Day-ahead Pricing Model for Smart Cloud

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Software Engineering

Submitted By

Chetan

(801331006)

Under the supervision of:

Dr. Inderveer Chana

Associate Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

July 2015

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*Day-ahead Pricing Model for Smart Cloud*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Inderveer Chana* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Chetan

(Chetan)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

15/07/15

(Dr. Inderveer Chana)

Associate Professor

Computer Science and Engineering Department

Countersigned by

(Dr. Deepak Garg)

Head

Computer Science and Engineering Department

Thapar University

Patiala

(Dr. S. S. Bhatia)

Dean (Academic Affairs)

Thapar University

Patiala

Acknowledgement

During the course of this thesis I have been lucky to be blessed by a lot of kind people. Let me take a moment to thank each one of them.

To begin with, I am thankful to God for his blessings and for consistently showing me the right direction.

I wish to express my sincerest gratitude towards the guidance and help that I have received from my supervisor Dr. Inderveer Chana. I shall ever remain indebted for her consistent support and encouragement. She provided me help whenever needed, and also arranged for me the resources required to complete this thesis report on time.

I am also thankful to Dr. Deepak Garg, Head, Computer Science and Engineering Department, Thapar University for his help and cooperation. Along with that I express my gratitude to all the staff and faculty members of Computer Science and Engineering Department, Thapar University for providing me with the facilities required for this thesis.

I would like to acknowledge Mr. Aashish Jainani and Mr. Amir Mussanna for improving the overall language of the thesis.

I would like to thank all my friends especially Arindam, Megha, Mini, Suman, Sachin, Vishakha, Anita for their support. Also I want to express my appreciation to every person who contributed with either inspirational or actual work to this thesis.

Last but not the least I am very grateful to all my family members for their inspiration and moral support which kept me motivated to purse my studies.

Chetan

Chetan

Abstract

Cloud Computing is a large scale parallel and distributed computing architecture. Services like virtualized machines, computing power, storage, and software etc. are offered through Cloud Computing. On-demand/requested services are generally on the basis of ‘Pay-as-you-go’ model and Cloud service providers charge their consumers for the services they use. Consumers should have a guarantee that the services they are paying for should be delivered to them uninterrupted and this is ensured through Service Level Agreements (SLAs) between the consumers and providers.

Masses around the globe associated with information technology are now realizing the importance of Cloud Computing. Major problems associated with internet can be solved through Cloud Computing. Efficient Pricing in Cloud Computing is an emerging issue in this field. Smart Cloud is the solution for this problem, Smart Cloud provides a platform to its peers which are consumers and providers. Smart Cloud uses a Time Dependent Pricing (TDP) model which calculates the price of resources on the basis of its previous consumption. TDP helps to provide a pricing technique which balances the requirements among consumers and providers. This technique satisfies the needs of both consumers and providers.

The focus of this research work is to provide an efficient pricing technique along with and a scheduling policy which helps to satisfy the service provider and resource consumer. This thesis presents a Smart Cloud which constitutes a framework that can distribute Cloud resources over a communication network. In this model a Cloud Workload Management System (CWMS) has been presented which is an interface between the service consumer and the service provider. It clusters the Cloud resources using k-mean clustering algorithm. A Time Dependent Pricing (TDP) model is used to calculate the price of resources. The Cloud resource providers send resource pricing information from their records (database) to Compromised Cost-Time Based (CCTB) scheduling policy located at the CWMS. CCTB can observe and manage consumers’ resource requirements and schedules it at peak and off-peak periods. Resources can be scheduled automatically or manually by CCTB depending upon the pricing at various hours of the day thus ensuring minimum SLA violations.

Table of Content

List of Figures

List of Tables

Chapter 1
Introduction

The usage of internet has increased exponentially in the recent decades. The services offered by the resource providers over the internet are the next big leap in the communications field. One of the facilities of the internet is to provide remote execution of the task on faraway distributed machines with the help of the internet, known as Cloud Computing. This chapter focuses on the brief introduction of Cloud Computing by examining different service models, deployment models, open challenges and application areas. Motivation and the structure of the thesis are discussed at the end of this chapter.

1.1 Cloud Computing

Everyone in this universe needs a substantial alteration of their working area. As the time passes quickly, within seven decades from the making of the very first digital computer, ENIAC [1], people are breathing in the age of the Internet, and one cannot envisage one's life without Internet. As the network bandwidth is growing, the internet is becoming an essential utility for day to day work. The internet provides its consumers, many facilities (services) such as Emails, Banking Solutions, Social Networking, Stock exchange and E-Business etc. By the existence of this emerging technology different researcher see Cloud Computing as one of the new areas of utility computing environment. Cloud Computing can trace its roots from Grid and Cluster Computing. All these can be a promising paradigm to provide IT as a service to end user. But the task of these service providers is not just to provide some attributes or capabilities towards the consumers, but also provide these services in an efficient manner because this will be a business of more than a trillion dollar business in the coming years. The main motive behind these computing paradigms is to provide computing utilities to its consumers as a guaranteed Quality of Service (QoS). The way of computing also changes with the changing technology, which leads to the development of service-oriented computing.

To understand Cloud Computing there is a need to compare it with widely used computing terminology such as Cluster Computing and Grid Computing. A large number of researchers define Clusters and Grid Computing in different ways. According to Buyya et al. [2], Grid Computing is also a large-scale parallel and distributed computing architecture in which consumers receive resources from geographically service providers and allow to interact with them. As per Buyya et al. [2], Cloud Computing is a large-scale architecture in which virtualized machines, multi-tenancy, computing power, memory (storage), infrastructure for computing, software and platforms are delivered on demand to external customers over an internet connected medium in a business oriented environment. According to Pfister [3], a Cluster is a distributed and parallel system, inside which several computers work as a stand-alone single computing structure. From the definitions, it is clear that Cloud Computing overlaps several features of these computing technologies so by this it is clear that Cloud Computing can trace its roots from Grid and Cluster Computing which are well-known utility computing paradigms. Grid Computing traces its root from Electric Grids where they provide electricity as a metered service, by the implementation of the computing Grids consumers are charged for the services they use as a Pay-as-you-go model. The development of Supercomputers, Datacenters, Web 2.0, distributed system, Grids and Cloud Computing [4], can be seen with the help of Figure 1.1 how they scale up. These utility services come under in the distributed system, in this system computing machine interacts with other machines, communicate and coordinate with each other via a network. Each machine knows its work and performs the tasks given to it, even if a machine needs some resources to compute that work, then that machine can demand for resources over the network. Grid, Cluster and Cloud Computing are interconnected with each other in many ways because Cloud is the terminology derived from the clusters and grid, and all come into the category of distributed systems. Table 1.1 shows various characteristics of Grid, Cluster and Cloud Computing.

Cloud Computing offers services to its consumers over the internet with the help of infrastructure in their data centers. Cloud Computing is a large scale parallel and distributed computing architecture that offers a variety of services to its consumers such as virtualized machines, computing power, storage, and software etc. Cloud Computing offers on-demand/requested services generally on the basis of ‘Pay-as-you-go’ model. Cloud service providers charge their consumers for the services they use [5]. Consumers should have a guarantee that the services they are paying for should be delivered to them uninterrupted. This is provided through SLAs (Service Level Agreements) between the consumers and providers.

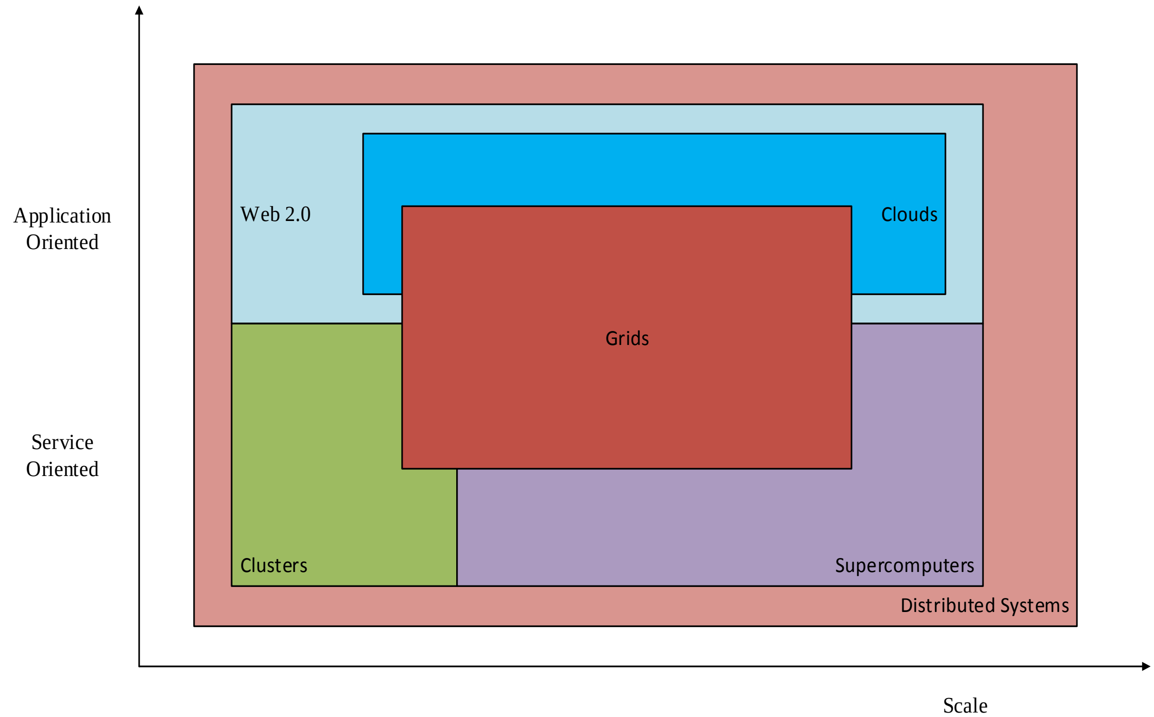


Figure 1.1: Overview of Utility Computing

Table 1.1: Different Characteristics of Grid, Cloud and Cluster Computing

Characteristics	**Grid Computing**	**Cloud Computing**	**Cluster Computing**
Residents	Servers and Clusters	Service User and Servers	Service User
Service Commitment	SLA's	SLA's	Limited
Resources Management	Distributed	Distributed or Centralized	Centralized
Resource Allocation	Decentralized	Centralized or Decentralized	Centralized
Privacy	Public or Private key for Authentication	Each user with its VMs Password	Old login Password Based
Application Areas	Scientific and Complex Computing Application	Dynamically and multi-tenant Web Services	Business Oriented

Cloud Computing, mainly focuses on two trends from the puddle of information technology – 1) IT Efficiency and 2) Swiftness of Business, from the areas of IT effectiveness, this approach should utilize the pricing techniques for contact establishment between resource providers and its end side consumers. Area of distributing computing provides its stakeholders to explore their business perspective, it provides them the thought in their business and it is done by the actual owner of these contracts. There is no need of any third party stakeholder [6]. From the business point of view, there are many hurdles in this ever changing environment of Cloud Computing. The economic aspect is a key factor in this area and pricing model techniques are very crucial in generating revenue of stakeholder of various Cloud service providers.

1.2 Characteristics of Cloud Computing

Many researchers explain the essential characteristics of Cloud Computing, some of them are discussed below [5] [9] [4] [8] [7]:

- Misconception of unlimited computing resource available on-demand- Cloud Computing delivers an infinite number of computing resources to its consumer on-demand over the internet. This service eases the working environment for consumers. All the scheduling and provisioning work is automated and dynamic, so there is no human interaction in scheduling and provisioning. Cloud Computing provides a flexible and scalable environment, consumers can lend IT resources and can customize their requirements.
- No capital cost to build Cloud Computing Services- Capital cost is the cost associated with adaptation of a new service. For organization that wants to build Cloud Computing there is no need of any initial investment. Before Cloud, providers need to invest in a capital cost to establish computing services. This feature of Cloud Computing helps small organizations to build new services and grow their businesses.
- Pay-as-you-go- "Pay-as-you-go" is a model which is used by several SaaS providers to provide services to their consumers. This model works on the principle of 'the more you consume, the more you have to pay'. This model features Cloud Computing as utility computing. Several Cloud Computing providers use this approach as a pricing technique including big giants of the IT

industry like Google App Engine, Amazon and Microsoft Windows Azure [10] [11] [12].

- Assurance for QoS- Service level agreement (SLA) provides the assurance of quality with the help of algorithms associated to compensate any violation. This agreement is signed by both service provider and service consumer. The user need not to focus on the aspect of quality, any violation is compensated by the SLA.

These are the main characteristics associated to Cloud Computing.

1.3 Cloud Computing Applications

Cloud Computing applications and consumers are increasing day by day over the internet due to its flexibility and scalability. Cloud Computing is now used by almost every industry, whether big or small. The main reason why people are shifting towards Cloud Computing is that there is no initial upfront commitment. Masses around the globe associated with information technology are now realizing the importance of Cloud Computing. Major problems associated with internet can be solved through Cloud Computing.

The main reasons behind success of Cloud Computing are:

- Adoption of new technologies associated with service computing.
- Decreasing infrastructure cost, increasing features associated with computing, like storage, computing power etc.
- Data size at the consumer side is very large and increasing exponentially.

This section presents some Cloud Computing applications on the basis of scientific and productivity applications.

1.3.1 Scientific Applications

These types of applications requires high computation power and throughput. Researchers in scientific field require very large amount of resources but for very short span of time. Cloud Computing serves best deal for these type of applications. Below some of successful applications of Cloud Computing in area have been discussed.

- ECG Analysis in the Cloud- Continuous monitoring of health is new buzz word in the field of medicines. In this application Cloud is used to analyse patients Electro Cardio Graph (ECG) and make prediction about any heart diseases. An instrument is placed near to patient heart which continuously monitor heart condition. Data from this instrument is received by mobile phone. Mobile phone forwards this data to Cloud which analysia this data.
- Satellite Image Processing- Spatial and Non-spatial data is collected by satellites. This raw data requires very large amount of processing to make it useful to Geographic Information System (GIS). A Cloud based application is developed By Indian Space Research Organization (ISRO) to process these images [52]. A Cloud based application was developed by ISRO to process these images.

1.3.2 Productivity Applications

Productivity applications use Cloud Computing for some commonly used desktop applications. Documents storage, office automation, website deployment like tasks can be done with ease and effectively using Cloud. Below some of productivity applications using Cloud Computing are :

- Drop Box- Dropbox [53] is a file hosting service operated by Dropbox, Inc., that offers Cloud storage, file synchronizations, and client software. Dropbox allows users to create a special folder on each of their computers, which Dropbox then synchronizes so that it appears to be the same folder regardless of which computer is used to view it.
- RoboEarth- It is a European project led by the Eindhoven University of Technology, Netherlands, to develop a WWW for robots, a giant database where robots can share information about objects, environments, and tasks. Researchers at ASORO (A-Star Social Robotics laboratory, Singapore) have built a Cloud Computing infrastructure that allows robots to generate 3-D maps of their environments much faster than they could with their on-board computers [54].

1.4 Cloud Computing Service Models

This section of the first chapter will discuss three service models or layers associated with Cloud Computing. Cloud platform deals with its end user as a request-response model, client (end user) sends a request (on demand service) to Cloud service provider and depending on the service they are requested, provider responds to consumers with the associated services model. Services model can be Software as a service, Platform as a service, Infrastructure as a service, and Business Process as a Service [13]. These Models contain several layers, each layer has its own specific work to be done. In this model, a Client requests for a service and depending on that service Cloud architecture provides those layers which are essential for that service. Cloud Computing used the internet as its network, any user can access any service which a service provider provides from anywhere. Figure 1.1 discusses various layers associated with the different Cloud Service Model. The details of models are:

- Software as a Service (SaaS) - As the name suggests that software's are used as utility services; in this user provides an application which is runs on client site. The user doesn't know where the actual application is running but the user can still store, deploy its data. Only need that a user should have a client interface that can be any web browser or and a stable internet connection. The consumer need not managed any of the internal structure of the Cloud Computing. Many Services are offered as a SaaS by big giants of IT like Google, Microsoft, Amazon, Zoho etc. [14].
- Platform as a Service (PaaS) - In this service, the model user provides a platform to build their applications over the web. In this case there is a lot of reduction in the development cost of the applications. In this service, users can develop and deploy an application. Some researchers consider PaaS as a part of SaaS with some fewer services than SaaS. In this compiler, operating system, office suites, etc.is provided to end user as a service. The user needs to maintain its data and application for access any platform provided as a service to end user. Several development models are also available as a service like WAMP and LAMP Model (AMP stands for Apache servers, Mysql, PHP) and 'W' is for the windows

environment and ‘L’ stands for Linux environment. Several PaaS are available in the market these are Windows Azure, Google App Engine, OpenStack, Salesforce etc. [14].

