

# **A Heterogeneous Workload Consolidation Technique for Green Cloud**

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

**Master of Engineering  
in  
Software Engineering**

*Submitted By*  
**Deep Mann**  
**(Roll No. 801031007)**

Under the supervision of:  
**Dr. Inderveer Chana**  
Associate Professor



COMPUTER SCIENCE AND ENGINEERING DEPARTMENT  
THAPAR UNIVERSITY  
PATIALA – 147004

**June 2012**

## Certificate

---

I hereby certify that the work which is being presented in the thesis entitled, "*A Heterogeneous Workload Consolidation Technique for Green 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.


  
(Deep Mann)

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

  
(Dr. Inderveer Chana)

Associate Professor,  
Computer Science and Engineering Department

Countersigned by

  
(Dr. Maninder Singh)  
Head  
Computer Science and Engineering Department  
Thapar University  
Patiala

  
(Dr. S. K. Mohapatra)  
Dean (Academic Affairs)  
Thapar University  
Patiala

## Acknowledgement

---

First of all, I am thankful to God for his blessings and showing me the right direction. With His mercy, it has been made possible for me to reach so far.

It gives me great pleasure to express my gratitude towards the guidance and help I have received from Dr. Inderveer Chana. I am thankful for her continual support, encouragement, and invaluable suggestion. She not only provided me help whenever needed, but also the resources required to complete this thesis report on time.

I am also thankful to Dr. Maninder Singh, Head, Computer Science and Engineering Department for his kind help and cooperation. I express my gratitude to all the staff members of Computer Science and Engineering Department for providing me all the facilities required for the completion of my thesis work.

I extend my thanks to Ms. Nidhi, Ms. Rajni and Ms. Pankajdeep Kaur for sharing their expertise and time to help me accomplish this work.

I would like to say thanks to all my friends especially Harnoorjit Kaur, Gurjot Singh, Mehak, Shveta, parul and Abhishek for their support. 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 highly grateful to all my family members for their inspiration and ever encouraging moral support, which enables me to pursue my studies.

**Deep Mann**

## Abstract

---

Cloud Computing offers service to end-users rather than a product, by sharing resources, software and other information under a usage based payment model. It enables hosting of various kinds of applications such as business, scientific, social network, etc. as it has key characteristics like multi-tenancy, scalability, performance, security, etc. Economic benefits are the main driver for the Cloud, since it promises the reduction of capital expenditure (CapEx) and operational expenditure (OpEx).

Cloud Computing is facing many challenges like Data Security, Energy Consumption, Server Consolidation, Virtual Machine Migration, etc. Existing approaches of workload management have investigated the consolidation of VMs using the similar type of workload. This research work focuses on the study of management of heterogeneous workloads in a cloud environment. Management of heterogeneous workloads helps in reducing the energy consumption of data centers, achieving the efficient resource usage and reduction in operational costs which benefits the end-users from decreased prices for resource usage.

In this thesis, available workload consolidation techniques have been compared. A Workload Consolidation Cloud portal has been designed, developed and presented for the implementation of heterogeneous workload consolidation technique for green cloud. This technique allows the dynamic adaption of VM allocation at run-time according to the current utilization of resources applying live migration, switching idle nodes to the sleep mode, and thus minimizing energy consumption. A simulated environment, CloudSim Toolkit has been used to validate the experimental results. The experimental results demonstrate that the proposed approach can effectively handle strict SLAs, heterogeneous infrastructure and heterogeneous VMs.

# Table of Content

---

---

Certificate.....	ii
Acknowledgement.....	iii
Abstract.....	iv
Table of Content.....	v
List of Figures and Tables.....	vii
Chapter 1 Introduction.....	1
1.1 Cloud Computing Evolution.....	1
1.2 Research Issues in Cloud Computing.....	7
1.3 Research Motivation.....	9
1.4 Organization of Thesis.....	10
Chapter 2 Literature Review.....	11
2.1 Virtualization.....	11
2.2 Resource Allocation Policies.....	13
2.3 Workload Consolidation Techniques.....	14
2.4 Green Cloud Architecture.....	21
Chapter 3 Problem Analysis.....	24
3.1 Gap Analysis.....	24
3.2 Problem Statement.....	25
Chapter 4 Proposed Technique for Heterogeneous Workloads.....	26
4.1 Design of Solution.....	26
Chapter 5 Experimental Results.....	32
5.1 Tools for setting Cloud Environment.....	32
5.2 Implementation of the Proposed Technique.....	34
5.3 Experimentation Work.....	42
5.4 Results and Conclusions.....	46
Chapter 6 Conclusion and Future Scope.....	47
6.1 Conclusion.....	47
6.2 Thesis Contribution.....	47
6.3 Future Scope.....	47

References.....	48
List of Papers.....	53

## List of Figures and Tables

---

### Figures

Figure 1.1: Evolution of Cloud Computing [6].....	1
Figure 1.2: Non-Exhaustive view on the main aspects forming a cloud system [7].....	3
Figure 1.3: Services of cloud computing.....	4
Figure 1.4: Cloud Computing Models [8].....	6
Figure 2.1: Life of a virtual machine: boot, run, and halt [18].....	12
Figure 2.2: Energy Consumption and SLA violation by ST Policy [19].....	13
Figure 2.3: Distribution of energy in data center.....	20
Figure 2.4: High Level Green Cloud Architecture [19].....	22
Figure 4.1: Layered Architecture of Cloud Computing.....	26
Figure 4.2: Context Level DFD.....	28
Figure 4.3: Use Case Diagram for WC Cloud Portal.....	29
Figure 4.4: Sequence Diagram of user and administrator.....	29
Figure 4.5: Sequence Diagram of member with server and administrator.....	30
Figure 4.6: Activity Diagram for query Processing.....	30
Figure 4.7: Activity Diagram showing activities after a member login.....	31
Figure 4.8: Deployment Diagram.....	31
Figure 5.1: Layered CloudSim Architecture [20].....	33
Figure 5.2: Home Page of the Portal.....	34
Figure 5.3: Login Error: user name or password do not matches.....	35
Figure 5.4: Registration Page for New User.....	35
Figure 5.5: Registration done by the New User.....	36
Figure 5.6: Registration Error: Required field left empty.....	36
Figure 5.7: Registration Error: Email id do not matches.....	37
Figure 5.8: Query entered by the user.....	37
Figure 5.9: Missing required field in Query page.....	38
Figure 5.10: page to give the inputs to calculate the energy consumed by homogeneous workloads.....	38
Figure 5.11: Shows the energy consumed by homogeneous workloads.....	39

