the cost of services offered by the providers. Due to the benefits of cloud many users will be using the cloud leads to traffic delay. This will create latency in job completion. In order to deal this problem the server with low working rate is set to primary server and the job is completed by that server.

Consider two sub servers with its working rate. Here one sub server is considered as Amazon and other sub server as Azure. If working rate for Amazon is low, its priority is assigned as one. The webpage is updated automatically and so the priority will be changing in seconds from Amazon to Azure and vice versa. When job request is send to the main server, the main server checks the priority and the job is assigned to that sub server whose working rate is low compared to the other sub server.
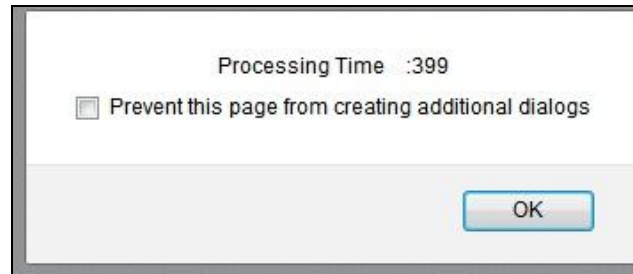

Figure 3: Processing time

Fig 3 shows the processing time of a sub server (it can be Amazon or Azure) to complete a particular job in milliseconds.
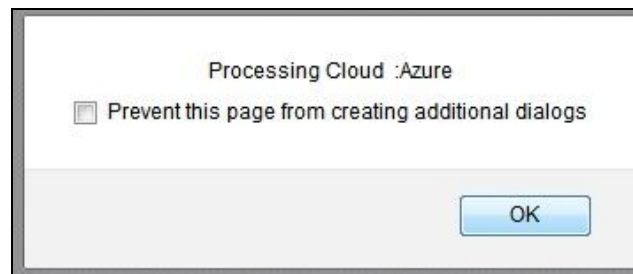

Figure 4: Processing cloud

Fig 4 shows to which subserver the job is allocated. Here show Azure.i.e., when the job request is submitted, at that time the server Azure got low working rate and the resources are allocated by this subserver. Thus multiple resource provisioning can be done easily. If the reserved resources are not sufficient cloud user purchase it through on-demand. In both cases optimal solution can be obtained by the proposed method.

6. Conclusion
In this paper, we are discussing mainly about the reduction of cloud usage cost. By using the OCRP algorithm an optimal solution can be obtained by paying only for the time spent on utilizing the cloud. When heavy network traffic occurs, time spent on cloud increases leads to more amount of money. The proposed method can efficiently manage the traffic and provides the resource provision with optimized cost.

7. References
1. Wikipedia:Cloudcomputing, http://en.wikipedia.org/wiki/Cloud_computing
2. G.Juve and E.Deelman, "Resource provisioning options for large-scale scientific workflows," IEEE foruth Int'l conference, 2008.
3. D.Kusic and N.Kandasamy, "Risk-aware limited look ahead control for dynamic resource provisioning in Enterprise Computing Systems, "Proc.IEEE Int'l conf. Autonomic computing,2006.
4. David Villegas, Athanasios, Seyed, "An Analysis of provisioning and allocation policies for IaaS clouds: ExtendedResults" http://www.pds.ewi.tudelft.nl/, 2011.
5. Ishtiaq Ali, Natarajan M, "Virtual Machines and Networks- Installation, Performance, Study, Advantages and virtualization options", IJNSA, vol.3, no.1, Jan 2011.
6. Sivadon Chaisiri, Bu-Sung lee, Dusit Niyato, "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE transactions on service computing, vol.5, no.2, April-June 2012.
7. Wikipedia:virtualization, http://en.wikipedia.org/wiki/Virtualization