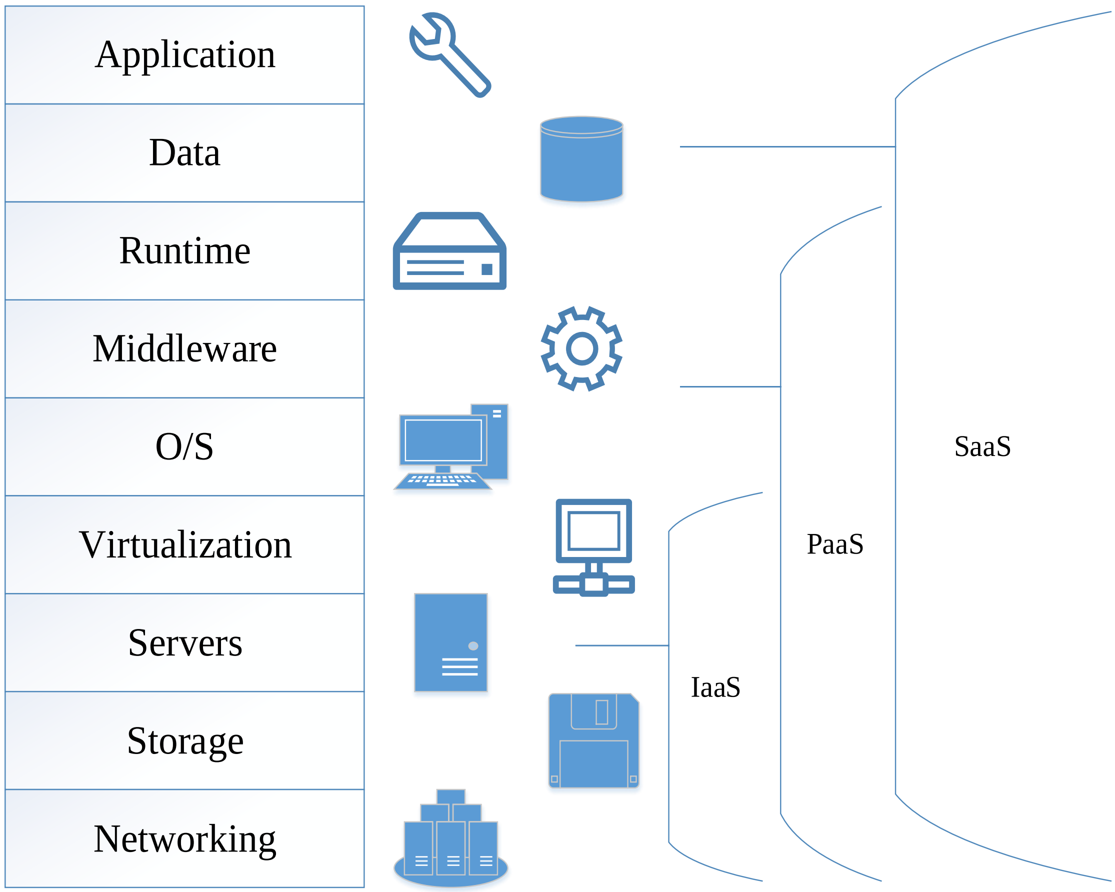


Figure 1.2: Layers of Each Service Model

- Infrastructure as a Service (IaaS) - As the name suggested in this computing machine’s infrastructure is provided as a service to the consumer. With this hardware, storage, network, processing power, servers and other computing resources, etc. are provided as a service. This service model deals with the virtualization because these resources are provided virtually to its end user. Consumers did not know the exact location of these resources but can access them from anywhere around the globe as an on-demand service. Various companies provide IaaS to consumers, which are AT&T, BlueLock, Joynet, Logicworks, HP etc. [14].

- Business Process as a Service (BPaaS) - Any business process which an organization does if that is provided to its end user with the help of Cloud then this is known as BPaaS. Several business processes like BPO, printing, payroll, M-Commerce, E-Commerce, etc. Services provided by this model to its consumer is on the basis of multi-tenancy in which only one server application is there and several multiple clients (tenants) are served by that application. BPaaS comes above the all other services provided by the Cloud architecture and BPaaS is integrated service consuming from other Services. BPaaS provides by TCS, Cognizant, and Wipro etc. [14].

These are the service models used in the Cloud Computing.

1.5 Cloud Computing Deployment Models

In this section, different deployment models used in Cloud Computing are studied. As Cloud Computing is an emerging technology and it is found that about 81% businesses are dealing with Cloud Computing either directly or build their business operation over Cloud this study is conducted by KPMG [15]. Every business has their own need depending upon the size, access and possession which tends to choose a different type of Cloud environment. Cloud Computing is more than a technology. It offers various services to its end consumers. Cloud Computing deals with various infrastructures associated with Computing and it contains various layers associated with different service models. To choose Cloud Computing as a business utility efficiently, each business should know their requirements before moving to Cloud interface. Cloud deployment/delivery model can be briefly classified into four categories. Figure 1.3 shows a pictorial representation of different Cloud deployment models.

- Public Cloud
- Private Cloud
- Community Cloud
- Hybrid Cloud

- Public Cloud- It is a classical model which provides Cloud as an infrastructure, platform, application/software, Information, business process as a service to the common public with the help of internet. The services are provided as a metered service by which pre investments are not required in these service models. Services are provided by a third party vendor using the internet. Gmail, Google, Facebook, is the example of public Clouds. Applications given by public Clouds are online storage, blogs, online backup, data-intensive workloads etc. [16].
- Private Cloud- The Cloud infrastructure within an organization is managed by a private organization for their employees. This service is used by a single organization privately. This service works inside the firewall of an organization and runs on its own website servers. Private Clouds are a centralized approach by which security and resilience are better than public Cloud. Services given by private Cloud are virtualization, multi-tenancy, security access and they are not just organizing centric nowadays they consume their services for public Clouds [16].

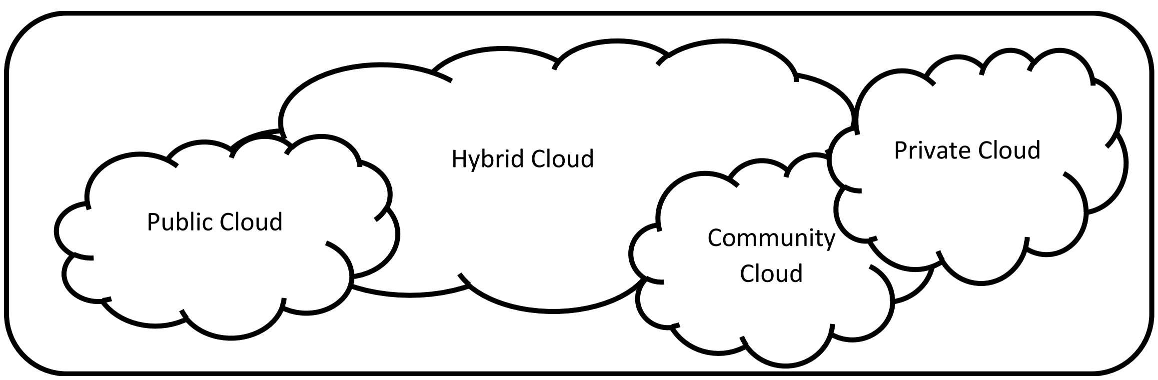


Figure 1.3: Different Deployment Model for Cloud Computing

- Community Cloud- This Cloud Computing infrastructure is not just an organization-centric, in this many organizations having same fundamentals share common Cloud Computing infrastructure. All organizations need not establish all their resource essentials for computing, they can share their resources and provide a better computing model. This is an expensive approach because resources are more than the private Cloud and business growth is involved in this. But this provides maximum security and policies for the services they are offering [16].

- Hybrid Cloud**-** As the name is suggesting this approach provides you the fusion of public and private Cloud Computing to solve some computing requirements. In this end user can outsource the public/open services whereas business applications can be processed with the help of private Clouds. This Cloud structure faces the problem of load balancing in the field of computing applications. There can be trusted or the untrusted environment as per the demand of service. Elasticity is the most common attribute for hybrid Cloud Computing. Secure data can be accessed by private Cloud Computing [16].

1.6 Cloud Computing Challenges

This section discusses the challenges and issues associated with Cloud Computing. Cloud Computing [2] is a vast technology, which offeres on-demand service to its consumer and this will affect the whole globe in the coming 20 years.

Some of these challenges are:

- Fault Tolerance- With the advancement of Cloud Computing's scalability, the fault tolerance is an essential need in the development of next generation computing models. Today's fault tolerant algorithm works on the demand of today's workloads. Cloud customers increases exponentially this failure tolerate algorithms will be outdated soon. There can be an internal factor or external factor which leads to failure in the system. The challenge in this emerging area is to design a fault tolerated system which should be easy to maintain.
- Security and Privacy- Cloud Computing market is going to be a more than a trillion dollar business in coming years. Security and privacy of user data is a big concern in the growing of Cloud market. Cloud providers have the access to all consumers' data over a Cloud, so consumer hesitates to provide all the data. The feature of accessing consumer's data from any geographical location, the privacy of consumer is violated. Information security helps the consumer to keep their useful data as a secret. But dealing with confidential data by the third party makes consumer unsecure. Consumers' information over the Cloud platform is in a decrypted form, but the ability to access it from all virtual machines hosted on

that platform makes it vulnerable. The trust also plays an important role in the enhancement of Cloud Computing. Consumers' have to choose which provider is trustable and which is not. Numerous agreements are signed for any violation by the service provider and consumer. But if the violations of security and privacy hits and any illegal action are taken place against the consumer data, it is a difficult task to find out, who is responsible for this in a virtualized environment. The issues in this area is to provide a secure and trustable Cloud environment.

- Scalability- Cloud Computing is having one of the attractive features which is on-demand scalability of resources. This feature leads to some of the issues occurring in the scalability of Cloud environment. This leads to vendor lock-in if the interoperability is not provided to scalable Cloud Computing. Any working Cloud should design with scalability in mind. The scalability algorithms running nowadays will be outdated soon because the population of Cloud Consumers is increasing exponentially. The issues in this area is to design a highly scalable Cloud environment.
- Load Balancing- Load balancing is an approach to balance the load among the components of Cloud Computing. The need of this approach is to provide an efficient environment in an emerging area which is enlarging day by day (Cloud Computing). Load has to be spread uniformly among all the participating nodes; the load can be from different perspectives like memory perspective, network perspective, and processing power perspective. So to provide efficiency in the computation of Cloud the emergence of load balancing approaches in the Cloud is needed. The issue of load balancing among Cloud is new emerging area of research in Cloud Computing. A centralized approach is not efficient in the handling of load balancing of resources among the nodes of Cloud Computing more decentralized approach is needed.
- Pricing in Cloud Computing- Cloud Computing is an on-demand service providing approach, so the consumer should have guaranteed that the services they are paying for should be delivered to them uninterrupted. This is provided through SLAs (Service Level Agreements) between the consumers and providers. Some economic models in literature survey are consumer-centric whereas some

are provider centric. The economy problem is the latest issue in the field of Cloud Computing, to resolve this emerging challenge Cloud application should serve both consumers and providers. Cloud Computing environment should be designed to minimize the SLA violations.

- Green Computing- Cloud Computing deals with several Data centers to provide the Iaas to its consumers. Data centers cannot be limited because as the number of users is increasing rapidly, the information they are sharing on the internet needed to be stored and for that large storage space is necessary. Although CPU is the major energy consumption part of the computing system, but in case of data centers concentration is moving towards memory management. In case of power failures, big generators are used for providing power supply to the data centers. The whole procedure consumes an ample amount of electricity being generated all over the world. To work efficiently in the Cloud more green computing approaches should be adopted.

There is a requirement to provide flexibility in pricing, rapid provisioning of resources, green solutions with low carbon footprint, scalability, testability and stability over old-style IT deployment models.

1.7 Research Motivation

With the ease of internet one can access any information from anywhere within a few seconds, this technology consists of various data centers, their work is to provide content to requested end user accurately and consistently. Cloud Computing uses this approach to establish its architecture and it provides business application capabilities as a service to end consumer that can be accessed over the internet. Cloud service providers charge their consumers with ‘Pay-as-you-go’ Model i.e. the consumer has to pay for the services he/she uses. Cloud Computing is an on-demand service, so consumers are guaranteed that the services they are paying for get delivered to them uninterrupted. This is provided through Service Level Agreements (SLAs).

Cloud provider provides the data and computing resources dynamically to the Cloud users based on their requirements over the internet. It is a challenging task to maintain the required Quality of Service (QoS) to fulfill the expectations of Cloud consumers [17]. The complexity of management of resources in Clouds is increasing day by day, so an efficient technique is required for management of resources. To solve this problem, there should be a focus on provisioning before actual scheduling of resources. Resource provisioning techniques help to improve the efficiency of computing resources in such a way so as to reduce the execution time and cost while considering other QoS parameters like availability, network bandwidth, serviceability and customer confidence level.

There are several challenges [18] in Cloud Computing. An efficient economic and pricing model is the emerging challenge for an enterprise working in the field of Cloud Computing. The motivation of this research work stems from the challenges in management and an efficient utilization of the resources. In real life situations, there are many constraints including (i) satisfying the QoS (ii) increasing service providers' revenue (iii) minimizing the provider cost problem (iv) improving customer satisfaction. To schedule the resources along with the QoS constraints is the main objective of this research work. This leads to develop a pricing model using an economic model called the Time Dependent Pricing (TDP) model. This thesis work aims to develop a Smart Cloud Computing environment, which constitute a framework which is automated and can distribute Cloud resources over a communication network that uses two-way communication between service providers and consumers. Smart Cloud Computing environment works on calculating prices of Cloud resources using TDP. Such approach helps resource providers to practice a versatile pricing to increase their profits/revenues by covering off-peak demand and executing a better Cloud resource consumption.

1.8 Organization of Thesis

Rest of the thesis is organized as follows:

Chapter 2 – This Chapter summarizes in detail the literature survey done to study the concept of pricing technique in Cloud Computing, economic models of resources in Cloud Computing and Various Secheduling Algorithm in Cloud Computing.

Chapter 3 – This Chapter focuses on gap analysis of this thesis. This chapter also gives the problem statement.

Chapter 4 – This chapter of thesis provides solution to the problem stated in previous chapter by proposing the architecture, schematic framework model and explaining all its component. Designing of system using UML diagram has also been presented.

Chapter 5 – This Chapter demonstrates proposed framework model by implementing it using CloudSIM, MS Visual Studio. Result of experimental setup is shown.

Chapter 6 – This Chapter describes the conclusion, contribution of work done and future research possible.

Chapter 2

Literature Survey

In the previous chapter, Cloud Computing and its application areas followed by its service models and deployment models have been introduced. The crucial challenge and issue in the field of Cloud Computing is the pricing of resources between the service provider and service consumer. Cloud Computing has two peers: resource providers and resource consumers. Resource providers have a different aspect of cost minimization, i.e. an increase in revenue and better utilization of resources. Resource consumers have different aspects, i.e. better Quality of Service (QoS), more rewards and discount, no SLA violations. Focusing on this new emerging challenges, this chapter discusses the research work carried out in the field of pricing in Cloud Computing. In this chapter firstly various research questions are discussed which help to analyze various research papers and currently advised pricing models. After that, this chapter discusses state-of-the-art in the pricing models between Cloud and Grid Computing.

2.1 Research Questions

This literature review consists of the present state of the art in the field of pricing in Cloud Computing by proposing the answers to the various present research questions.

1. What is the current status of pricing models in Cloud Computing?
2. How to minimize the providers' cost minimization problem?
3. Which economic model is identified in the existing pricing models?
4. How to overcome the issues of scheduling resources with efficient cost and execution time?
5. Which pricing model does justice to its peers (i.e. service providers and service consumers)?
6. How to validate the various economic models through different tools?

The literature review system started with definite research questions and after that the search is followed by defining the specific keywords related to the research questions.

Table 2.1 specifies the keywords used to find the literature survey. Using these specific keywords, the in-depth question of research was allowed to find issues in the pricing of resources in Cloud and Grid Computing.