Figure 5.12 shows the energy consumed when threshold is set to 50%.....	39
Figure 5.13 shows the wrong value of inputs entered by the user.....	40
Figure 5.14: page to give the inputs to calculate the energy consumed by homogeneous workloads.....	40
Figure 5.15: Shows the energy consumed by heterogeneous workloads when variability in length is 5000.....	41
Figure 5.16 Shows the graph when variability in heterogeneous workloads is 1000....	41
Figure 5.17 Shows the graph when variability in heterogeneous workloads is 5000....	42
Figure 5.18: Showing energy consumption and SLA violation of case 1.....	43
Figure 5.19: Showing energy consumption and SLA violation of case 2.....	44
Figure 5.20: Showing energy consumption and SLA violation of case 3.....	44
Figure 5.21: Showing energy consumption and SLA violation of case 4.....	45
Figure 5.22: Showing energy consumption and SLA violation of case 5.....	45
Figure 5.23: Showing energy consumption and SLA violation of case 6.....	46
<b>Table</b>	
Table 2.1: Comparison of Workload Consolidation Techniques.....	18
Table 3.1: Showing Heterogeneous Workload Consolidation Techniques.....	24
Table 5.1: Value of Threshold according to the heterogeneity between the applications.....	46

# Chapter 1

## Introduction

---

This chapter introduces Cloud computing, its evolution and various related technologies like distributed computing, grid computing, Cloud characteristics and the services offered by Cloud along with motivation of research and organisation of thesis.

### 1.1 Cloud Computing Evolution

The term Cloud has been used historically as a metaphor for the Internet. Cloud computing has evolved through a number of phases which include Grid and Utility computing, Application Service Provision (ASP), and Software as a Service (SaaS) as shown in Figure 1.1. The underlying concept of Cloud computing dates back to the 1960s, when John McCarthy opined that computation may someday be organized as a public utility.

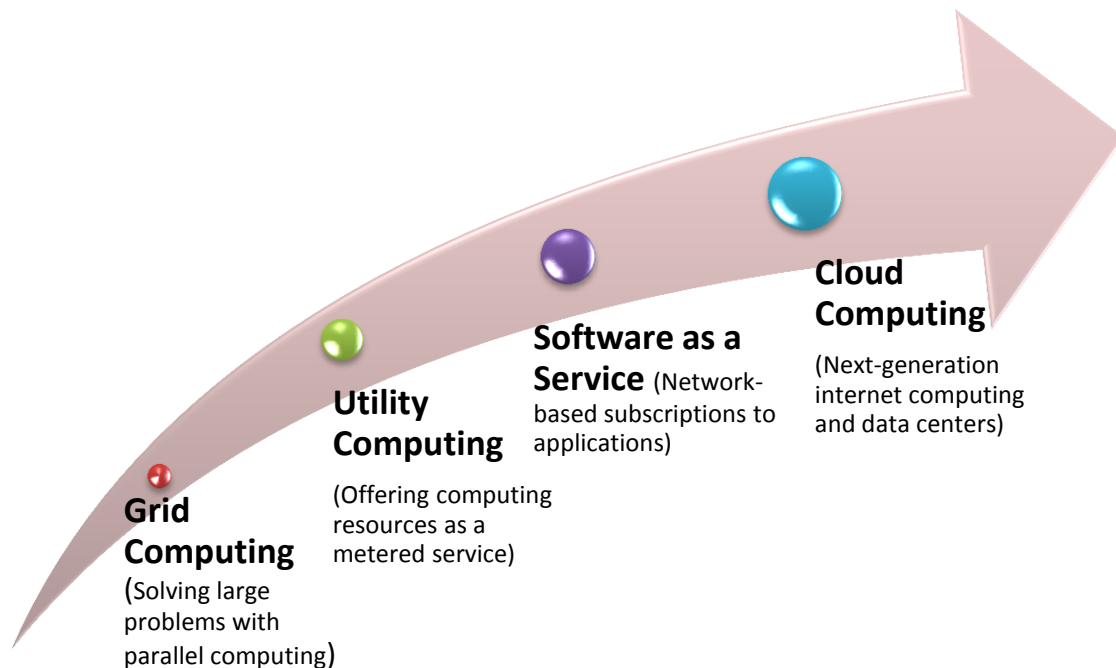


Figure 1.1: Evolution of Cloud Computing [6]

It is also said that Cloud computing concept of showing the entire relevant computer network was forwarded by J.C.R Licklider and John McCarthy. The idea of an "intergalactic computer network" was introduced in the sixties by J.C.R. Licklider, who

was responsible for enabling the development of ARPANET (Advanced Research Projects Agency Network) in 1969, his vision, was for everyone on the globe to be interconnected and accessing programs and data at any site, from anywhere. “Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services” [1].

The data in a Cloud is stored on to centralized location called data centers having a large size of data storage. The data as well as processing is somewhere on servers. So, the clients have to trust the provider on the availability as well as data security. The SLA is the only legal agreement between the service provider and client. “The basic principle of Cloud computing is to distribute the computing tasks to many distributed computers, not local computer or remote servers” [2].

In 1999 salesforce.com put the idea of Cloud computing in an application. They used a simple website to deliver its enterprise applications to users. Then in 2002 amazon launched their Cloud based web services. Cloud computing does not stop here it was then used by Microsoft to promote the Cloud computing services by Azure Services platform in October 27, 2008. Google in 2009 came into market by using Cloud computing concept by delivering Google Apps.

Cloud Computing offers service to end-users rather than a product, by sharing resources, software and other information under a usage based payment model. “Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services” [A15]. Figure 1.2 shows the main aspects of Cloud computing. There are many proposed definitions of the Cloud computing due to its growing popularity defining its characteristics. Some of the definitions given by many well-known scientists and organizations are:

- Rajkumar Buyya defines the Cloud computing in terms of its utility to end user as “A Cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers” [3].

- National Institute of Standards and Technology (NIST) defines Cloud computing as follows: “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This Cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models” [4].