Table 2.1: Keywords

Keywords	**Synonyms**
Cloud Computing	Grid Computing, Utility Computing
Pricing	Cost, SLA, Economic Model
Scheduling	Provisioning

2.2 Existing Pricing Models in Cloud Computing

In this section of the literature review, various existing pricing models used by Cloud Computing giants like Amazon, Google etc. are studied and then various existing pricing models by different researchers are studied and their pros and cons are discussed.

There are several problems which providers, as well as consumers bear during the usage of Cloud Computing as a utility service. Pricing of resources is one of the crucial issues in the field of Cloud Computing. Several pricing models have been used by big giants of Cloud Computing. "Pay-as-you-go" is a model which is used by several SaaS providers to provide services to their consumers. This model works on the principle of 'the more you consume, the more you have to pay'. This model features Cloud Computing as utility computing. Several Cloud Computing providers use this approach as a pricing model which includes big giants of the IT industry like Google App Engine, Amazon, and Microsoft Windows Azure [10] [11] [12]. This model makes Cloud Computing environment very flexible because the cost is calculated on the basis of usage. vThunder Pay-as-you-go Licensing Model provides various IaaS providers to offer servers L4 and L7 advanced networking tenant details and services like reporting, billing, license management and metering [19]. The service it provides makes it elastic and flexible for various Cloud Service Providers (CSP). Consumers can adopt vThunder with automated licensing possibilities like Rental Billing Model (RBM), Utility Billing Model (UBM). The working of RBM and UBM is shown in Table 2.2.

Subscription and Pay-as-you-go models are the pricing models used by various Cloud Computing Service Providers. Table 2.2 explains the live example of how service is charged over the internet. Next section explains these pricing models in detail and discusses their advantages and disadvantages to its peers (service provider and service consumer).

Table 2.2: Difference between Different Pay-As-You-Go Business Models

RBM (Rental Billing Model)	UBM (Utility Billing Model)
This model generates bills with a fixed bandwidth, license for a given period of time say a month.	This model generates bills based on consumption, license per byte.
Bandwidth is fixed with some MBPS increments.	Billing is based on actual data consumed.
Deterministic approach for billing.	Non Deterministic approach for billing.

2.2.1 Pay-as you-go Pricing Model

Cloud service providers charge their consumers with 'Pay-as-you-go' Model by which the consumer has to pay for the services he/she has used. Providers can make profits by charging their customers for the usage of the service which they provide over the internet. The pay-as-you-go model is a generic model and various 3rd party agents use this approach to calculate the bill for various resources consumed. This approach is not flexible for all CSP (for example vThunder is flexible for networking tenant and L4-7 advanced networking layers as a service). A model is needed which can be reusable and is flexible to various CSPs.

Table 2.3: Study of Pay-as-you-go model in Cloud Computing

Pricing Model	Pricing Scheme	Justice	Pros	Cons
Pay-As-You-Go	In this method authority to set price is with the service provider.	This model is not justifiable to service consumer because he has to pay more than he/she consumed as in RBM	Resources are kept for the consumers for a time period for which he/she paid.	Provider has to reserve resources for a longer time period and provider cannot vary the prices as the need of time.

The pay-as-you-go model is used by various service providers. This pricing model is not beneficial for the service providers. The service provider has to reserve the resources even if the consumer is not consuming it. The service provider can't increase the price of resources even if the demand is rising for any of the resources. This will not help the provider to minimize their cost problem. There is no provision to shift load of the consumer to off-peak load periods.

2.2.2 Subscription Pricing Model

A subscription is an arrangement by which something is received by regularly paying in advance. In Cloud Computing consumers subscribe for various resources available over the network. This approach is used with the Pay-as-you-go model. [11, 12, 13] have commonly used this approach. Qun Zhou et al. [20] have developed CCOA (Cloud Computing Open Architecture) in which they state the seven principles of Cloud Computing. In 5th principle, the IaaS providers have to manage all the subscription as well as provisioning resources. This architecture leverages the providers to support subscription. In their architecture, subscription of the Cloud is enabled by order creation and offering a creation processes.

Table 2.4: Study of Subscription model in Cloud Computing

Pricing Model	Pricing Scheme	Justice	Pros	Cons
Subscription	Pricing scheme can vary depending upon subscription but remains fit for entire period.	Consumer may pay more or less depending upon subscription period.	Consumer has to pay less if the resources are reserved and uses them extensively.	Provider has to pay more if the resources are subscribed for long period.

The subscription model has some advantages and disadvantages as well. Subscription model causes a lock-in situation. Consumers have to pay more in the subscription model even if they are not using any resources. This model also does not solve the problem of cost minimization and neither allows the providers to maximize their revenue. There is no scope for any rewards and discount in the existing pricing models.

2.2.3 State-of-the-art in Pricing Solutions

Some of the existing pricing models used and simulated by different researcher are explained in this section.

- Buyya et al. [21] proposed a pricing model in which they design, simulate and develop with two constraints in mind, i.e. providing dynamic flexibility in terms of QoS for consumers' perspective and profitability for the providers' perspective. Various Cloud resources which the vendors or providers providing they call it as C3 (Cloud Compute Commodity). They manage C3 as assets and deploy using three steps a) they model C3 model as a pricing problem option; b) QoS is the backbone for providing assets in an efficient manner; c) they calculate price using approach of continuous time monitoring. They used this methodology to develop and simulate the model with two forms, one from economics background BSM (Black-Scholes-Merton) [22] [23] and the other from the technological evolution of Moore’s Law [24]. Using models proposed in [22] [23] one can calculate a closed form of a solution which is from European put and European call. In [21] they used the approach to calculate the European call to calculate the pricing model with an upper bound and lower bound. A price that is in between a lower and upper bound would be beneficial for both user and provider. The result in [21] shows various Cloud parameters like initial investment from providers, QoS, Contract time, resource age and rate of depreciation with Cloud resources which is resource price generally.

- Hao Li et al. [25] proposed a Cloud banking model’s architecture. In this architecture, they simulated a 5 layer architecture in which all flexibility is achieved in the middle layer which consists pricing strategy, information server, and Cloud dictionary server. There are two layers for an agency which communicates with providers (Cloud Service Agent), (Cloud Resource Agent) and consumers (Resource Consumer Agency). Each Agent has its own task which has to be maintained for the dynamic approach of price calculation. They borrowed this idea from G Commerce [26] and proposed an economic approach for pricing calculation. In this model, they provide a pricing algorithm by which they calculate the cost/price say P from described task obtained from agents. Then it sends this cost to various agents. In this approach three agents are used and obtain the user information from various agencies and perform tasks of the user and get the benefit.
- Mihailescuet al. [27] generates the dynamic pricing approach for resources via a third party component called as market-maker. The results they simulated were on the basis of user welfare and successful requests. For the consumer welfare, they simulate the results in a centralized market-maker where each consumer and resource provider connects to a peer to peer network which can be achieved by FreePastry [28] an open-source DHT network environment. The economic

perspective in [27] is achieved for resources allocated via a third party Market-Maker using Individual rationality. Using this each participant whether user or provider acts as a rational participant and gains higher utility because they are participating in the resource allocation method. Incentive Compatibility thus ensures that each participant in this algorithm is truth-telling. Budget Balance confirms that the total sum of all the payments made by the user is equal to the payment received by the seller. Pareto Efficiency can be achieved for an allocation as no improvement can be made by no participant better off without making other participants a worse off. It simulates the results with dynamic pricing which shows an increase in buyer's welfare and average performance for dynamic pricing is better than fixed pricing.

- Mario et al. [29] present a genetic model for calculating the exercise price for the Cloud Computing markets. The genetic algorithm they simulated; one of the advantages of this is that it can easily build in the decentralized architecture. When the genetic model is applied to price calculation, it implies to solve the three issues in the calculation of accurate naive pricing function. Naïve function calculates the price in the actual system. The three issues are *Define a Chromosome, Evaluating the Chromosome, Selection and Reproduction of Chromosome.*

- Sewook Wee [30] studied the SI (Spot Instance) launched by AWS (Amazon Web Service) over a time period of one year. Some of the services are fixed price till now in AWS. But before SI every service that AWS leases to users are on the basis of fixed price. So they studied the historical data of AWS instances. You can achieve an even more optimised price for a service instance if the user shifts his/her load from peak hours to fewer load hours. Wee's study suggests that on an average only 3.7% cost is saved whereas 52.3% cost is saved when you shift from fixed pricing scheme to Dynamic pricing as AWS launches its dynamic pricing for their instances on December 2009. No extra cost is calculated in this study like the cost of transferring the loads. Ideal Reliability mechanism is used to handle the resources when the price increases to maximum bid price.

- Dhawas et al. [31] focuses on the security breach and outrage of user's data over the provider's side. The secure model focus on the how a data backup is to be maintained by other providers if any one of the providers goes unavailable or goes bankrupt. The proposed model uses the linear programming problem to propose the given secured cost model. They develop an economic destitution of user's data over several providers with the availability of data and secure storage take place over several providers.
- Sururah et al. [32] proposed ideas about how flexibility in the pricing model associated with Cloud Computing can be achieved for SaaS users. They implement different use cases to understand the possible timings in the pricing schemes in the Cloud Computing. The rewards given to the user on the basis of its usage of Cloud are also discussed.

Table 2.5 explains the various existing pricing models proposed by various researchers.

Table 2.5: Study of various Existing Pricing Models used in Cloud Computing

Pricing Model	Pricing Scheme	Justice	Pros	Cons
A Novel Financial Economic Model (C^3) [21]	Pricing scheme is calculated on the basis of dynamic usage with initial investment and strike investment.	Beneficial for both if the price is in between the upper and lower bound.	(i) It helps providers to recover its investment before the age of resources. (ii) QoS is provided to users.	(i) It assumes the volatility (σ) of market is constant which is not true in real market scenario. (ii) Cost of maintenance of resources is not included in the investment
Pricing algorithm for Cloud Computing Resources [25]	Pricing scheme is calculated in real time. Dynamic in nature.	Beneficial for the provider, because it maximizes the revenue and reduces the cost.	(i) It helps the provider to increase revenue.	(i) This pricing algorithm is fixed and can't adapt to the rapid changes in the demands. (ii) The amount of revenue increased by service provider is not explained in this approach. Statics needed to justify the work.

Dynamic resource pricing on federated Cloud [27]	It uses the auction based concept which is dynamic in nature.	It is beneficial for both the users and the providers, because in this the price is generated by a third party	(i) Its simulated result shows that it is better than fixed pricing approach and increases the buyer's welfare up to 200%.	(i) The results of simulation shows that fixed pricing have a better result in the case of over demand. (ii) The scalability is less for the high demand in the market
A Genetic Model for Pricing in Cloud Computing Markets [29]	Pricing scheme is calculated real time. Dynamic in nature.	It is biased towards the providers to increase its revenue, the genetic approach simulated increase the revenue.	(i) Increases the revenue. (ii) Steady in noises. (iii) Genetic algorithms are easy to implement. (iv) Flexible due to naïve function.	(i) It assumes that the participants in the markets are rational which degrades this model. (ii) It assumes that market is near about equal demand but underperforms in the very high and low demand of resources.
Debunking Real-Time Pricing in Cloud Computing [30]	Pricing is calculated on the RTP (Real Time Pricing).	In this a historic data is studied and that checks if the user shifts their loads to cheaper hours. No extra cost is calculated in the study like cost of transferring the loads.	(i) In this incentive is not measured for users to shift loads from peak to cheap hour. (ii) Shifting favors the providers in their resource management.	(i) The additional cost saving price can be achieved through scheduler. (ii) Ideal Reliability mechanism to handle the resources when the price increased to maximum bid price.
A Secured Cost Effective Multi-Cloud Storage (SCMCS) [31]	Each user has a specific number of providers to choose a QoS provider. Using this approach they proposed a linear programming assignment problem to solve the security of user data and user has to pay per unit storage data.	The main focus of the research is to overcome the breach of privacy and outage of service of users' data. If any of the service provider goes bankrupt or become unavailable user can able to access his data to other providers.	(i) The main focus of this model is to overcome the security issues and breach in privacy reports [33]. (ii) Secured cost model divides the data among the various providers within the budget.	(i) The dynamic nature of the market is not suggested in the given model. (ii) Threats on Cloud market are not suggested in this model. (iii) A few threats have been discussed in this model.

Flexible pricing model for Cloud services [32]	The authors focus on the varying nature of pride in Cloud Computing.	The main focus in this is towards the user. Approaches and rewards are more user-centric in this paper. Main focus is on improving the pricing of SaaS and PaaS.	(i) The use cases and rewarding approaches are the key in the flexibility of pricing strategy in the given Cloud Computing.	(i) In this study no specific implementation is done. (ii) Only a theoretical aspect of pricing is discussed in the pricing model.

Buyya et al. [34] discuss various pricing models for various resources (goods and services) used for Grid Computing and Cloud Computing with economics perspective. Following table 2.6 shows the complete perspective of these models.

Table 2.6: Study of various Economic Models

Economic Model for Resources	Description
Bargaining Model	This class of models represents those models in which resource (Grid or Cloud) providers and resource consumers discuss with each other until they meet their aims and agree on a price.
Commodity Market Model	In this model, various pricing model can be used, the approach of this model is ‘more you consume more you have to pay’. The price of resources is calculated on the basis of consumption. Different pricing model which comes under this class are- Flat fee, Subscription, Usage Duration.
Posted Price Model	The model is similar to the commodity market model but the additional paradigm this model provides is known as advertising. Providers attract their consumers through advertisements.
Auction Model	Auctioneer set the rules which are accepted by both providers and consumers. Different consumers auction for the resources and negotiate and agree on a final price.
Monopoly Oligopoly	This model shows the dominance of a Grid Service Provider (GSP). GSP is the only provider of particular resource, service or protocols. User have to restrict to the prices set by the GSP.-
Contract-Net model	In order to control the usage of services and resources, the model is used for contracting between providers and consumers.
Bartering Model	This model works on the perspective hybrid Cloud, individual or institutions create a computing environment and share their resources.
Bid Based Model	This model works on a bid value or a threshold value, which states the amount of share customer had for various resources.

Some of the existing economic models used by various pricing models are explained in following section.

- Guan et al. [35] simulated the double auction bidding mechanism to allocate the resources and increase the revenue of a Cloud Service Provider (CSP). The mechanism used in the approach is of truthfulness because they apply Second

Auction Mechanism due to which dynamic resources increase as well. The main focus of this approach is provider-centric to increase revenue.

- Kento et al.[36] proposed a mechanism by which users can enable the current and future services in the market which can be spot/forward market. Users can also order a mixture of co-allocations and services of resources together. The authors of this paper suggest and simulated a model from the aspect of future need of Cloud Computing where thousands of customers bid for a suitable resource.
- Shifeng et al. [37] study the recent flows in the Cloud market and propose a double auction-based pricing scheme. With this approach, they first develop a Cloud market which is uniform and provides a competitive nature to several Cloud service providers. They focus on how a pricing scheme can be more flexible in resource trading.
- Yang et al. [38] provide an autonomous pricing calculation on the basis of resource bidding and quoting in which the price of resources is fixed by the providers. The simulated results show how this approach of pricing can be concurrent with the partial equilibrium theory.
- Pourebrahimi et al. [39] proposed a pricing model which is based on one of the traditional pricing economic models. The authors simulate their pricing algorithm on the continuous double auction mechanism. They simulate their result on the grid environment with three aspects: A Balanced network, task intensive network, resource intensive network.
- Michael et al. [40] provides two types of pricing models in a combinatorial grid where the scheduler provides the allocation of resources such as memory and CPU time simultaneously. The types of pricing models are Shadow Pricing and Scarcity Pricing. The pricing model used on the auction economic model which is English Auction. The Shadow pricing based on the linear distribution of resource whereas the Scarcity pricing overcome the problem associated with linear pricing program.
- Xiangang Zhao et al. [41] simulated a pricing model which is based on demand prediction and task classification. As stochastic processes with a specific property know as Markov Chain, by which the authors are able to predict the demand of

resources. As grids are dynamic in nature so as the available resources in the grid market and they are also stochastic in nature so the Markov Chain can be applied. By the help of Markov Chain the prediction of demand/load of the given resources can be done in the future and then a mechanism to adjust the prices of the resources. As price and demand are mutually dependent on each other, the pricing models which are considered for the profit of resource consumers as well as load balancing of the available resources is to be applied. They also define the tasks which are an exclusive task and shared task depending upon this they provide two different pricing strategies.

Table 2.7 explains the pros and cons of existing economic models proposed by various researchers.