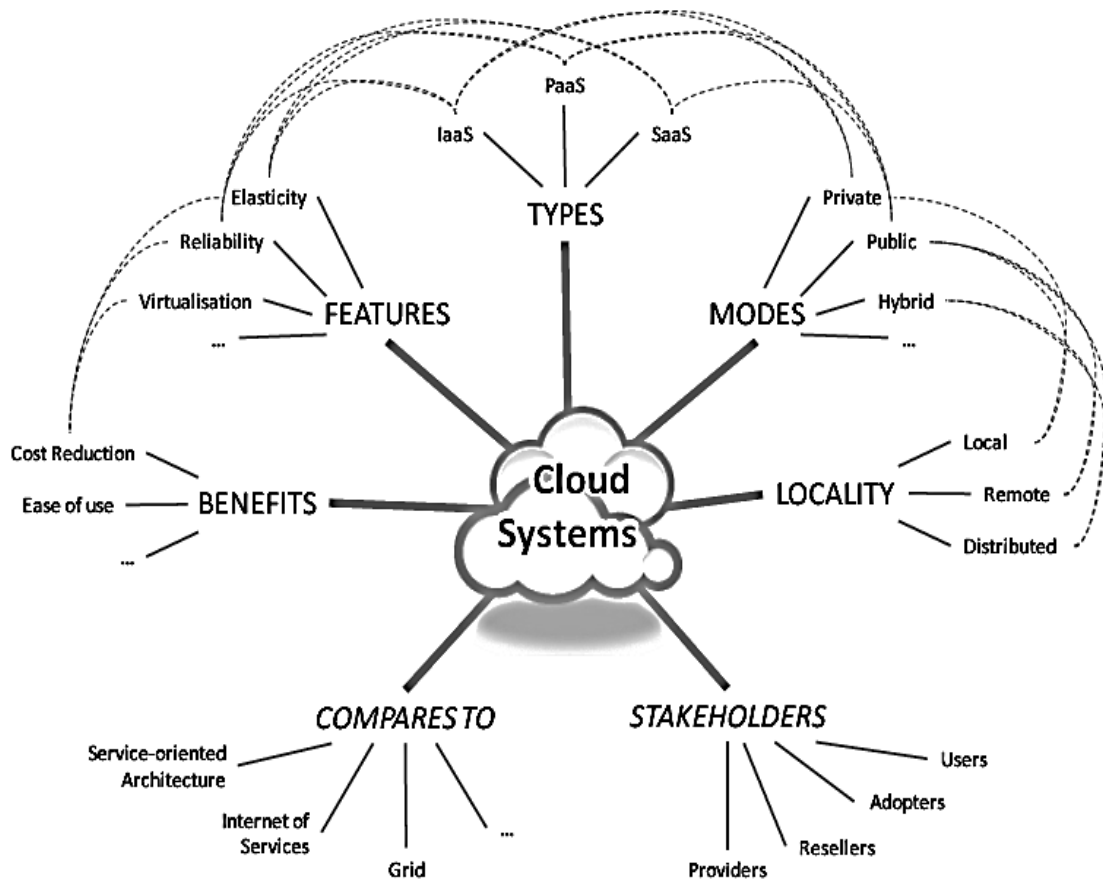


Figure 1.2: Non-Exhaustive view on the main aspects forming a cloud system [7]

### 1.1.1 Characteristics of Cloud Computing

The different characteristics of Cloud computing [5] are described below:

- **Reduced cost**: There are a number of reasons to attribute Cloud technology with lower costs. The billing model is pay as per usage; the infrastructure is not purchased thus lowering maintenance. Initial expense and recurring expenses are much lower than traditional computing.

- **Increased Storage:** With the massive Infrastructure that is offered by Cloud providers today, storage & maintenance of large volumes of data is a reality. Sudden workload spikes are also managed effectively & efficiently, since the Cloud can scale dynamically.
- **Flexibility:** Cloud computing stresses on getting applications to market very quickly, by using the most appropriate building blocks necessary for deployment.
- **Reliability:** It is the capability to ensure constant operation of the system without disruption, i.e. no loss of data, no code reset during execution etc. Reliability is typically achieved through redundant resource utilization.
- **Location independence:** It enables users to access systems using a web browser regardless of their location or what device they are using.

### 1.1.2 Cloud Computing Services

Cloud computing is typically divided into three levels of service offerings [5] as shown in Figure 1.3.

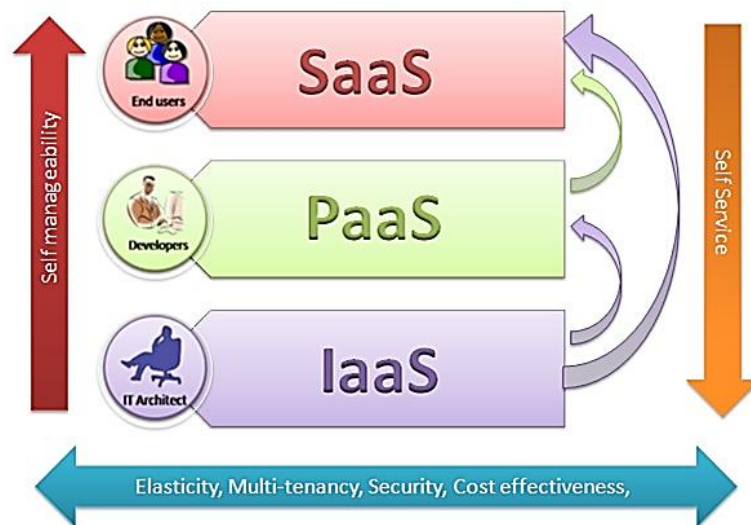


Figure 1.3: Services of cloud computing

- **Software as a Service (SaaS)**

A complete application is offered to the customer, as a service on demand. A single instance of the service runs on the Cloud & multiple end users are serviced. On the customer side, there is no need for upfront investment in servers or software licenses, while for the provider, the costs are lowered, since only a single application needs to be

hosted & maintained. Today SaaS is offered by companies such as Google, Salesforce, Microsoft, Zoho, etc.

- Platform as a Service (PaaS)

PaaS is an application development and deployment platform delivered as a service to developers over the Web. It facilitates development and deployment of applications without the cost and complexity of buying and managing the underlying infrastructure, providing all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available from the Internet. This platform consists of infrastructure software, and typically includes a database, middleware and development tools. PaaS providers offer a predefined combination of OS and application servers, such as LAMP platform (Linux, Apache, MySQL and PHP), Google Apps Engine is a PaaS offering where developers write in Python or Java.

- Infrastructure as a service (IaaS)

Infrastructure as a Service is the delivery of hardware (server, storage and network), and associated software (operating systems virtualization technology, file system), as a service. It is an evolution of traditional hosting that does not require any long term commitment and allows users to provision resources on demand. Unlike PaaS services, the IaaS provider does very little management other than keep the data center operational and users must deploy and manage the software services themselves – just the way they would in their own data center. Amazon Web Services Elastic Compute Cloud (EC2) and Secure Storage Service (S3), GoGrid, 3 Tera are the examples of IaaS offerings.

### **1.1.3 Deployment Models of Cloud Computing**

The different deployment models of Cloud Computing [5] are shown in Figure 1.4.

- Public Cloud

Public Cloud is the most common deployment model where services are available to anyone on Internet in a pay-as-you-go manner. To support thousands of public domain users, datacenters built by public Cloud providers are quite large comprising of thousands of servers with high speed network. Some of the famous public Clouds are Amazon Web Services (AWS), Google App Engine, and Microsoft Azure. Amazon EC2 provides infrastructure as a service, Google AppEngine provides platform as a service, and

Salesforce.com provides software as a service. The characteristics of Public Clouds are multi-tenancy and scalability.

- Private Cloud

Private Clouds are deployed within the premise of an organization to provide IT services to its internal users. The private Cloud services offer greater control over the infrastructure, improving security and service resilience because its access is restricted to one or few organizations. Such private deployment poses an inherent limitation to end user applications i.e. inability to scale elastically on demand as can be done using public Cloud services. An organization can buy more machines according to expanding needs of its users, but this cannot be done as fast and seamlessly as with public Clouds.

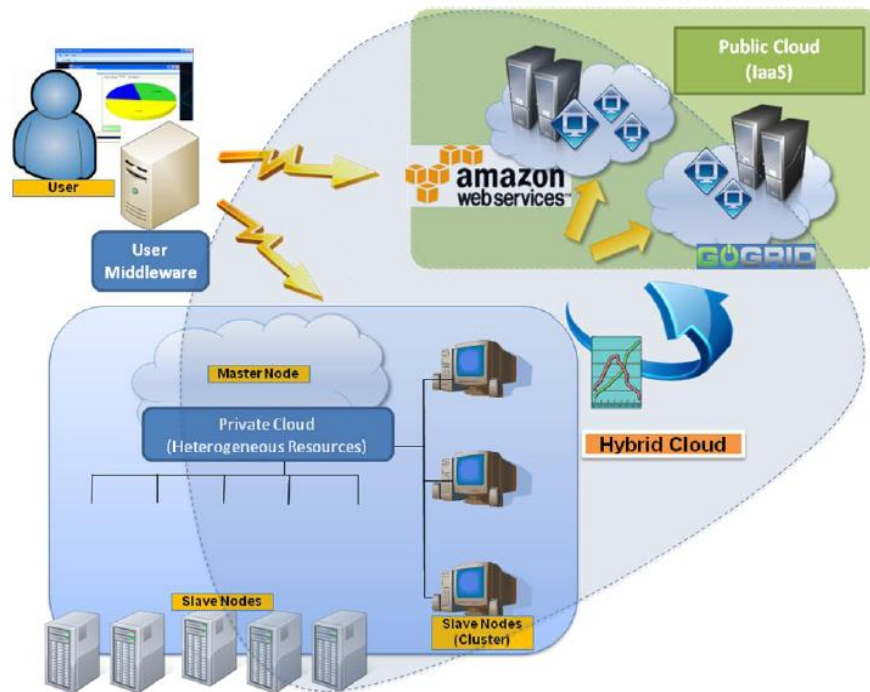


Figure 1.4: Cloud Computing Models [8]

- Hybrid Cloud

Hybrid Clouds is the deployment which emerged due to diffusion of both public and private Clouds advantages. In this model, organizations outsource non-critical information and processing to the public Cloud, while keeping critical services and data in their control. Therefore, organizations can utilize their existing IT infrastructure for maintaining sensitive information within the premises. With a Hybrid Cloud, service

providers can utilize 3rd party Cloud Providers in a full or partial manner thus increasing the flexibility of computing.

## **1.2 Research Issues in Cloud Computing**

The emergence of Cloud computing has made a tremendous impact on the Information Technology (IT) industry over the past few years. Currently IT industry needs Cloud computing services to provide best opportunities to real world. Cloud computing is in initial stages, with many issues [11], [9], [10] still to be addressed.

- **Automated Service Provisioning**

The objective of a service provider in this case is to allocate and de-allocate resources from the Cloud to satisfy its service level objectives (SLOs), while minimizing its operational cost. These approaches typically involve: (i) Constructing an application performance model that predicts the number of application instances required to handle demand at each particular level, in order to satisfy QoS requirements; (ii) Periodically predicting future demand and determining resource requirements using the performance model; and (iii) Automatically allocating resources using the predicted resource requirements. Application performance model can be constructed using various techniques, including Queuing theory, Control theory and Statistical Machine Learning. The proactive approach uses predicted demand to periodically allocate resources before they are needed. The reactive approach reacts to immediate demand fluctuations before periodic demand prediction is available.

- **Virtual Machine Migration**

Virtualization can provide significant benefits in Cloud computing by enabling virtual machine migration to balance load across the data center. In addition, virtual machine migration enables robust and highly responsive provisioning in data centers. Authors have pointed out that migrating an entire OS and all of its applications as one unit allows to avoid many of the difficulties faced by process-level migration approaches, and analyzed the benefits of live migration of VMs [12]. The major benefit of VM migration is to avoid hotspots. Detecting workload hotspots and initiating a migration lacks the agility to respond to sudden workload changes.

- Server Consolidation

Server consolidation is an effective approach to maximize resource utilization while minimizing energy consumption in a Cloud computing environment. Live VM migration technology is often used to consolidate VMs residing on multiple under-utilized servers onto a single server, so that the remaining servers can be set to an energy-saving state. The problem of optimally consolidating servers in a data center is often formulated as a variant of the vector bin-packing problem, which is an NP-hard optimization problem. Various heuristics have been proposed for this problem. Server consolidation activities should not hurt application performance. For server resources that are shared among VMs, such as bandwidth, memory cache and disk I/O, maximally consolidating a server may result in re-source congestion when a VM changes its footprint on the server. Hence, it is sometimes important to observe the fluctuations of VM footprints and use this information for effective server consolidation. Finally, the system must quickly react to resource congestions as and when they occur.

- Energy Management

Improving energy efficiency is another major issue in Cloud computing. It has been estimated that the cost of powering and cooling accounts for 53% of the total operational expenditure of data centers [13]. In 2006, data centers in the US consumed more than 1.5% of the total energy generated in that year, and the percentage is projected to grow 18% annually [14]. Hence infrastructure providers are under enormous pressure to reduce energy consumption. The goal is not only to cut down energy cost in data centers, but also to meet government regulations and environmental standards. Designing energy-efficient data centers has recently received considerable attention. This problem can be approached from several directions. For example, energy-efficient hardware architecture that enables slowing down CPU speeds and turning off partial hardware components has become commonplace. Energy-aware job scheduling and server consolidation are two other ways to reduce power consumption by turning off unused machines. A key challenge in all the above methods is to achieve a good trade-off between energy savings and application performance.

- Data Security

Data security is another important research topic in Cloud computing. Since service providers typically do not have access to the physical security system of data centers, they must rely on the infrastructure provider to achieve full data security. Even for a virtual private Cloud, the service provider can only specify the security setting remotely, with-out knowing whether it is fully implemented. It is critical to build trust mechanisms at every architectural layer of the Cloud. Firstly, the hardware layer must be trusted using hardware trusted platform module. Secondly, the virtualization platform must be trusted using secure virtual machine monitors. VM migration should only be allowed if both source and destination servers are trusted.

### **1.3 Research Motivation**

Green Computing [39], or Green IT, is the practice of implementing policies and procedures that improve the efficiency of computing resources in such a way as to reduce the energy consumption and environmental impact of their utilization [40] [41].

Cloud computing is a current advancement where applications and IT infrastructure are provided as ‘services’ on a usage based payment model. There are many issues in Cloud computing such as Automated Service Provisioning, Virtual Machine Migration, Energy Management, Server Consolidation, Data Security, etc. as discussed in previous section that have not been fully addressed. Central to these issues is the Energy Management. There is a rapid expansion in data centers, due to the exponential growth of the Cloud computing. This expansion has triggered the dramatic increase in energy used and its effect on the environment in terms of carbon footprints. In order to reduce power consumption, it is necessary to consolidate the hosting workloads.