Table 2.7: Study of various Existing Economic Models used in Cloud Computing

Pricing Model	**Pricing Scheme**	**Justice**	**Economic Model**	**Pros**	**Cons**
A Dynamic Auction Mechanism [35]	This algorithm develops a resource allocation mechanism with double and efficient allocation of resources to the users.	Beneficial for the providers with the best dynamically allocated resources and time efficient background computing.	Bidding methods which are marginal and sealed bid.	(i) It helps providers to improve its revenue by setting equilibrium in the tasks. (ii) The proposed algorithm used Gaussian distribution.	(i) As the scale and scope increases so there is the need to make some changes to proposed approach.
A Combinational Double-Sided Auction to Resource Allocation in Cloud Computing [36]	Pricing scheme W-Mart is develop after the architecture of U-Mart [42]. W-Mart contrasts only two market aspect which is spot/forward.	The goal of this pricing model is to develop a well-organized marketplace for Cloud atmosphere. Benefit for the users as they can order as per their requirements.	Double Sided Action Model using K pricing scheme.	(i) It helps both parties as it is a combinational auction so each participant will be beneficial to this auction. (ii) The deal is done when both user and a service provider is satisfied.	(i) As the spot and forward market are independent, if the situation demands to interact between spot and forward market, which is not done in this pricing scheme.

DABGPM [37]	They used a double auction model to provide a flexible pricing model in the Cloud competitive environment.	This approach gives is more centric towards consumer, but it also helpful for providers.	Double Sided Auction Model.	(i) Double auction model supports both the user and providers in the pricing scheme. (ii) This model makes it more flexible, more competitive in the market.	i) This architecture should follow in a competitive environment where many providers compete for one or two clients.
An Autonomous Pricing Strategy [38]	Pricing scheme is fixed by the service providers, the price is set on the basis of decision making.	Beneficial for the providers due to decision making approach.	Bidding	(i) It helps providers to recover its revenue because tries to achieve partial equilibrium theory.	i) It is not scalable as other pricing models. ii) More control of price is in provider's hand.
Market based Continuous double auction resource pricing algorithm [39]	Pricing scheme is calculated on the knowledge of consumer and producer.	Beneficial for the providers due to decision making approach.	Double Auction Approach with continuous aspect.	(i) It helps Consumer and providers an efficient resources utilization from both aspects.	i) It shows better result in the simulated environment but real time changing factor is not considered in this.
Dynamic Resource allocation with combinatorial Auction [40]	This approach is an agent based. Three agents are proposed which are Resource agents, Task agents and Mediating Agents.	It is beneficial for the user perspective as it provides two pricing calculation because some time it is very expensive to calculate the shadow pricing.	Double Auction Approach which is English Auction	(i) It helps allocation of resources and pricing of these resources in a combinatorial auction environment of biding.	i) The approach they proposed in which the shadow approach of pricing becomes very complex and tedious.
Pricing Model with Task Classification and Demand Prediction in Grid [41]	The pricing model used Markov Chain to predict the uses of resources in grid.	This dynamic approach is suitable for real time market of grid resources. This model gives fulfillment of both the aspects of users as well as the providers.	Commodity Market Model	(i) This model is simulated over a grid market but result shows this can be associated with real time system as well. .	i) As Demand and Prices are mutually dependent so continuous monitoring is needed to both factors of resources.

2.3 Scheduling Policies

The resource scheduling techniques in distributed systems frequently have the objectives of distributing the workload on resources and increasing their resource consumption but reducing the time of workload execution. Existing resource scheduling/provisioning techniques have been presented in this section.

2.3.1 Resource Scheduling Techniques – without Quality of Service

- Quiroz et al. [43] proposed a distributed approach to determine resource scheduling of workload on innovativeness Clouds. To handle with erroneous request for resource that causes over utilization of resources delivered by Cloud workload request, their approach has revealed a design-based technique for approximating the workload execution time specified. It is a scheduling technique but the behaviour of a particular workload is not identified.
- Nguyen Van et al. [44] addresses the issue of autonomic VM administration for presenting Cloud services although improving an overall effectiveness function that adds the functional charge of the Cloud provider and the workload level SLAs. This approach tries to improve the efficiency of the running workloads with an optimum functional charge for the Cloud provider and controls the scaling of workloads executing automatically and effectively.
- Silva et al. [45] offered a heuristic based mechanism for consuming lesser amount of resources for execution of Cloud workload with fixed amount of applications that better utilization and performance can be attained within a predefined budget. Though, the nature of workloads coming is random and dynamic; therefore prediction of an optimal number of resources required is difficult to estimate along with the fulfilment of SLAs in desired deadline.

The scheduling policies used with QoS are discussed in next section.

2.3.2 Resource Scheduling Techniques – with Quality of Service

- Lua et al. [46] proposed QoS based autonomous SLA model of violation-filtering for IaaS and PaaS applying a SLA appraising many ways & penalty architecture and presenting a resource scheduling mechanism to manage resource efficiently without specifying the QoS parameters.

- Wu et al. [47] proposed SLA based scheduling technique to reduce resource price and SLA deviations. The management of customer requests, mapping them with resources is defined along with the supervision of different types of workloads by considering QoS such as execution time.
- Li et al. [48] presented a virtual environment based framework for better supervision of infrastructure which offers separation among workloads running concurrently with similar resources and sanctions dynamic horizontal and vertical scalability to realize Service Level Objectives. But it is difficult to allocate resources efficiently due to variation of an execution time of different workloads.
- Chieu et al. [49] proposed a framework for dynamic workloads based on a threshold number of dynamic periods for dynamic assigning and quickly scheduling of virtual resources to users based on their QoS requirements.
- Herbst et al. [50] described self-adaptive resource scheduling approach to finding the appropriate predicting ways for a given perspective though the use of decision tree. This approach considers QoS parameters like relative error and SLA violation and minimizes both but does not consider execution time and execution cost.
- Qavami et al. [51] proposed a dynamic behaviour based resource scheduling technique to assign an adequate number of resources to workloads. This approach reduces the execution cost of resources only without considering execution time.

2.4 Conclusion

Different pricing algorithms that are focused on the economic aspect and various scheduling policies have been studied and reviewed in this chapter. The study is classified into three categories existed pricing algorithms, proposed algorithms and scheduling polices. It shows the economic models used in the pricing algorithms that help to improve the providers' cost minimization problem and QoS based scheduling policy helps to improve the service consumer satisfaction. Next chapter discusses the gaps found in the literature and the objectives of this research work.

Chapter 3

Research Gap and Problem Statement

The previous chapter explained the literature survey of various pricing techniques used by various service providers and given by several researchers in the field of Cloud Computing and Grid Computing. Based on the literature survey, various gaps in the pricing techniques of Cloud Computing have been recognized in this chapter. In this chapter, the problem statement of this thesis is discussed. Objectives of this thesis are enlisted at the end of the chapter.

3.1 Gap Analysis

As in the previous chapter, literature survey of various existing pricing techniques is conducted, there are many gaps found on the basis of that literature survey. More than 70 percent pricing techniques are somewhat in favor of the consumer of Cloud Computing. Existing literature reports that almost all pricing models are consumer oriented. Only a few of them focus on maximization of providers' revenue. None of the existing models have considered the providers' cost minimization problem which is the one of the most challenging problem from the providers' aspect. An efficient scheduling technique which could minimize the execution time and the cost of resources is the need of the hour. All the existing work focuses on the pricing aspect of the resources but none of the proposed model is able to explain any scheduling policy in addition to pricing model.

The existing economic models help to improve consumers' satisfaction levels. But in these models there is no provision of shifting Cloud consumers' workload to off-peak periods where the availability of resources is more which helps the providers to maximize their revenues by utilizing the off-peak resources. Each existing model has its own gap. Table 3.1 shows various today's needs or perspectives of Cloud providers and Cloud consumers. Literature survey of existing pricing models based on these parameters of consumers and providers has done in Table 3.2. In Table 3.2 providers' and consumers' perspective in various pricing models has been mentioned. Table 3.2 is created using these questions in mind suggested in Table 3.1.

Table 3.1: Different Perspectives of Consumers and Providers in Cloud Computing

Provider Perspective	**Consumer Perspective**
Cost Minimization Problem	Consumer satisfaction on the basis of QoS
Increase in profit and revenues	Rewards and Discounts
Better Resource Utilization	Efficient Cost and Execution Time

Table 3.2: Comparison between Different Pricing Economic Models

Pricing Model	**Economic Model**	**Provider Perspective**	**Consumer Perspective**
Buyya [21]	Commodity Market Model	✓	✓
Hao Li [25]	Cloud Banking Model	✓	×
Mihailescu [27]	Market Model	×	✓
Mario [29]	Naïve Pricing Model	✓	×
Sewook [30]	Real Time Pricing	×	✓
Dhawas [31]	Posted Price Model	×	✓
Sururah [32]	Commodity Market Model	×	✓
Guan [35]	Second Priced Auction Model Using Sealed Bid	×	✓
Kento [36]	Double Side Auction Model using K pricing	×	✓
Shifeng [37]	Double Side Auction Model	×	✓
Yang [38]	Bidding	×	✓
Pourebrahimi [39]	Double Auction Approach with continuous aspect	✓	×
Michael [40]	Double Auction Approach(English Auction)	×	✓
Xiangang [41]	Commodity Market Model	×	✓

Keeping these perspectives in mind, there is a need of a model which could satisfy both provider's and consumer's needs.

3.2 Problem Formulation

An efficient pricing model is a need for both service providers as well as for service consumers. The main challenge is to develop a pricing algorithm which fulfills the perspective of both consumers and providers that helps to maximize the revenue and is able to solve the providers cost minimization problem. If the consumer can shift their workload to the off-peak workload hours in advance; this facility will provide consumers to earn rewards and discounts. After this, there is the need of an efficient scheduling algorithm which will schedule the resources with minimum execution time and minimum cost thus ensuring minimum SLA violations.

3.3 Objectives

The objectives of this thesis are

- To study the existing pricing models and scheduling techniques of Cloud Computing.
- To develop a Day-ahead time dependent pricing technique which maximizes the revenue of resource providers and minimizes the execution time and cost of the resources, thereby providing rewards/discounts to the consumers.
- To Develop a Compromised Cost Based Scheduling Technique which could minimize the execution time and cost of resources.
- To validate the developed model in Cloud environment.

3.4 Conclusion

Gap analysis always helps to find out the failings of different existing models. This chapter of thesis makes the foundation of proposed thesis. This chapter focuses on the needs of service providers and service consumers. Next chapter presents the proposed solution.

Chapter 4
Proposed Solution

The gaps and the various objectives of this thesis have been explained in previous chapter. This chapter presents the architecture and model of the proposed system. In this chapter the design of the proposed model will be discussed through UML diagrams. Also in this chapter the architecture of the Day-ahead Pricing Model using Time Dependent Pricing and CCTB Scheduling Technique has been proposed. Each component of the architecture has been explained in detail.

4.1 Design of Proposed Solution through UML

The Unified Modeling Language (UML) is a standard language used by software developers for writing software proposals. The advantage with using UML diagrams is that it provides a view of the entire system functionalities. The UML is a very expressive and useful language. This section of the thesis will explain the proposed solution using UML diagrams.

4.1.1 Use Case Diagram

Use case diagrams show the overall behavior of the proposed day-ahead time dependent pricing architecture, Use case diagrams show relationship among various actors and use cases. Figure 4.1 shows the pictorial representation of the system. Actors in the use case are Cloud Consumer, SLA_DB, CloudSim, Scheduler, QoS Metric_DB, Cloud Workload Analyzer, Workload_DB, Workload resource Pool. All the use cases and actions are connected to actors. Figure 4.1 shows the actual core of the day-ahead time dependent pricing model.

4.1.2 Sequence Diagram

A sequence diagram is like an interaction diagram that emphasizes the structural organization of the objects which send and receive messages. Graphically a sequence diagram is a table that shows objects arranged along the X-axis and messages, ordering in increasing time along the Y-axis. Figure 4.2 shows the sequence diagram of the day-ahead time dependent pricing model.