Various existing techniques manage the heterogeneous workloads but are not energy efficient for the Cloud computing platform. Aim of the thesis is to consolidate the heterogeneous workloads in an efficient way so that the resource utilization can be maximized and the energy consumption of the data center could be minimized that can further result in reducing carbon footprints and hence assist in achieving Green Computing.

## 1.4 Organization of Thesis

The rest of the thesis is organized as follows:

**Chapter 2** – This chapter describes in detail the literature survey done to study the concept of virtualization, existing techniques of workload consolidation and the green Cloud architecture.

**Chapter 3** – This chapter describes the problem Analysis of the thesis work. It gives the gap analysis and problem statement.

**Chapter 4** – This chapter describes in detail the solution of the problem with the help of PDL of the Proposed Technique and UML diagrams.

**Chapter 5** – This chapter focus on the implementation details and experimental results – description of CloudSim, Eclipse and snapshots of the Cloud portal designed to study the technique.

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

## Chapter 2

### Literature Review

---

This chapter discusses the state of the art and research issues of Virtualization, Resource Allocation policies, Workload Consolidation Techniques and Architecture of Green Cloud.

#### 2.1 Virtualization

Virtualization is another term for abstraction. It is a technology that allows running two or more operating systems side-by-side on just one PC or embedded controller [34]. It is rapidly adopted in the engineering world. It helps them to better utilize their available processing hardware to build more efficient systems. As multicore processors with 4, 8, and 16 cores on a chip become a common place, many processor cores are likely to be underutilized in a typical system. Most applications will have only a finite amount of parallel tasks that can be executed at a given time, leaving many processor cores idle. Virtualization can solve this challenge by allocating a group of processor cores to individual operating systems running in parallel. Thus benefits of virtualization can be enumerated as:

- (i) Save hardware cost and footprint.
- (ii) Take advantage of Operating System services.
- (iii) Make use of Multicore Processors.

Virtualisation technology allows datacentres to lease virtual machines to end users, rather than traditional dedicated machines. This system is the basis of Cloud Computing's on-demand, utility model. Virtualisation can also increase the efficiency of existing machine rooms, reducing the number of physical servers required through consolidation of existing applications by introducing multiple virtual machines per server and thereby increasing resource utilisation. It is assumed that this increases efficiency; it allows you to do more work with less IT equipment. Increase utilisation of a machine can also increase its energy efficiency, as a typical blade server consumes 50% of its peak power drawn when idling. Well utilising that machine will allow the idle cost to be amortized.

The energy consumption of a virtual machine [18] is shown in Figure 2.1. The figure shows the boot up (t=10 to t=30), CPU burn (t=40 to t=100), shut down (t=110 to t=122) of a virtual machine. It is also shown that the boot and the shutdown consume really less energy than a CPU burn. So we should put a virtual machine to sleep mode to save the power consumed when it does nothing.

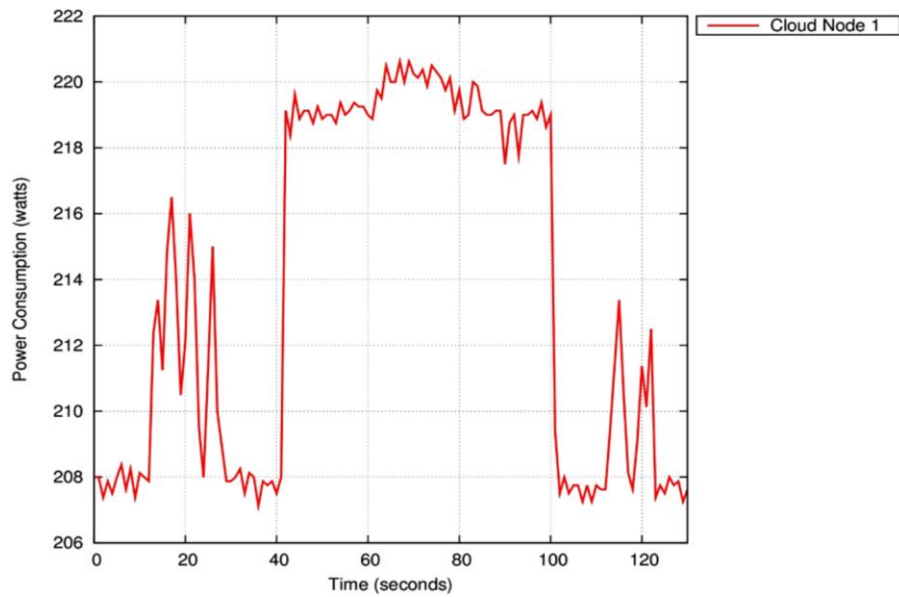


Figure 2.1: Life of a virtual machine: boot, run, and halt [18]

Virtual machine migration, which is used to transfer a VM across physical computers, has served as a main approach to achieve better energy efficiency of Internet Data Centers (IDCs). This is because in doing so, server consolidation via VM migrations allows more computers to be turned off. Generally, there are two types of migrations [37]: regular migration and live migration. The former moves a VM from one host to another by pausing the originally used server, copying its memory contents, and then resuming it on the destination. The latter performs the same logical functionality but without the need to pause the server domain for the transition. In general when performing live migrations the domain continues its usual activities and from the user's perspective—the migration should be imperceptible. It shows great potential of using VM and VM migration technology to efficiently manage workload consolidation, and therefore improve the total IDC power efficiency [22].

## 2.2 Resource Allocation Policies

There are different policies to allocate the resources to the coming cloudlets so as to maximize the resource utilization and minimize the energy consumptions. Some of the resource allocation policies are discussed below:

- Single Threshold

This policy is based on the idea of setting upper utilization threshold for hosts and placing VMs while keeping the total utilization of CPU below this threshold. In this policy upper utilization threshold is set for the hosts and place the VMs by keeping the total utilization of CPU below this threshold. “The aim is to preserve free resources to prevent SLA violation due to consolidation in cases when utilization by VMs increases” [19]. The energy consumption and SLA violation by Single Threshold Policy is shown in Figure 2.2. The figure shows that the SLA violation increases with the growth of utilization threshold and energy consumption decreases.