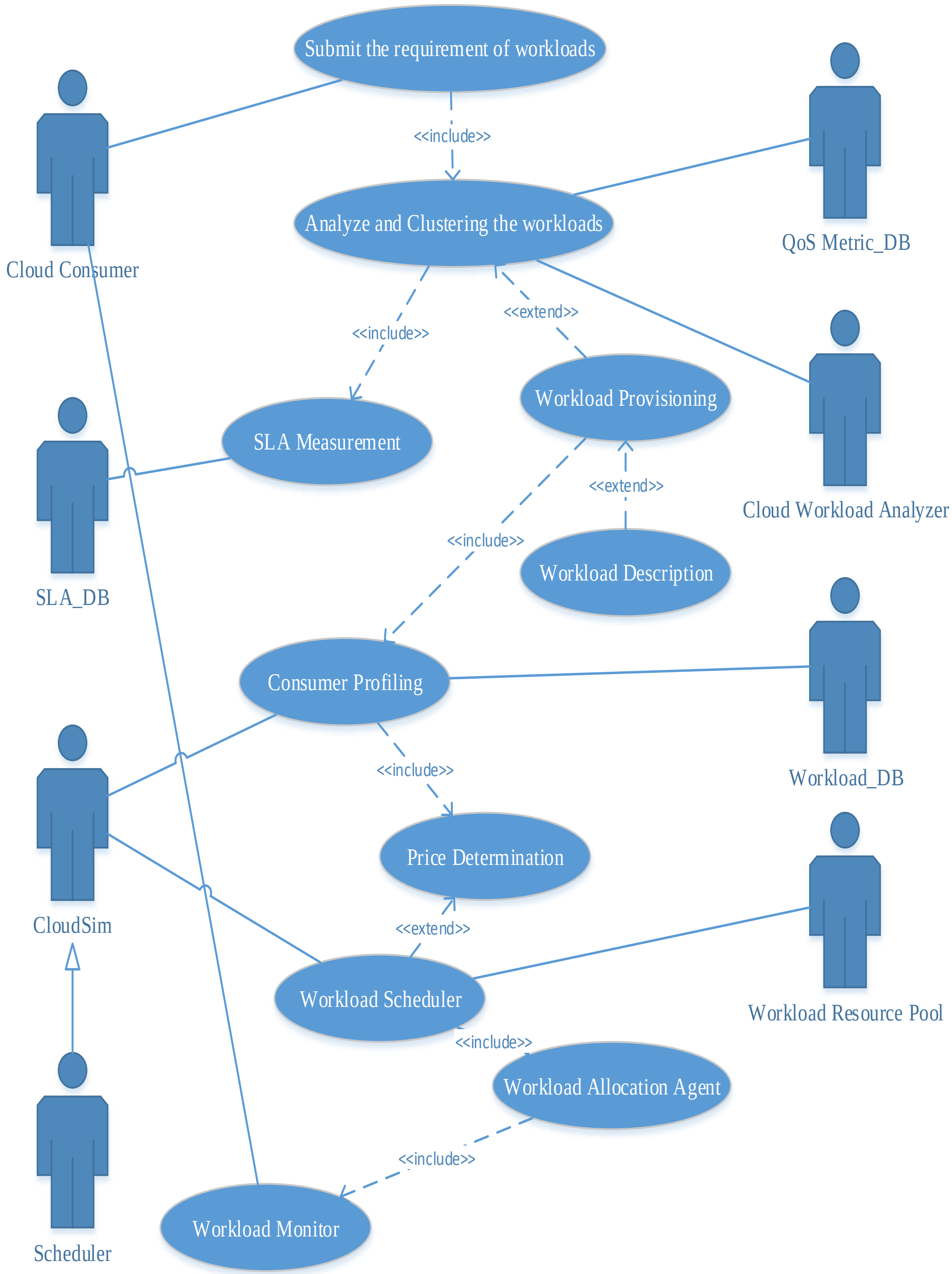


Figure 4.1: Use Case Diagram of Proposed Approach

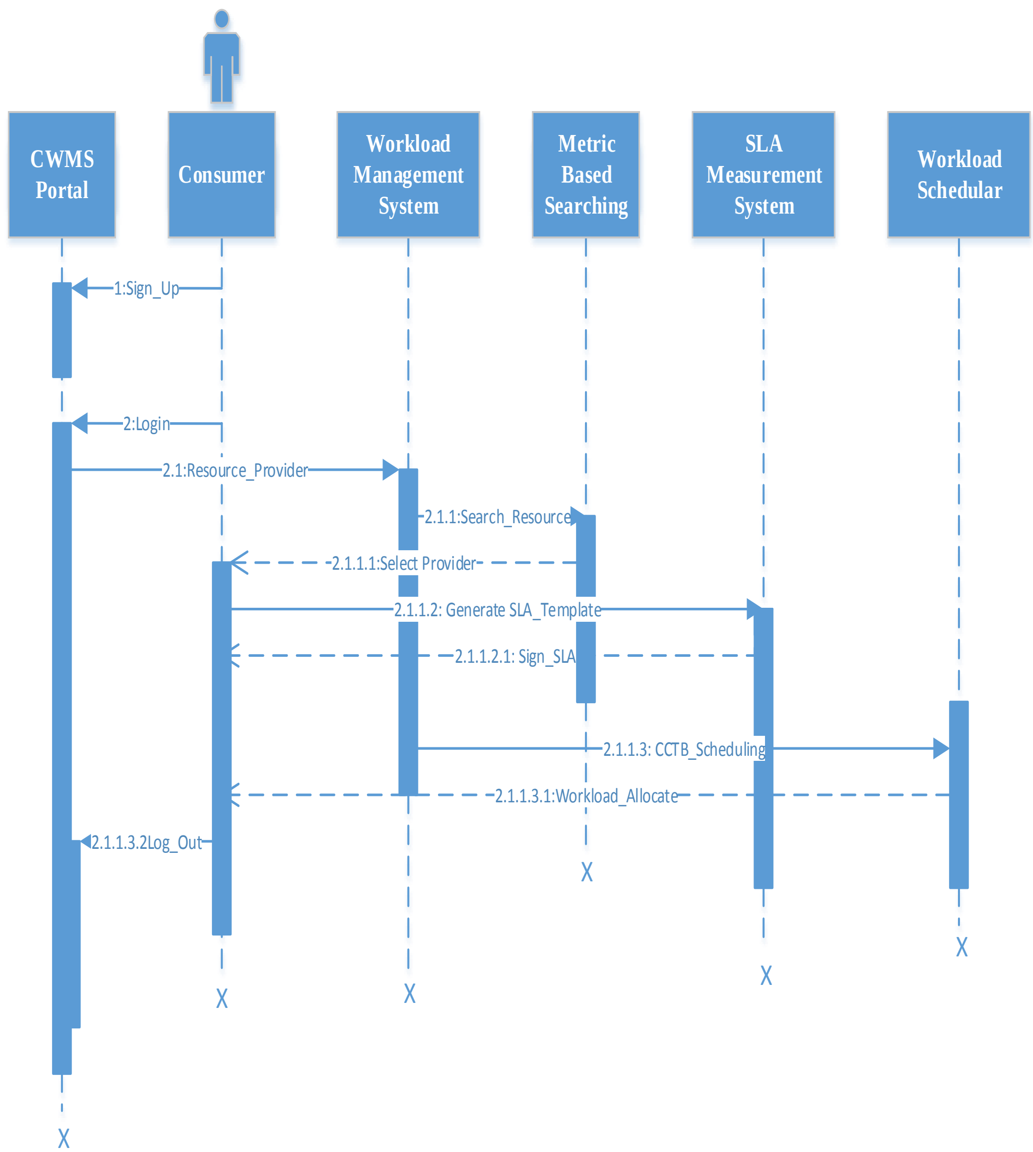


Figure 4.2: Sequence Diagram of Proposed Approach

4.2 Day-ahead Pricing Model for Smart Cloud

The basic motivation behind smart Cloud Computing is to constitute a framework which is automated and which can distribute Cloud resources over a network which uses two-way communication between service providers and consumers. In this model a Cloud Workload Management System (CWMS) has been used which acts as an interface between the consumer and the service provider (who manages the customer's resource requirements and clusters the Cloud resources using the k-mean clustering algorithm). This provides more support for demand response of resources using an economic model/technique called the Time Dependent Pricing (TDP) model. Smart Cloud Computing environment works on the principle of calculating prices of Cloud resources using TDP. The Cloud resource providers are required to send resource pricing information from their records (database) to Compromised Cost-Time Based (CCTB) scheduling policy is located at the CWMS interface. CCTB can observe and manage a consumers' resource requirement (automatically or manually) and schedule it at peak and off-peak periods. Such an approach helps resource providers to use a versatile pricing model to increase their profits/revenues by covering off-peak demands and executing better Cloud resource consumption.

4.3 Day-ahead Time Dependent Pricing and CCTB Scheduling Technique

In this architecture, TDP which is a versatile economic model acts as a feedback loop between the providers and the consumers as shown in Figure 4.3. The resource provider observes the Cloud workload to evaluate the amount of demand that can be shifted; the consumers' response to these prices is scheduled by CCTB. This helps the provider to refine the estimate to maximize revenue. The architecture of the proposed model is as shown in the Figure 4.4. This architecture focuses on the providers cost minimization problem and to improve the consumer satisfaction. In this systematic feedback loop the architecture contains three layers, namely: consumer response, Cloud Workload Management System and provider server.

Smart Cloud allow every consumer and Cloud service provider a two-way communication, thus enabling the Cloud service provider to generate a time dependent pricing technique using a feedback loop. This model charges a consumer more during peak periods and lesser during off peak periods, which encourages consumers to reschedule their workload to lesser traffic (off-peak) periods. This helps service providers to practice a versatile pricing technique to increase their profits by covering off-peak demand and minimizing the provider's cost optimization problem. It also minimizes the execution time in setting these prices by Compromised Cost-Time Based (CCTB) scheduling. Shifting workload is a probabilistic function which tells consumers to shift their workload.

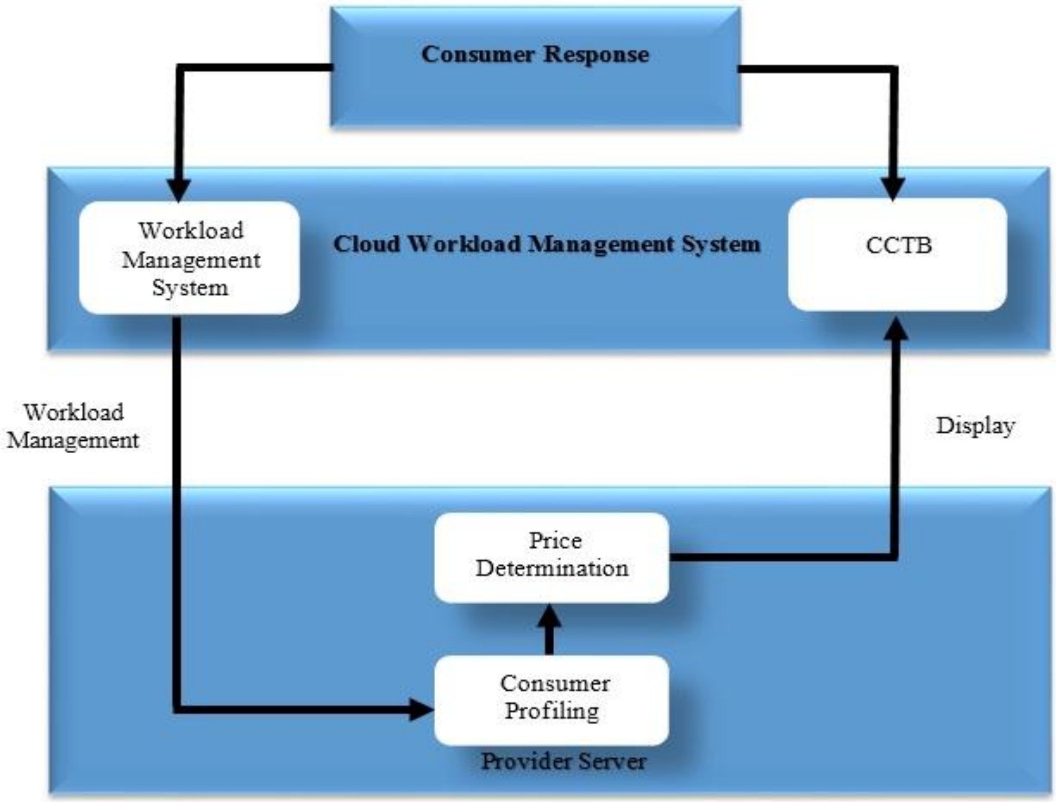


Figure 4.3: Schematic Behavior of Feedback Loop between the Resource Provider and Cloud Consumer

4.3.1 Cloud Consumer

This layer of architecture shows how consumer interacts with the model. This is a two way communication model where the consumer interacts with the models and models generate a response for the consumer.

4.3.2 Cloud Workload Management System

CWMS is an interface between the Service Providers and Resource Consumers. First the consumers interact with Workload Management System (WMS) in the environment of CWMS.

4.3.2.1 Workload Management System

The main task of the workload management system is that the Cloud service provider provides the workloads associated with its smart Cloud Computing environment to its requested consumer on the Cloud Workload Management Portal.

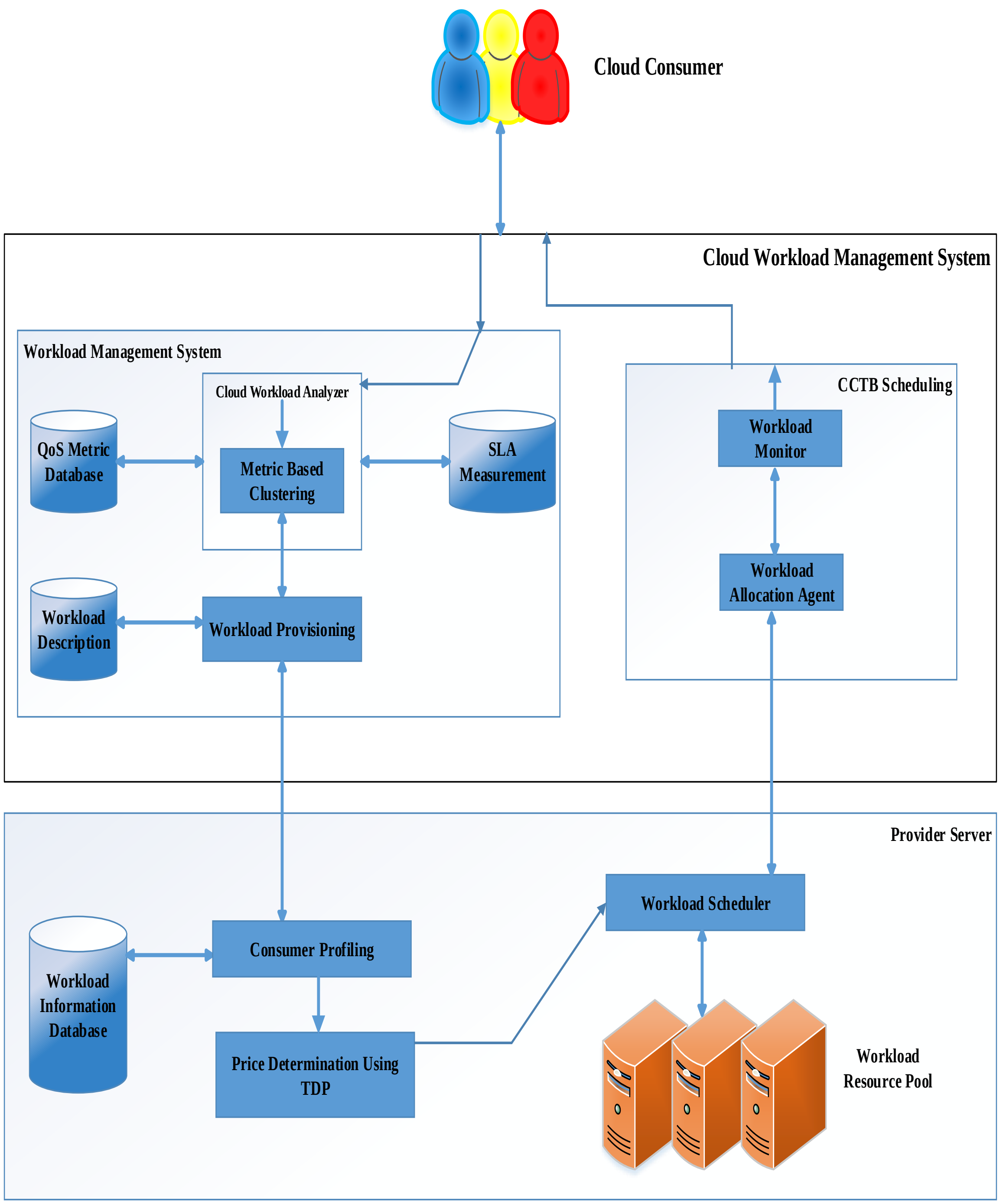


Figure 4.4: Cloud Workload Management System using Time Dependent Pricing and CCTB Scheduling

CWMS uses the K-means algorithm for clustering the various workloads specified in Table 4.1. Each workload has its own demand on the basis of QoS requirements which cannot be assigned on the basis of a simple static provisioning technique. Hence a k-means clustering algorithm which is proposed in [55] is used to cluster the associated workloads for this day-ahead pricing technique. “K-means” is a non-hierarchical algorithm and in its initial step creates a number of clusters equivalent to the number of population of workloads.

Table 4.1: Cloud Workload Categorization in WMS [55]

Workload Category	**Cloud Workloads in CWMS**	**QoS Parameters-Weights**
Server Oriented Workload	Hosting websites (W1)	Storage efficiency-3 Network bandwidth-3 Availability-5
	Technological Cloud Computing (W2)	Capacity of computing-5
	Endeavor software (W3)	Security- 5 Availability-5 Confidence level of customer-3 Correctness-3
	Computing performance testing (W4)	Capacity of computing-5
	online transaction oriented application (W5)	Security- 5 Availability-3 Accessibility of internet-5 Usability-3
	e-commerce (W6)	Load of computing-5 Customizability-3
	Online financial services (W7)	Security- 5 Availability-3 Changeability-1 Integrity-5
	Backup and storage services (W8)	Reliability- 5 Persistence-3
Client Oriented Workload	Production application for Cloud (W9)	Network bandwidth-2 Latency-3 Backup of data-4 Security-5
	Project development and testing (W10)	Self-service rate-4 Flexibility-4 Infrastructure service-1 Testing time-5
	Graphics project (W11)	Network bandwidth-3 Latency-3 Visibility-4
	Critical applications (W12)	Availability-5 Usability-3 Serviceability-4
Mobile Oriented Workload	Mobile services (W13)	High Availability-3 Reliability-5 Portability-2

Then it assigns weights on the basis of QoS associated with that workload on a scale from 1 to 5.The various types of workloads which are recognized by this thesis are hosting websites, technological Cloud Computing, endeavor software, computing performance testing, online transaction oriented application, e-commerce, online financial services, backup and storage services, production application for Cloud, project development and testing, graphics project, critical applications and mobile services [56]. Table 4.1 summarizes these Cloud workloads. This table also gives a unique identification to each workload. Various Cloud workloads which a consumer can demand from the provider are categorized on the basis of various QoS metrics. Metrics can be security related, bandwidth need, computing capacity etc. Time dependent pricing is applied on the listed workloads in Table 4.1. These workloads are listed in three categories on the basis of QoS requirements. The three categories are: server oriented workloads, client oriented workloads and mobile oriented workloads.

The weights associated with each Cloud workload are shown in Table 4.1. Singh *et al.* enhanced these thirteen workloads associated with one or more QoS requirements from the listed twenty-four [55]. The weights associated with each workload QoS parameter is based on an average value assigned to that parameter in different research papers. WMS manages these workload resources efficiently using k-means clustering algorithm and the distance is calculated using the four values of seeds [57].The result is taken from the tables given in [55]. The cluster formation is calculated using two iterations and these thirteen workloads are classified into four seed clusters. After two iterations the number of workloads remains same in each cluster. Singh et al. provided an efficient way of clustering the given. It also contains SLA measurement which connects with Cloud workload analyzer which measures any violation that may occur by the service provider while managing workloads.

4.3.2.2 CCTB Scheduling

All the workloads are scheduled for the consumer using CCTB [55]. Information regarding the workloads and its computational resources is sent to a workload allocation agent. The QoS factor is also represented in scheduling of each workload. The fundamental success behind using CCTB scheduling policy is that it focuses on minimizing the execution time and cost. The CCTB can schedule consumers' resources according to the negotiated cost and execution time because the prices are known to providers one day in advance. Most techniques do not heed to day-ahead pricing and also do not use TDP as a baseline for the economic model. The assumption of this work is that user will have the same amount of workload as he had before the introduction of the TDP technique. By the ease of TDP, the workload is shifted from one period to another and the workload is not lost. This is also valid when scheduling is done through CCTB. Subsequently, the workload is automatically made to schedule over a period on the basis of a compromised cost and time based scheduling [55].

4.3.2.3 Provider Server

The main task of provider server is to calculate the prices of resources and schedule the workload. A variable R is used to denote usage price that a Cloud service provider charges from its consumers without TDP. The TDP can be used in the terms of discount/reward to the usage price R. Discount (r_i) for every hour i, $i= 1, 2, \ldots, n$ where n=24. The day-ahead pricing technique assists consumers to shift their workloads to earlier or later period from the "actual period" of computing to get maximum rewards. The maximum reward across all n period is taken as R, because resource providers never charge negative prices from their customers.

There has been a trend of paradigm transition from the periods originally calculated by shifting workload function $p_j(r,t)$, which calculates the probability of shifting each consumer usage to a given time t, whereas r denotes the discount/reward in shifted period. The shifting workload function may vary from consumer to consumer. The subscript to shifting function j denotes consumer device parameterization. This function assumes to increase in discount (in terms of shifting the load with the law of diminishing marginal utility) and there has been a decrement in the time period (i.e. there is preference for shifting workloads to off-peak periods by the consumer). As shifting functions are probabilistic, it can normalize every shifting function over the entire days' time t if $p_j(R, t)$. Shifting of workload never exceeds the actual consumption ensured by normalization.

This thesis assumes each hour of the day as one unit of time and the representation of each hour period is done by i, where i equals to $1,2,3,4\ldots,n$. if a transition in consumption of the consumer occurs from hour i to hour k; where k can be greater than i or less than i. The total amount they shift is given by $k - i$ hours, where $k - i$ lies between [1, n] ϵq. It is equivalent to $k - i \equiv q \pmod{n}$. There is congruency between q and $k - i$ modulo n. Consumers transfer their workload to next day's hour k if $i > k$. With this the consumers gets the opportunity to shift their workload to k hours before or after hour i. There are some workloads which are never turned off. In this algorithm, it assumes a demand of baseline w_i in every hour i. This is the amount of workload which can never be shifted. With this an optimized problem of Cloud Computing service provider has been formulated.

Next, an expression of cost of discount/reward offered is formulated and later this will express the formulation for consumer satisfying cost. Discounts/rewards are given by r_1 ,r_2,........r_n in each hour over a day. First, calculate shifted demand of workload from each hour i into each hour k. Then the amount of workload needed by each consumer to be calculated in the hour i, then multiply it by shifting probability. The overall sum for each user is given by

$$\sum_{j \in W_i} p_j\left(r_{k,}|k - i|\right)u_j, \qquad (1)$$

W_i which includes all group of consumers workloads initially (i.e. without having Time Dependent Pricing) in hour i. u_j is the total workload demanded by consumers' device j (this refers to heterogeneity at the consumer end, a consumer can have more than one interface and can use more than one workload simultaneously). The total shifting of workload into an hour i is given by

$$\sum_{k \neq i} \sum_{j \in W_k} p_j\left(r_{i,}|k - i|\right)u_j. \qquad (2)$$

The sum of all cost (discounted/rewarded) offered is given below

$$\sum_{i=1}^{n} r_i \sum_{k\neq i} \sum_{j\in W_k} p_j\left(r_{i,}|k-i|\right)u_j + r_i w_i \qquad (3)$$

In this section a method to calculate the cost of satisfying consumer demand is introduced. Cloud service providers can categorize their servers into two categories: base-load servers and on-demand servers. The based-load servers have a low operating cost as compared to on-demand servers. Base-load servers have a higher capital cost than operational cost and these servers run continuously to serve the Cloud consumers need, whereas on-demand servers have a higher operational cost as compare to base-load servers. In any given hour, if a consumer needs workload and the base-load server is not able to fulfill the demand of consumers, then an on-demand server must fulfill the needs of the consumer.

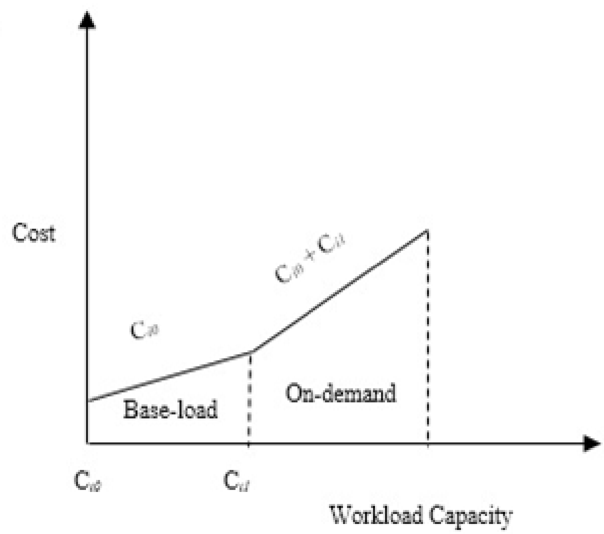


Figure 4.5: Linear Cost Structure of on-demand and Base-load Servers

Linear type of behavior for different operational cost (slopes) has been modeled in this algorithm for the cost of load servers. Usage charges and operational costs are taken into consideration for calculating these incremental costs. Operational cost is assumed to be uniform throughout the study while the usage cost is made to vary throughout.c_{i1} denotes the incremental cost of using an on-demand server without having base-load in hour i.

The random variables are a type of incremental cost; to optimize the provider's cost problem, random variables values can be solve exogenously. The linearity of cost structure of both servers' on-demand and base-load servers) is shown in Figure 4.5; c_{i0} shows the slope of base-load server workload handling cost.

Incremental cost and base-load workload handling capacity can vary from hour to hour. So C_{i0} and C_{i1} show the workload handling capacities in respective base-load and on-demand servers in hour *i*. The capacities taken here are the samples of random variables taken from price independent (i.e. exogenous) distribution. These instances are applied to the time-series algorithms like triple-exponential to calculate the capacities of base-load and on demand servers on the basis of historical data. This prediction can help to calculate the provider's cost minimization if consumers demand exceeding the on-demand and base-load servers.

The demand in each hour *i is given* below, using (1)

$$W_i - \sum_{k \neq i} \sum_{j \in W_i} p_j(r_{k,}|k-i|)u_j + \sum_{k \neq i} \sum_{j \in W_k} p_j(r_{i,}|k-i|)u_j. \qquad (4)$$

The cost of satisfying consumers' workload demand in respectively hour *i* is given by (5)

$$\sum_{z=1} c_{iz} \left[W_i - \sum_{k \neq i} \sum_{j \in W_i} p_j(r_{k,}|k-i|)u_j + \sum_{k \neq i} \sum_{j \in W_k} p_j(r_{i,}|k-i|)u_j - C_{iz} \right]^+ \qquad (5)$$

Where $[x]^+$ shows the maximum of 0 and x. Combining (3) and (5) and expression for providers cost minimization.

$$\min_{ri} \sum_{i=1}^{n} r_i \sum_{k \neq i} \sum_{j \in W_k} p_j(r_{i,}|k-i|)u_j + r_i w_i + \sum_{z=1} c_{iz} \left[W_i - \sum_{k \neq i} \left(\sum_{j \in W_i} p_j(r_{k,}|k-i|)u_j \right.\right.$$
$$\left.\left. + \sum_{j \in W_k} p_j(r_{i,}|k-i|)u_j \right) - C_{iz} \right]^+ \qquad (6)$$

$$s.t. r_i \geq 0, i = 1,2,3, \dots, n \qquad (7)$$

$$var.r_i, i = 1,2,3, \dots, n \qquad (8)$$

The equations (6-8) explain the service provider's optimization problem with large number of consumers and hours. As the shifting workload function and capacity vary frequently, the prices of service provider's cost minimization also vary. So, a dynamic pricing algorithm that adapts to these changes is needed.

4.3.2.4 Proposed Algorithm

The algorithm which uses provider's cost optimization problem is given below:

Algorithm 1: Calculating Day-Ahead Pricing- TDP

---

1. Clustered every workload using k-means algorithm.
2. Calculate the capacity of servers for next n hours.
3. Solve steps (6-8), using the capacities computed in the earlier step to get optimal discount/reward.
4. **while** the Cloud resource providers' implements TDP
   do
5. Evaluate for the upcoming n hours the capacities of servers.
6. Calculate steps (6-8) for estimating the discount for n hours in advance with the help of servers' workload predicted and earlier calculated rewards/discount for the following *n* -1 hours.
7. **if** it is hour *n* **then**
8. Compute shifting workload function method by equation (10).
9. **end if**
10. **end while**

By the help of this model, the different service provider sets up the time-dependent pricing by predicting every hour's capacity over the next day's hour and prices. In the initial stage, the provider sets the TDP (Time Dependent Pricing) for hour 1 of the first day. The discount (reward and prices) on each day for each hour 1 to *n*, is computed prior to the first hour. On the first day, at the end of hour 1, the provider calculates the capacities of hour one on the next day. Calculations of the discount of hour 1 for day 2 are done by enhancing the expected cost from hour 2 of day one to hour 2 of day 2. These calculations are done for as long as TDP runs for calculating the day-ahead discount.

4.3.2.5 Shifting Workload Function Estimation

This function is used in algorithm, for applying this pricing algorithm, the provider needs to calculate consumers' shifting workload functions. This section uses a method to calculate the same with a function using best-fit curves. Only aggregate data has been used to calculating this function; it doesn't track every individual usage in real time.

For calculation of shifting workload function, a function has been used, in which consumers discount/reward amount increases but time decreases exponentially as:

$$p_\gamma(r,t) = C_\gamma \frac{r}{(t+1)^\gamma}, \qquad (9)$$

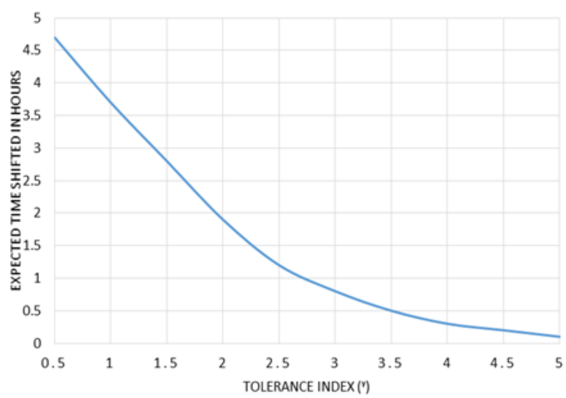


Figure 4.6: Average Shifting Hours over Several Different Values γ under a Flat Rate Pricing

C_γ, is a normalization factor constant. γhas been used as tolerance index, it calculates consumers' willingness to shift its workload usage to after or before its original operational hour. γ is a changeable parameter, higher value γ indicates more intolerant consumer, the consumer which is less prepared to move his workload for extended hours of the day. Figure 4.6 gives a clear understanding of different values of γ with an average period of shifting workload in a flat-rate pricing technique in the absence of TDP. In this algorithm the main focus that shifting workload function is simply a mathematical expression to its consumer performance, so this can be well calculated with functions like equation (9).

This estimation algorithm is used to predict the individual γ values for various different consumers. For simplicity, all values of γ are included for a given hour into single *collective shifting workload function*, by adding the all shifting workload functions for each instance in the given hour, subjected to workload consumption by each instance (device), in hour *I*, *the collective shifting workload function* is:

$$p_i(r,t) = \sum_{j \in W_i} p_j(r,t)\delta_j. \qquad (10)$$

Where, δ_j denotes the workload usage because of instance *j*. From hour *i* to hour *k*, the quantity of workload consumption delayed is noted as:

$$Q_{i,k} = p_i(r_k, |k-i|)(W_i - w_i)$$

And also for a given set of rewards and hour shifted time, the shifting workload function depends completely on the variables, γ_j& $\delta_j$where γ_j is tolerance index. In this similar tolerance index parameters are located together. As in (10) the similar index has become one which is $p_j(r,t)\delta_j$. Note that $W_i - w_i$ is the usage demand shifted from hour *i*. Let S_i show the difference between the usage demand W_i without using TDP and the usage demand using TDP. Each S_i can be expressed with collective shifting workload function:

$$S_i = \sum_{k \neq i} Q_{k,i} - Q_{i,k}$$

Each S_i, where *i* 1,2,3,….n could be stated as a $\frac{n(n-1)}{2}$ linear functions of S_i. Starting from S_1 , S_2 to S_{n-1} these n linear equations can be calculated by $Q_{1,2}$, $Q_{2,3}$,$Q_{3,4}$,……, $Q_{n-1,n}$. For every, *i* there is a value of *k* as stated in this calculation. This linear function is calculated for every day at the providers end and stored in the dataset. Now the best fit algorithm is applied to calculate the γ_j and δ_j which determine $Q_{i,k}$ and give the shifted workload function. Non-linear least square method is used as the best fit algorithm in these calculations. This algorithm can run offline, rather than in real time.

4.4 Conclusion

Day-ahead time dependent pricing model has been proposed in this chapter, which minimizes the providers' cost minimization problem, minimize the execution time, increase the revenue, provides rewards and discounts to the consumers. Next chapter verifies the proposed algorithm through experimental results.

Chapter 5

Implementation & Experimental Results

This chapter of the thesis focuses on the experimental setup, tools used to setup the environment, and implementation of proposed architecture. Lastly the experimental results of proposed architecture and comparison of the results with other existed techniques. Results show that how this technique improves the execution time and minimizes the cost of resources.

5.1 Tools for Setting Cloud Environment

Cloud applications have different compositions, configuration and deployment requirements. The tools being used to build the proposed Cloud architecture are explained below.

5.1.1 CloudSim

CloudSim is a new, generalized and an extensible simulation toolkit enabling modeling, simulation and experimentation of Cloud Computing systems and application services. Researchers and industry based developers use CloudSim to focus upon specific issues related to system design without any concern regarding low level details of Cloud based infrastructure and services. It supports both behavior and system modeling of components of Cloud system like virtual machines, datacenters and resource provisioning. Generic application provisioning techniques can be easily extended and implemented using CloudSim. Novel features offered by CloudSim are:

- Modeling and simulation support for Cloud Computing infrastructure, including datacenters existing on single physical computing node.
- Self-contained platform for modeling datacenters, scheduling, service brokers and allocation policies.
- Support for modeling and simulation of energy aware computational resources.

- Support for modeling and simulation of federated Clouds.
- Allows dynamic insertion of simulation elements.
- Support for modelling and simulation of massage passing applications and data center network topologies.
- Support for user defined policies for host allocation to virtual machines.

Multi-layered design and architectural components are shown in Figure 5.1. The simulation layer of CloudSim supports modeling and simulation of virtualized data center environment on Cloud that includes management interfaces for memory, virtual machine, bandwidth and storage.

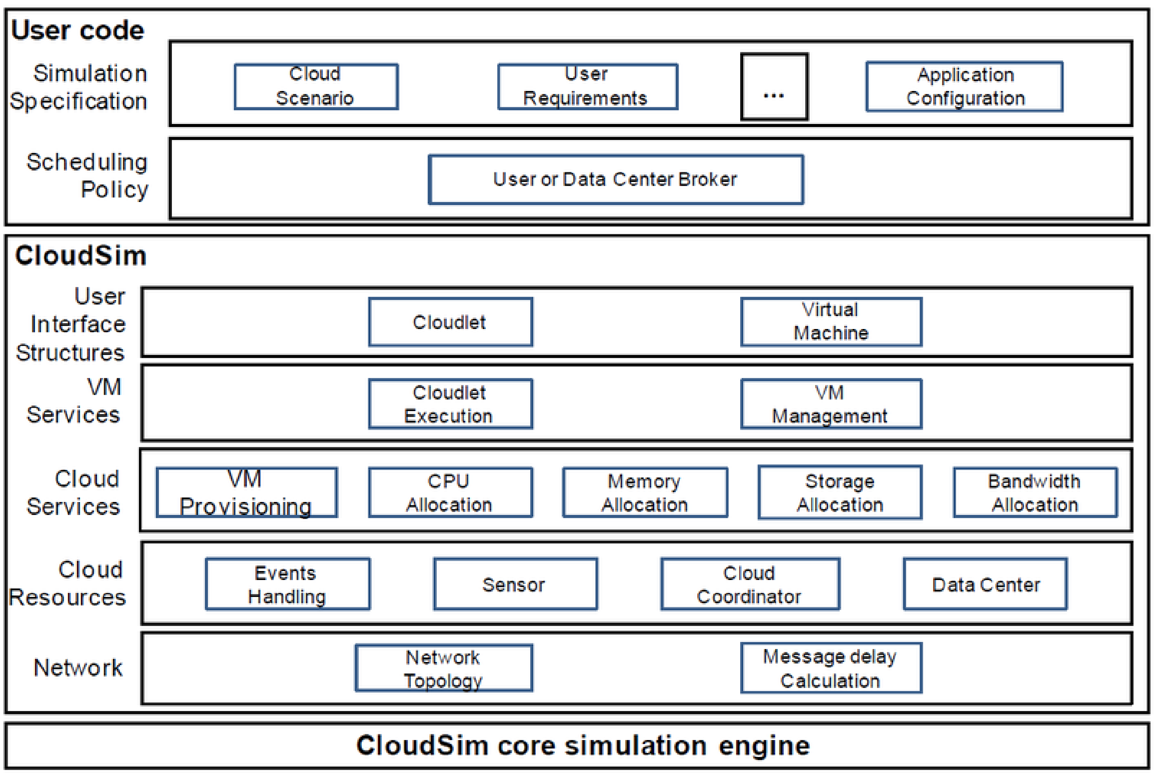


Figure 5.1: Layered CloudSim Architecture [58]

5.1.2 MS Visual Studio

Microsoft Visual Studio is an IDE from Microsoft. This tool is used to build the interface between service provider and service consumer. Visual studio helps developer to choose from variety of programing languages. Microsoft Visual Studio 2010 is used for developing interface between service provider and service consumers.