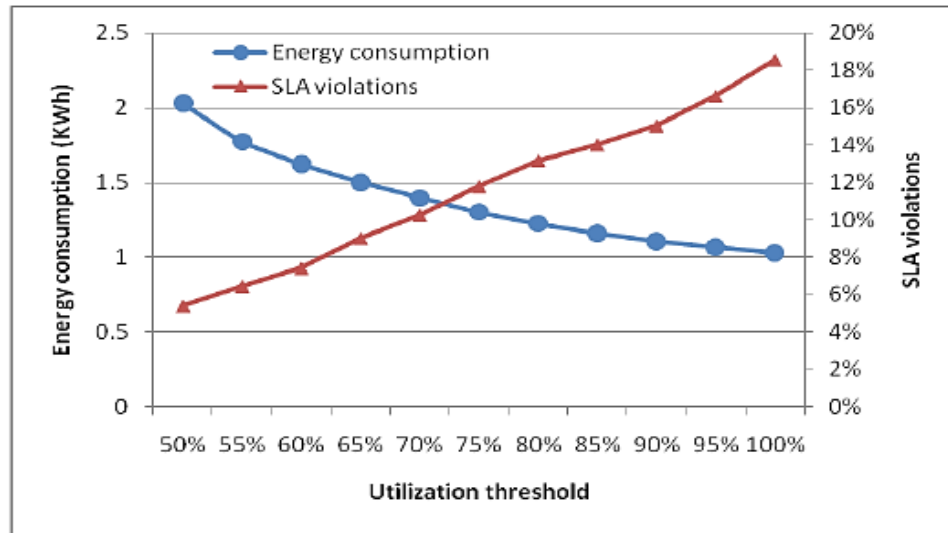


Figure 2.2: Energy Consumption and SLA violation by ST Policy [19]

- Minimization of Migrations (MM):

In this policy we migrate the least number of VMs to minimise migration overhead. In this we set upper and lower utilization thresholds for hosts and keeping total utilization of CPU by all VMs between these thresholds. If the utilization of CPU for a host goes below the lower threshold, all VMs have to be migrated from this host and the host has to be

switched off in order to eliminate the idle power consumption. If the utilization goes over the upper threshold, some VMs have to be migrated from the host to reduce utilization in order to prevent potential SLA violation. The algorithm for minimum migration policy is shown in the Figure 2.2. The comparison between Minimization of migration, High Potential Growth and Random Choice Policies is that “the MM and RC policies lead to significantly less SLA violations than the HPG policy. The comparison is shown in the Figure 8. Moreover, the usage of the MM policy results in a significantly reduced number of VM migrations in comparison to the other policies” [21].

- **Highest Potential Growth (HPG):**

When the upper threshold is violated, the Highest Potential Growth (HPG) policy migrate VMs that have the lowest usage of CPU relatively to the requested in order to minimise total potential increase of the utilization and SLA violation.

- **Random Choice (RC):**

In this policy we choose the necessary number of VMs by picking them according to a uniformly distributed random variable, to decrease the CPU utilization by a host below the upper utilization threshold.

- **Non Power-Aware (NPA):**

This policy does not apply any power-aware optimizations and implies that all hosts run at 100% CPU utilization and consume maximum power all the time.

- **Dynamic Voltage Frequency Scaling (DVFS):**

Dynamic Voltage Frequency Scaling policy adjusts the voltage and frequency of the CPU according to the current utilization.

## **2.3 Workload Consolidation Techniques**

Workload consolidation techniques are classified according to the type of workload:

### **2.3.1 Homogeneous Workload Consolidation Techniques**

These techniques include the similar kind of workloads which is having fixed number of parameters such as length, number of CPU's required and buffer size of input and output

files, etc. the different techniques to manage homogeneous workloads are discussed below:

- Nathuji et al. [22] have explored the problem of power efficient resource management in large scale virtualized data centers and given a new technique soft resource scaling. They pointed out the number of benefits of virtualization such as: improved fault and performance isolation between applications sharing the same resource; ability to relatively easy move of VMs between physical hosts applying live or offline migration; support for hardware and software heterogeneity. The experimental results show that reduction in power consumption up to 17% by exploiting power management heterogeneity.
- Kim et al. [23] have proposed a real-time virtual machine model in which provider provisions VMs for requested real-time applications and ensures the meeting of specified deadline constraints. The problem is addressed using different levels. For scheduling real-time VMs in a data center, the authors have proposed three policies:
  - (i) The lowest DVFS policy,
  - (ii) The  $\delta$ -Advanced-DVS policy,
  - (iii) The Adaptive-DVS policy.

These policies help to reduce the energy consumption, maximizes the acceptance rate of provisioning requests, while meeting the deadline constraints. To evaluate the proposed techniques CloudSim Toolkit is used.

- Buyya et al. [19] have focused on the development of dynamic resource provisioning and allocation algorithms that consider the synergy between various data center infrastructures and work to boost data center energy efficiency and performance and proposed “(i) architectural principles for energy-efficient management of Clouds; (ii) energy-efficient resource allocation policies and scheduling algorithms considering quality-of-service expectations, and devices power usage characteristics; and (iii) a novel software technology for energy-efficient management of Clouds.” They used the CloudSim Toolkit for the evaluation of the performance evaluation and gave the Green Cloud Architecture. Their simulation results shows that MM policy leads to the best energy savings and dynamic reallocation of VMs according to current utilization of CPU provides higher energy savings compared to static allocation policies.

- Gmach et al. [24] have explored the problem of energy efficient consolidation of VMs dynamically in enterprise environment. The authors had proposed the technique, which is a combination of a trace-based workload placement controller and a reactive migration controller. They had conducted the simulation-driven evaluation by using the three-months of real-world workload traces for 138 SAP applications and had shown that by applying both the optimization controllers together we can achieve the best results rather than applying separately.
- Verma et al. [25] have explored the problem of dynamic placement of applications in virtualized systems, while minimizing the power consumption and maintaining the SLA. The authors have proposed the pMapper application placement framework, which consists of three managers and an arbitrator, which coordinates their actions and makes allocation decisions. The experimental results show that the approach allows saving about 25% of power relative to the Static and Load Balanced Placement algorithms. The researchers suggest several directions for future work, such as consideration of memory bandwidth, more advanced application of idle states and extension of the theoretical prove of the problem.
- Kusic et al. [26] have investigated the problem of power and performance of efficient resource management in virtualized computing systems. They focused to maximize the resource provider profit by minimizing the SLA violation and power consumption. They stated the problem as a sequential optimization and addressed using Limited Lookahead Control (LLC). The experimental results show that a server cluster managed using LLC saves 26% in the power consumption costs over a 24 hour period when compared to an uncontrolled system. Power savings are achieved with 1.6% SLA violations of the total number of requests.
- Sharifi et al. [27] have consolidated the Virtual Machines in order to efficiently use their hardware resources and save the power and presented the four models to identify the performance interferences between processor and disk utilizations and the costs of migrating VMs. The models are :
  - (i)The target system model,
  - (ii)The application model,

(iii)The energy model,

(iv)The migration model.

They also presented “a consolidation fitness metric to evaluate the merit of consolidating a number of known VMs on a PM based on the processing and storage workloads of VMs. We then propose an energy-aware scheduling algorithm using a set of objective functions in terms of this consolidation fitness metric and presented power and migration models [28].” The results of Consolidation fitness of disk intensive workloads and processor intensive workloads are also presented and their effect on eachother. Simulative results of the algorithm shows that 24.9% power savings compared to two other scheduling algorithms.

### **2.3.2 Heterogeneous Workload Consolidation Techniques**

These techniques include different kind of workload which is having varying parameters like length, number of CPU's required and buffer size of input and output files. The different techniques to manage heterogeneous workloads are described below:

- Carrera et al. [29] have explored the problem of managing heterogeneous workloads in a virtualized data centers. They have proposed a technique which dynamically modifies the heterogeneous workload placement on a same physical hardware. The utility functions used in this technique helped them to compare the performance of different workloads. This technique maximizes the performance of heterogeneous workloads.
- Zhan et al. [30] proposed the Phoenix Cloud, which provides the cooperative resource provision and management policies to manage the heterogeneous workloads for same organization. The simulation results have shown that this technique results in increased number of completed jobs.
- Tian et al. [31] have given the new MapReduce model to manage the heterogeneous workloads .They have categorized the workload into three categories namely CPU-bound, CPU-bound without shuffle and I/O-bound. The simulation results show the improved throughput of Hadoop by 30%, by using the triple queue scheduler mechanism.

- Steinder et al. [32] have explored the technique that automatically manages the data centers running the heterogeneous workloads. This technique allows to collocate the heterogeneous workloads on the any server, uses the high level performance goals for resource allocation and is also compatible with homogeneous workloads. Authors have used the both real system as well as simulation to evaluate the technique.
- Goiri et al. [33] have proposed the scheduling technique to manage heterogeneous workloads such as HPC jobs and web-based applications on the virtualized data centers, which aims to maximize the providers benefit. Looking to the consumers it provides energy efficiency, fault tolerance and SLA violation penalties. The simulation results show that provider benefit is increased by 15% and improvements in energy and performance efficiency.

The various techniques mentioned above are summarized in the table 2.1 below, which manage both the homogeneous and heterogeneous workloads. Existing heterogeneous workload consolidation techniques are not energy efficient for the Cloud computing platform. The aim of this thesis is to consolidate the heterogeneous workloads in an efficient way so that resource utilization can be maximized and energy consumption of the data center could be minimized resulting in the less number of carbon footprints and hence assisting in achieving Green Computing.

Table 2.1: Comparison of workload Consolidation Techniques

<b>Technique (Year)</b>	<b>Power Saving Technique</b>	<b>System Resources</b>	<b>Workload Type</b>	<b>Findings</b>
(i) Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing (2011)	Leveraging heterogeneity of Cloud data centers, DVFS	CPU	HPC Applications	(i) Minimized energy consumption. (ii) Satisfied performance requirements.
(ii) Energy Efficient Resource Management in Virtualized Cloud Data Centers (2010)	VM Consolidation	CPU	Arbitrary	(i) Reduces the no. of active servers. (ii) Substantial energy savings while ensuring QoS.
(iii) Multifaceted Resource Management for Dealing	VM	CPU	(i) HPC jobs	(i) Provides 15%

with Heterogeneous Workloads in Virtualized Data Centers (2010)	consolidation		(ii) Web based applications	benefit to the provider. (ii) Reduces power.
(iv) vGreen: A System for Energy Efficient Computing in Virtualized Environments (2009)	VM consolidation, VM scheduling across different PMs	CPU, Memory	Arbitrary	(i) Improves overall performance and energy efficiency. (ii) Balances the power of the system.
(v) A Dynamic MapReduce Scheduler for Heterogeneous Workloads (2009)	VM consolidation	CPU, Memory	(i) CPU bound (100%utilization) (ii) CPU bound (without shuffle) (iii) I/O bound	Improve Hadoop throughput by 30%.
(vi) Power-aware Provisioning of Cloud Resources for Real-time Services (2009)	DVFS, VM consolidation	CPU	Arbitrary	Datacenters can reduce power consumption by using DVS schemes.
(vii) Resource Pool Management: Reactive versus Proactive (2009)	VM consolidation, server power switching	CPU, Memory	Arbitrary	(i) Maximize resource utilization. (ii) Satisfy performance requirements.
(viii) Power and performance management of virtualized computing environments via lookahead control (2009)	DVFS, VM consolidation, server power switching	CPU	Online services	(i) Minimize power consumption. (ii) Minimize performance loss.
(ix) pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems (2008)	DVFS, VM consolidation, server power switching	CPU	Arbitrary	(i) Minimize power consumption. (ii) Minimize performance loss.
(x)Coordinated Multi-level Power Management for the Data Center (2008)	DVFS, VM consolidation, server power switching	CPU	Arbitrary	(i) Minimize power consumption. (ii) Minimize performance loss. (iii) Meet power budget.
(xi) Managing SLAs of Heterogeneous Workloads using Dynamic Application Placement (2008)	DVFS, VM consolidation	CPU	(i) Transactional Applications (ii) Long Running Jobs	Maximized performance of heterogeneous Workloads

(xii) Phoenix Cloud: Consolidating Heterogeneous Workloads of Large Organizations on Cloud Computing Platforms (2008)	Server Consolidation	CPU	(i) Parallel Batch Jobs (ii) Web Service	Increased no. of completed jobs
(xiii) VirtualPower: Coordinated Power Management in Virtualized Enterprise Systems (2007)	DFVS, soft scaling, VM consolidation, server power switching	CPU	Arbitrary	(i) Minimize energy consumption. (ii) Satisfy performance requirements.
(xiv) Server virtualization in autonomic management of heterogeneous workloads (2007)	Server Consolidation, VM consolidation	CPU	(i) Web Application, (ii) Non-interactive workloads	(i) Heterogeneous workloads can be collocated on any server machine. (ii) Used high level performance goals for resource allocation.

With the gained popularity of Cloud computing, power dependency has also increased. It has been estimated that the data centers consume 0.5 percent of the world's total electricity usage [15]. If current demand continues, it is projected to be augmented by 2020. In 2005, the total energy consumption (servers, cooling units) was projected at 1.2% of the total U.S. energy consumption, amplifying every 5 years [16]. Amazon's estimates [17] about its data center are shown in Figure 2.3. The major part of the energy used in society nowadays, is generated from fossil fuels which results in harmful CO<sub>2</sub> emissions. To manage the energy consumption of the Cloud, Green Cloud architecture is required, which is described below.