5.2 Implementation of Proposed Model

This section of this chapter deals the interface and implementation aspect of the proposed architecture.

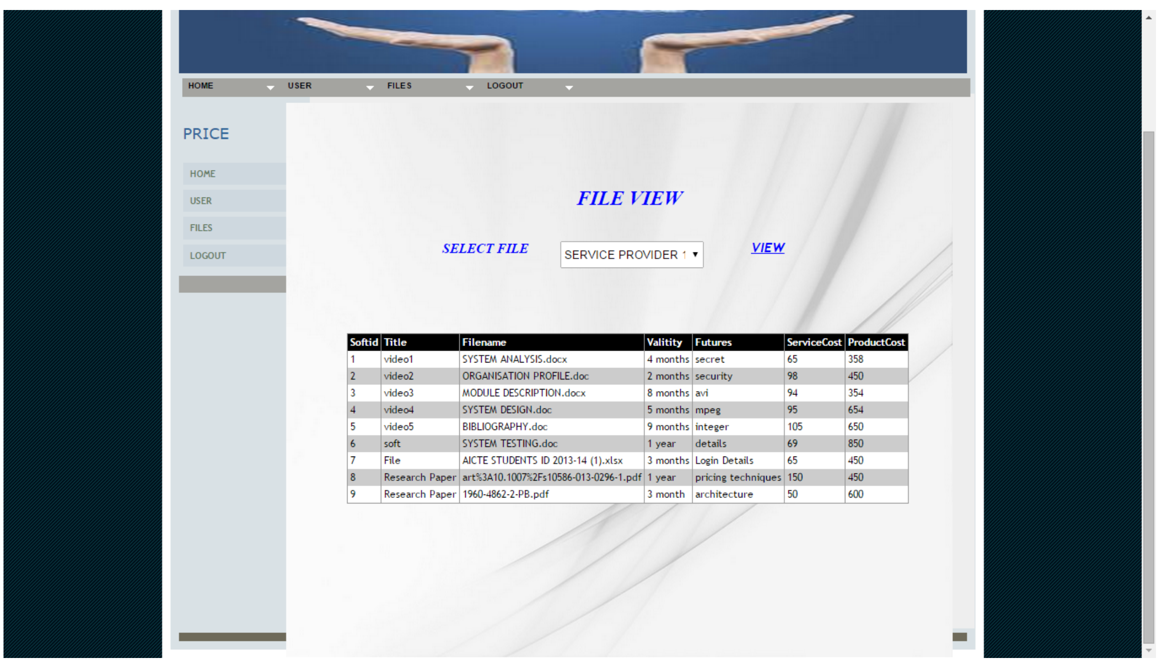
5.2.1 Interface

Cloud Workload Management System (CWMS) acts as an interface between the consumer and the service provider which manages the customer's resource requirements and clusters the Cloud resources using the k-mean clustering algorithm. This leads to more support for demand response of resources using an economic model called the Time Dependent Pricing (TDP) model. Smart Cloud Computing environment works on calculating prices of Cloud resources using TDP. The Cloud resource providers are required to send resource pricing information from their records (database) to Compromised Cost-Time Based (CCTB) scheduling policy located at the CWMS interface.

Figure 5.2: CWMS Portal Service Provider Login Page

Service Provider can login to portal and can set different workloads to be available for the smart Cloud. The various resources available in this portal are visible to various consumers who are looking for a resource. Payment System is received after setting the TDP and after that scheduler schedule the resources.

Figure 5.3: Service Provider Setting up the Prices of Resources



Softid	Title	Filename	Valitity	Futures	ServiceCost	ProductCost
1	video1	SYSTEM ANALYSIS.docx	4 months	secret	65	358
2	video2	ORGANISATION PROFILE.doc	2 months	security	98	450
3	video3	MODULE DESCRIPTION.docx	8 months	avi	94	354
4	video4	SYSTEM DESIGN.doc	5 months	mpeg	95	654
5	video5	BIBLIOGRAPHY.doc	9 months	integer	105	650
6	soft	SYSTEM TESTING.doc	1 year	details	69	850
7	File	AICTE STUDENTS ID 2013-14 (1).xlsx	3 months	Login Details	65	450
8	Research Paper	art%3A10.1007%2Fs10586-013-0296-1.pdf	1 year	pricing techniques	150	450
9	Research Paper	1960-4862-2-PB.pdf	3 month	architecture	50	600

Figure 5.4: Various Resources Availability by Different Providers

Figure 5.5: Consumer Registration to Smart Cloud

Figure 5.6: Consumers' Payment Gateway for Scheduled Resources

5.2.2 Experimental Results

The simulation of this model in CloudSim is done as Cloudlets. More than 3000 Cloud workloads were used for these calculations [58]. While studying the feasibility of this model a calculation for the execution time of workloads with or without using TDP is prepared. Along with that a comparison of CCTB with other existing scheduling policies will be done. Table IV shows the characteristics of resources that are used in the experimental setup. The performance of this algorithm is done with around 3000 workloads and 50-250 Cloud Computing resources.

Table 5.1: CloudSim Parameters and their Values

Parameters	Values
Number of resources	50–250
Number of Cloudlets (workloads)	3,000
Bandwidth	1,000–3,000 B/S
Size of Cloud workload	10,000+ (10–30%) MB
Number of PEs per machine	1
PE ratings	100-4,000 MIPS
Cost per Cloud workload	$3-$5
Memory size	2,048–12,576 MB

To compare this scheduling approach with the existing scheduling algorithms: Compromised Time Cost (CTC) [59] and Deadline and Budget Distribution-Cost Time Optimization (DBD-CTO) [60] these steps are followed. The aggregate of shifting workload function is taken as a sum of function, as in (9), where γ = 0.5, 1, 1.5……, 5 (shifting parameters are chosen collectively).

Test Case 1: Execution time and Cost comparison of Cloud workloads with or without using TDP

The execution time for same Cloud workload and having same no. of resources is calculated by CloudSim, TDP took 1,321 seconds whereas the flat rate pricing took 1,468 seconds. The execution time in TDP is less than the time taken by flat rate pricing. Figure

5.7 shows the result of this evaluation. Figure 5.8 shows the cost structure of TDP and flat rate pricing. It cost $195 using TDP whereas $238using flat rate pricing. The results are calculated using the same number of resources and same workload.

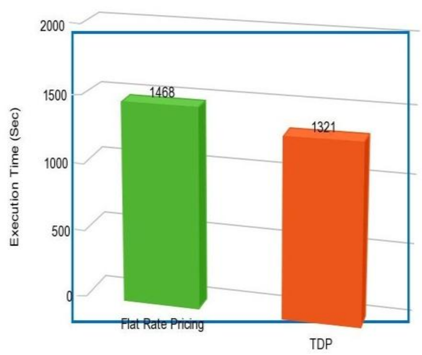


Figure 5.7: Execution Time Comparison of Same Workload and Same no. of Resources for Different Pricing Schemes.

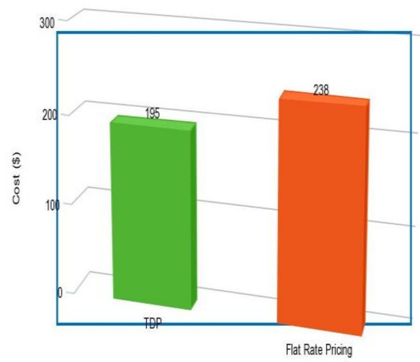


Figure 5.8: Cost Comparison of Same Workload and Same no. of Resources for Different Pricing Schemes.

Test Case 2: Execution Time and Cost comparison of Cloud workloads using TDP with different Scheduling policies

Figure 5.9 and Figure 5.10 show the result using different scheduling policies with TDP. Execution time and cost for the same number of workloads and same number of resources are calculated differently using scheduling policies. The result shows that CCTB improves the execution time and cost for different workloads. Each scheduling algorithm works with homogenous Cloud workloads. These scheduling policies are implemented in CloudSim using VMSchedular.java. Results show that CCTB scheduling policy combined with TDP improves the execution time and cost of Cloud Computing, which subsequently overcomes service providers' cost minimization problem and helps the provider to increase their revenue.

Test Case 3: Execution time for different number of resources and Cloud workloads using three scheduling policies:

Outcome of increasing the resources while keeping the workload constant in CWMS is shown in Figure 5.11.The outcome shows that as the resources increase the corresponding time of execution decreases. CCTB performs better than other two scheduling policies. Figure 5.11 shows that as the resources increase, execution time decreases in the same proportion in CTC and CCTB. The results of increasing the workload at execution time are shown in Figure 5.12. Approximately 3,000 workloads are used to show this result. As the outcome states that the execution time reduces using the CCTB policy as compared to other scheduling policies. The average reduction time is around 12-28%. The execution time generally increases with increase in number of workloads. The Figure 5.12 results show that CCTB is a better approach than the other two. Table 5.2 and Table 5.3 shows the outcome this results and the graph is plotted on the basis of these results.

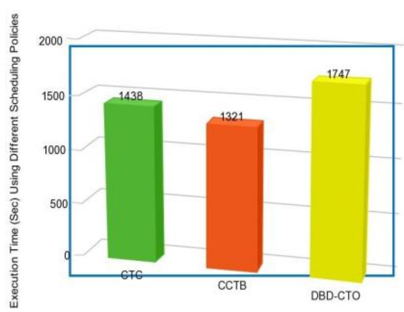


Figure 5.9: Execution Time Comparison of Same Workload and Same no. of Resources for Different Scheduling Policies.

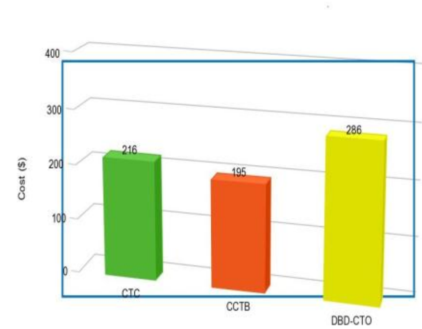


Figure 5.10: Cost Comparison of Same Workload and Same no. of Resources for Different Scheduling Policies.

Test Case 4: Cost for different number of resources and Cloud workloads using three scheduling policies

The cost for Cloud workload execution varies for every scheduling policy. Figure 5.13 shows the outcome that the cost of execution decreases with increase in the number of resources. Figure 5.13 illustrates that CCTB scheduling policy consumes less cost for the same number of resources as compared to other scheduling policies. Figure 5.14 shows that the cost per workload increases as the Cloud workload increases for a given constant number of resources. The existing scheduling policies seem to be expensive as compared to the CCTB scheduling policy. The overall cost for handling such amount of Cloud workload is less by using the CCTB scheduling policy as compared to CTC and DBO-CTO. CCTB performs better with increase in the number of workloads. Table 5.4 and

Table 5.5 shows the outcome this results and the graph is plotted on the basis of these results.

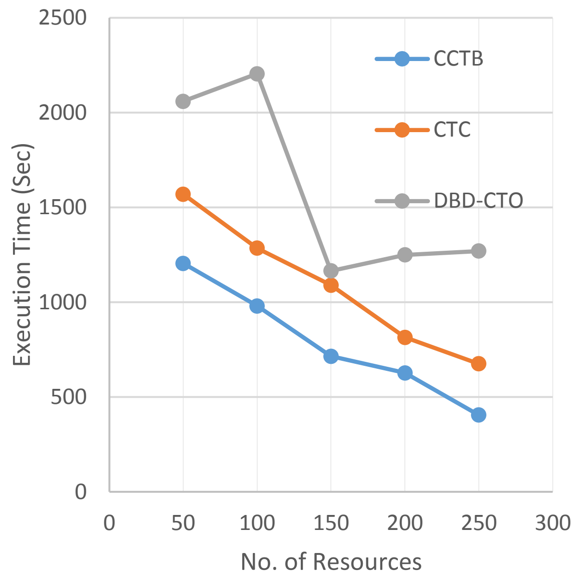


Figure 5.11: Execution Time Comparison of Same Number of Workload with Different Number of Resources on Different Scheduling Policies.

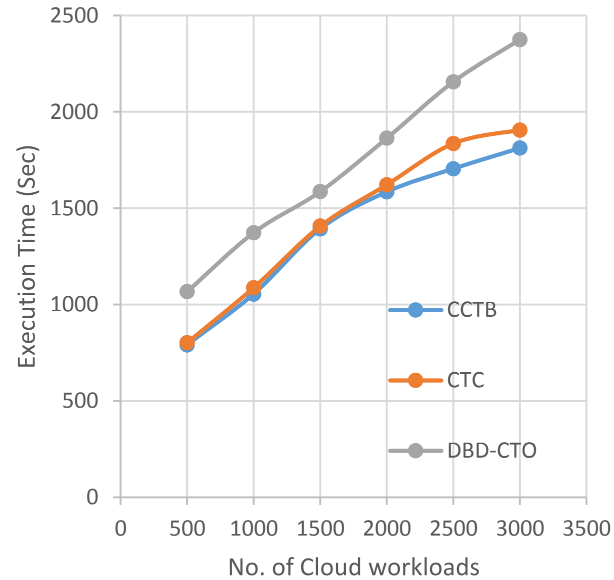


Figure 5.12: Execution Time Comparison of Same Number of Resources with Different Number of Workload on Different Scheduling Policies.

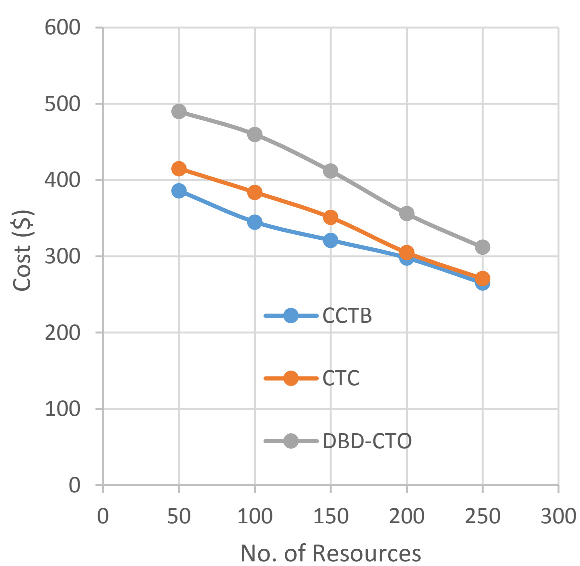


Figure 5.13: Cost Comparison of Same Number of Workload with Different Number of Resources on Different Scheduling Policies.

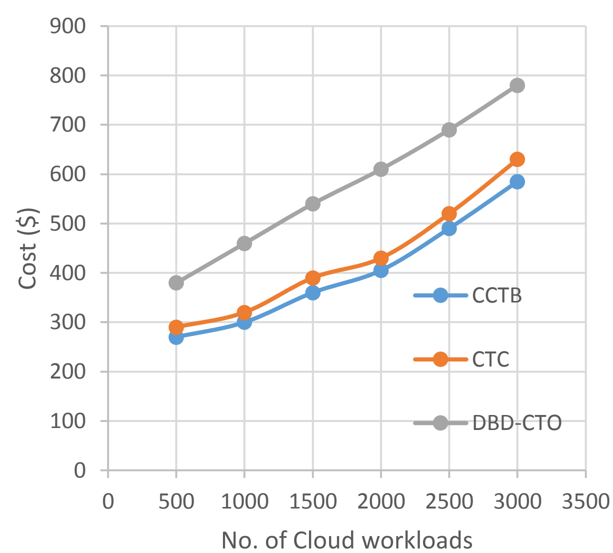


Figure 5.14: Cost Comparison of Same Number of Resources with Different Number of Workload on Different Scheduling Policies.

Table 5.2: Results of Various Resources w.r.t Execution Time Keeping Number of Cloud Workloads Constant

Tasks	**Execution Time(Sec)**		
No. of Resources	**CCTB**	**CTC**	**DBD-CTO**
50	1205	1570	2060
100	980	1285	2205
150	715	1090	1165
200	627	815	1250
250	405	675	1270

Table 5.3: Results of Various Cloud Workloads w.r.t Execution Time Keeping Number of Resources Constant

Tasks	**Execution Time(Sec)**		
No. of Cloud Workloads	**CCTB**	**CTC**	**DBD-CTO**
500	792	803	1068
1000	1056	1089	1374
1500	1394	1408	1587
2000	1586	1623	1864
2500	1706	1837	2157
3000	1813	1906	2375

Table 5.4: Results of Various Resources w.r.t Cost Keeping Number of Cloud Workloads Constant

Tasks	**Cost($)**		
No. of Resources	**CCTB**	**CTC**	**DBD-CTO**
50	386	415	490
100	345	384	460
150	321	351	412
200	298	305	356
250	265	271	312

Table 5.5: Results of Various Cloud Workloads w.r.t Cost Keeping Number of Resources Constant

Tasks	**Cost($)**		
No. of Cloud Workloads	**CCTB**	**CTC**	**DBD-CTO**
500	270	290	380
1000	300	320	460
1500	360	390	540
2000	405	430	610
2500	490	520	690
3000	585	630	780

5.3 Conclusion

Simulation results of day-ahead time dependent pricing using CCTB scheduling prove that this technique is efficient than other existing model. Comparative analysis shows that this models helps provider to maximize their revenue. Setup tools for this architecture are also discussed in this chapter. Next Chapter concludes the thesis.

Chapter 6

Conclusion and Future Scope

This chapter concludes the research work presented in this thesis. At the end of this chapter future research directions have been suggested which can be used to improve the providers' cost minimization problem and can improve the scheduling approaches used.

6.1 Conclusion

This thesis works on the aspects of Smart Cloud Computing, which is designed, developed and simulated in this thesis. It also introduces an economic model using time dependent pricing with a day ahead constraint. The smart Cloud schedules their resources using compromised cost and time scheduling for calculating the prices. This model shows the consumers' willingness to transfer their workloads to lower price off-peak hours. This thesis expresses the need of providers' cost minimization problem. The architecture of this framework is a feedback loop between consumers and service providers with workload management system and real time scheduling using CCTB. Results show that improved prices help service provider to increase their revenue.

6.2 Thesis Contributions

Some of thesis contributions are:

- Extensive literature review of various existing pricing and economic models of Cloud and Grid Computing has been presented.
- Developed a feedback loop architecture to solve the problem of providers' cost minimization problem.
- A compromised cost and time based scheduling technique is used to improve the cost and execution time.
- This proposed model tries to increase the providers' revenue and also provide QoS to its consumers by offering rewards and discounts as well.

- Designed architecture has been implemented using software such as CloudSim, Visual Studio.

6.3 Future Scope

This study opens up the different areas in Cloud like scheduling, economic model, automated scheduling. This work can be further extended as:

- The day ahead concept can lead towards hour ahead pricing in future.
- The consumer can choose from a variety of scheduling techniques instead of using provider specific scheduling techniques which is CCTB in this thesis.
- In future, the proposed architecture can automate the scheduling with variety of new scheduling algorithms.
- The concept of on-demand resource scheduling in Smart Cloud can be done in the future.

References

[1] M. H. Weik, (1961). The ENIAC Story [Online]. Available: http://ftp.arl.mil/mike/comphist/eniac-story.html

[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg and I. Brandic , "Cloud computing and emerging IT platforms: Vision, hype and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, Vol. 25, Issue 6, pp.599-616, 2009.

[3] G.F. Pfister, *In Search of Clusters*, 2nd ed., Prentice Hall, Upper Saddle River, USA, 1998.

[4] I. Foster, Y. Zhao, and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," in *Grid Computing Environments Workshop (GCE '08)*, 2008.

[5] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, "Above the Clouds: A Berkeley View of Cloud computing," EECS Department, University of California, Berkley, Technical Report No. UCB/EECS-2009-28, February 2009.

[6] M. G. Avram (Olaru), "Advantages and challenges of adopting cloud computing from an enterprise perspective," in *7th International Conference Interdisciplinary in Engineering, INTER-ENG 2013*, vol. 12, 2013, pp.529-534.

[7] N. Sadashiv and D. Kumar, "Cluster, Grid and Cloud Computing: A Detailed Comparison," in *the 6th International Confrence on Computer Science & Education (ICSE 2011)*, SuperStar Virgo, Singapore, 2011, pp. 477-482.

[8] S. Kumar, *Cloud Computing- insight into New-Era Infrastructure*, 1st ed. India: Wiley India, 2011.

[9] R. Buyaa, C.Vecchiola, and T. Selvi, *Mastering Cloud Computing*, 1st ed. New Delhi, India: McGraw Hill Education (India) Private Limited, 2013.

[10] Google App Engine Pricing [Online]. Available: https://cloud.google.com/appengine/pricing

[11] Amazon Web Service Pricing [Online]. Available: http://aws.amazon.com/pricing

[12] Azure Pricing [Online]. Available: http://azure.microsoft.com/en-us/pricing

[13] A. Kaur and D.Singh, "Enlightening the Cloud Computing Domain," *International Journal of Computers &Technology*, vol. 4, no.2, 2013.

[14] M. Creeger, "Cloud computing: An overview", *ACM Queue*, vol. 7, no.5, June, 2009.

[15] Types of Cloud Computing [Online]. Available: http://www.appcore.com/types-cloud-computing-private-public-hybrid-clouds/

[16] K. Saurab, *Cloud Computing*, 1st ed. India: Wiey India, 2012.

[17] S. Singh and I. Chana, "Cloud Based Development Issues: A Methodical Analysis," *International Journal of Cloud Computing and Services Science (IJ-CLOSER)*, vol. 2, no. 1, pp. 73-84, 2012.

[18] P. Samimi, and A. Patel, "Review of pricing models for grid & cloud computing," in *ISCI 2011 - 2011 IEEE Symposium on Computers and Informatics*, 2011, pp. 634-639.

[19] vThunder Pay-as-you-Go Licensing Model [Online]. Available: https://www.a10networks.com/products/thunder-series-appliances/vthunder/vthunder-pay-you-go

[20] L.J. Zhang and Qun Zhou, "CCOA: Cloud Computing Open Architecture," in *7th IEEE International Conference on Web Services*, 2009, pp.607-616.

[21] B. Sharma, R. K. Thulasiram, P. Thulasiraman, S. K. Garg, and R. Buyya, "Pricing Cloud Compute Commodities: A Novel Financial Economic Model," in *12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE Computer Society*, 2012.

[22] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654 1973.

[23] R. Merton, "Theory of rational option pricing," *Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.

[24] G. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, 1965.

[25] H. Li, J. Liu, G. Tang," A Pricing Algorithm for Cloud Computing Resources," in *International Conference on Network Computing and Information Security IEEE*, 2011.

[26] R. Wolski, J. S. Plank, T. Bryan, and J. Brevik, “G-commerce: Market formulations controlling resource allocation on the computational grid,” in *15th IEEE Int. Parallel Distributed Processing Symposium*, 2001.

[27] M. Mihailescu and Y. M. Teo, “Dynamic resource pricing on federated clouds,” in *IEEE International Symposium on Cluster Computing and the Grid. Los Alamitos, USA: IEEE Computer Society*, 2010, pp. 513–517.

[28] A. Rowstron and P. Druschel, “Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems,” in *18th IFIP/ACM Int. Conf. Distributed Systems Platforms (Middleware 2001)*, 2001, pp. 329–350.

[29] M. Mac´ıas and J. Guitart, “A genetic model for pricing in cloud computing markets,” in *26th Symposium. Applied Computing*, 2011.

[30] S. Wee, “Debunking real-time pricing in cloud computing,” in*IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, 2011.

[31] V. N. Dhawas,P. Juikar,N. Patekar,N. Lendghar, and S. Vartak,“A Secured Cost Effective Multi-Cloud Storage in Cloud Computing,” *International Journal of Scientific & Engineering Research*, vol. 4, Issue 5, pp. 1017-1021, 2013.

[32] S. A. Bello, and G. A. Wakil, “Flexible Pricing Models for Cloud Services,” *Transactions on Networks and Communications*, vol. 2, no. 5, pp. 15-28, 2014.

[33] M Arrington, “Gmail Disaster:Reports of mass email deletions,” [Online] Available: http://www.techcrunch.com/2006/12/28/gmail-isasterreport-ofmass-email deletion/, 2006.

[34] R. Buyya, D. Abramson, J. Giddy, and H. Stockinger, "Economic models for resource management and scheduling in grid computing" *J. Concurrency Computation: Practice Experience (CCPE)*, vol. 14, no. 13, pp. 1507- 1542, 2002.

[35] W.-Y. Lin, G.-Y. Lin, and H.-Y. Wei, “Dynamic auction mechanism for cloud resource allocation,” in *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, 2010, pp. 591–592.

[36] I. Fujiwara, K. Aida, and I. Ono, “Applying Double-sided Combinational Auctions to Resource Allocation in Cloud Computing”, in *10th Annual International Symposium on Applications and the Internet, IEEE*, 2010, pp. 7-14, 2010.

[37] Shang, S., J. Jiang, Y. Wu, Z. Huang, G. Yang, and W. Zheng, “DABGPM: A Double Auction Bayesian Game-Based Pricing Model in Cloud Market,” Network and Parallel Computing, pp. 155-164, 2010.

[38] J. Yang, S. Yang, M. Li, and Q. Fu, "An Autonomous Pricing Strategy toward Market Economy in Computational Grids," in *IEEE International Conference on Information Technology Coding and Computing (ITCC'05)*, Las Vegas, Nevada, 2005, pp. 793-794.

[39] B. Pourebrahimi, K. Bertels, G. Kandru, and S. Vassiliadis, “Market-based resource allocation in grid,” in *2nd IEEE International conference on eScience and grid Computing*, 2006.

[40] M. Schwind, O. Gujo, and T. Stockheim, “Dynamic resource prices in a combinatorial grid system,” in *IEEE Joint Conference on E-Commerce Technology (CEC’06) and Enterprise Computing, E-Commerce and E-Services (EEE’06),* San Francisco, CA, 2006.

[41] X. Zhao, L. Xu, and B. Wang, "A Dynamic Price Model with Demand Prediction and Task Classification in Grid," in *6th international conference on grid and cooperative computing*, 2007, pp.775-782.

[42] H. Sato, Y. Koyama, K. Kurumatani, Y. Shiozawa, and H. Deguchi, "U-Mart: A Test Bed for Interdisciplinary Research in Agent Based Artificial Market", *Evolutionary Controversies in Economics*, pp. 179-190, 2001.

[43] A. Quiroz, H. Kim, M. Parashar, N. Gnanasambandam, and N. Sharma. Towards autonomic workload provisioning for enterprise grids and clouds. in *10th IEEE/ACM International Conference on Grid Computing (Grid 2009)*, 2009, pp. 50 – 57.

[44] H. N. Van and F. D. Tran, "Autonomic Virtual Resource Management for Service Hosting Platforms," in *ICES '09: Proceedings of the International Conference on Software Engineering Workshop on Software Engineering Challenges of Cloud Computing. IEEE Computer Society*, 2009, pp. 1-8.

[45] J. a. N. Silva, L. Veiga, and P. Ferreira, "Heuristic for resources allocation on utility computing infrastructures," in *6th International Workshop on Middleware for Grid*

Computing, ser. MGC '08, New York, NY, USA: ACM, 2008, pp. 1-9.

[46] K. Lu, R. Yahyapour, P. Wieder, C. Kotsokalis, E. Yaqub, A. I Jehangiri, and Y. Hamadi, "QoS-Based Resource Allocation Framework for Multi-Domain SLA Management in Clouds," *International Journal of Cloud Computing*, vol. 1, no. 1, 2013.

[47] L. Wu, S. Garg, and R. Buyya, "Sla-based resource allocation for software as a service provider (SaaS) in cloud computing environments," in *11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2011, pp. 195-204.

[48] Q. Li, Q. Hao, L. Xiao, and Z. Li, "Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control", in *1st International Conference on Information Science and Engineering (ICISE)*,2009, pp.99-102.

[49] T. C. Chieu, A. Mohindra, A. A. Karve, A. Segal, "Dynamic Scaling of Web Applications in a Virtualized Cloud Computing Environment", in *IEEE International Conference on e-Business Engineering (ICEBE 2009)*, 2009, pp.281-286.

[50] N. R. Herbst, N. Huber, S. Kounev and E. Amrehn, "Self-adaptive workload classification and forecasting for proactive resource provisioning," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 12, pp. 2053-2078, *2014*

[51] H. R, Qavami, S. Jamali, M. K. Akbari, and B. Javadi, "Dynamic Resource Provisioning in Cloud Computing: A Heuristic Markovian Approach," *Cloud Computing Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 133, pp 102-111, 2014.

[52] K. Raghavendra, A. Akilan, N. Ravi, K. P. Kumar and G. Varadan, "Satellite Data product Generation using Aneka Cloud," in *Research Demo at 10th IEEE International Symposium o Cluster, Cloud and Grid Computing (CCGrid 2010)*, Melbourne, Australia, 2010.

[53] Dropbox Inc. Dropbox. [Online] Available: http://www.dropbox.com/

[54] E. Guizzo, "Robots with their heads in the cloud," *IEEE Spectrum*, 2011.

[55] S. Singh, and I. Chana, "QRSF: QoS-aware resource scheduling framework in cloud computing," *The Journal of Supercomputing*, vol. 71, no. 1, pp. 241-292, 2015.

[56] S. Singh, and I. Chana, "Metrics based workload analysis technique for IaaS cloud," in *International Conference on Next Generation Computing and Communication Technologies*, Dubai, 23rd and24th April 2014.

[57] M. Su, and C. Chou, "A modified version of the K-means algorithm with a distance based on cluster symmetry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 674–680, 2001.

[58] G. Eason, B. R. N. Calheiros, R. Ranjan, C. A. F. D. Rose, and R. Buyya, "Cloudsim: A novel framework for modeling and simulation of cloud computing infrastructures and services," Computing Research Repository, vol. abs/0903.2525, 2009.

[59] K. Liu, H. Jin, J. Chen, X. Liu, D. Yuan and Y. Yang, "A Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-Intensive Cost-Constrained Workflows on Cloud Computing Platform," *International Journal of High Performance Computing Applications Sage*, pp.445-456, 2010.

[60] A. Verma, and S. Kaushal, "Deadline and budget distribution based cost-time optimization workflow scheduling algorithm for cloud," in *International conference on recent advances and future trends in information technology (iRAFIT 2012)*, 2012.

List of Publications

[1] Chetan Chawla and Inderveer Chana, “Day-ahead Pricing Model for Smart Cloud using Time Dependent Pricing,” *International Journal of Computer Network and Information Security (IJCNIS)*, vol. 7, no. 9, 2015. [**Accepted and to be Published in August 2015**]

[2] Chetan Chawla and Inderveer Chana, “Optimal Time Dependent Pricing Model for Smart Cloud with Cost Based Scheduling,” in *3rd ACM International Symposium on Women in Computing and Informatics (WCI-2015)*, Kochi, Kerala, India, to be held in August 2015. [**Accepted**]

[3] Chetan Chawla and Inderveer Chana, "Strategy-proof Pricing Approach for Cloud Market," presented at *International Conference on Communication, Information and Computing Technology (ICCICT-15)*, Amritsar, Punjab, India, 2015, DOI= arXiv:1506.06648. [**Published**]

Video Presentation

https://www.youtube.com/watch?v=nnlF9-MB01M