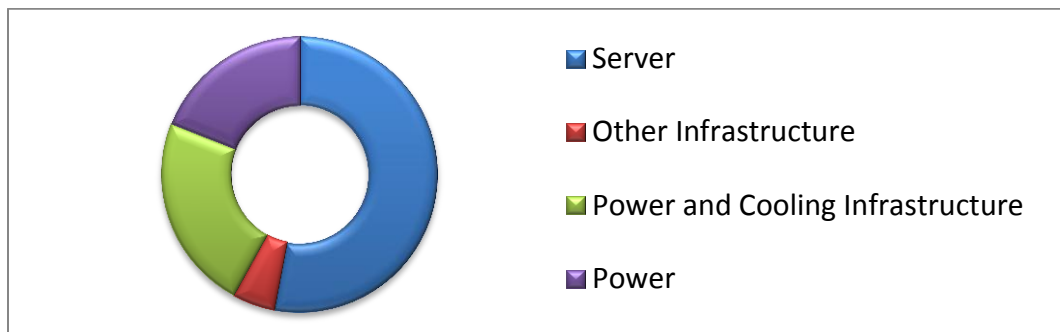


Figure 2.3: Distribution of energy in data center

## 2.4 Green Cloud Architecture

Green Cloud is an Internet Data Center architecture which aims to reduce data center power consumption, while at the same time guarantee the performance from users' perspective, leveraging live virtual machine migration technology. A big challenge for Green Cloud is to automatically make the scheduling decision on dynamically migrating/consolidating VMs among physical servers to meet the workload requirements meanwhile saving energy, especially for performance-sensitive applications [38]. Hence, a real-time VM consolidation is needed.

Figure 2.4 shows the high-level architecture for supporting energy-efficient service allocation in Green Cloud computing infrastructure [19]. There are basically four main entities involved:

- Consumers/Brokers:

Cloud consumers or their brokers submit service requests from anywhere in the world to the Cloud. It is important to notice that there can be a difference between Cloud consumers and users of deployed services. For instance, a consumer can be a company deploying a Web application, which presents varying workload according to the number of "users" accessing it.

- Green Resource Allocator:

Acts as the interface between the Cloud infrastructure and consumers. It requires the interaction of the following components to support energy-efficient resource management:

- Green Negotiator: Negotiates with the consumers/brokers to finalize the SLA with specified prices and penalties (for violations of SLA) between the Cloud provider and consumer depending on the consumer's QoS requirements and energy saving schemes. In case of Web applications, for instance, QoS metric can be 95% of requests being served in less than 3 seconds.
- Service Analyzer: Interprets and analyses the service requirements of a submitted request before deciding whether to accept or reject it. Hence, it needs the latest load and energy information from VM Manager and Energy Monitor respectively.

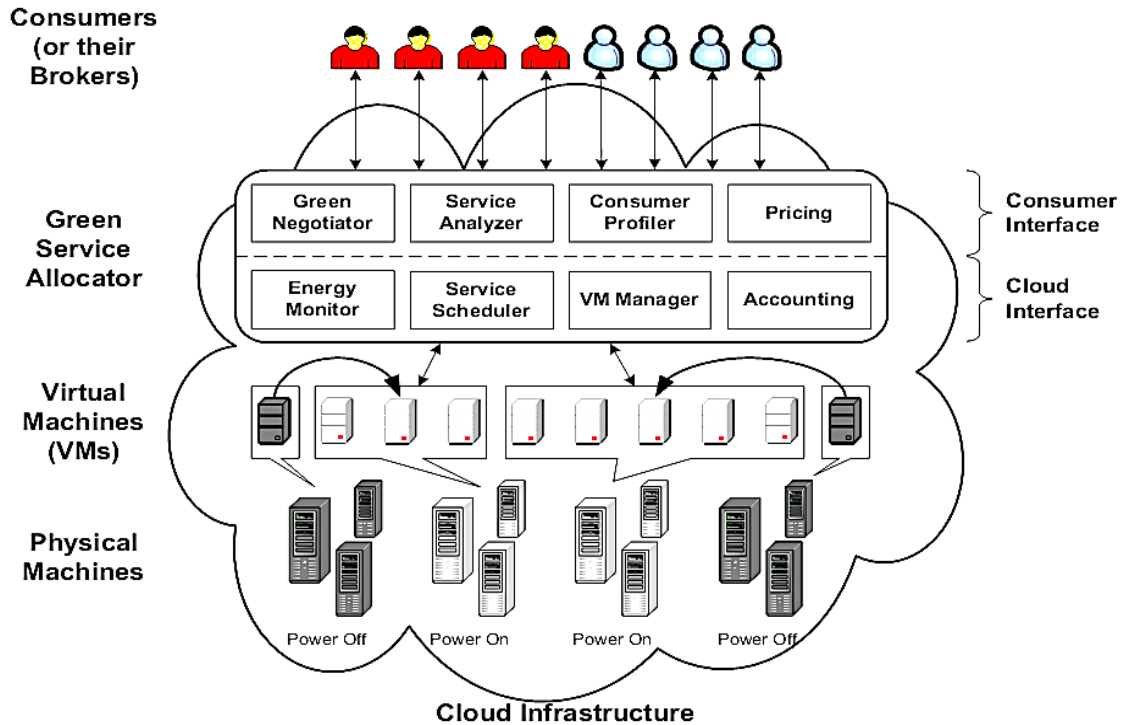


Figure 2.4: High Level Green Cloud Architecture [19]

- **Consumer Profiler:** Gathers specific characteristics of consumers so that important consumers can be granted special privileges and prioritized over other consumers.
- **Pricing:** Decides how service requests are charged to manage the supply and demand of computing resources and facilitate in prioritizing service allocations effectively.
- **Energy Monitor:** Observes and determines which physical machines to power on/off.
- **Service Scheduler:** Assigns requests to VMs and determines resource entitlements for allocated VMs. It also decides when VMs are to be added or removed to meet demand.
- **VM Manager:** Keeps track of the availability of VMs and their resource entitlements. It is also in charge of migrating VMs across physical machines.
- **Accounting:** Maintains the actual usage of resources by requests to compute usage costs. Historical usage information can also be used to improve service allocation decisions.

- VMs: Multiple VMs can be dynamically started and stopped on a single physical machine to meet accepted requests, hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests. Multiple VMs can also concurrently run applications based on different operating system environments on a single physical machine. In addition, by dynamically migrating VMs across physical machines, workloads can be consolidated and unused resources can be put on a low-power state, turned off or configured to operate at low-performance levels (e.g., using DVFS) in order to save energy.
- Physical Machines: The underlying physical computing servers provide hardware infrastructure for creating virtualized resources to meet service demands.

This chapter described the concept of Virtualization, Resource Allocation Policies, Workload Consolidation Techniques and Green Cloud Architecture. Next chapter presents the problem analysis.

Previous chapter analyzed various techniques of workload consolidation available. This chapter focuses on the gaps in literature review that has been reviewed.

### 3.1 Gap Analysis

Cloud infrastructure services provide users with the ability to provision virtual machines and allocate any kind of applications on them. This leads to the fact that different types of applications (e.g., enterprise, scientific, and social network applications) can be allocated on one physical computer node. However, it is not obvious how these applications can influence each other, as they can be data, network or compute intensive thus creating variable or static load on the resources.

Based on the literature survey there are various techniques (as shown in the Table 3.1), which manage the heterogeneous workloads but are not energy efficient as they provide high performance without meeting SLA violations and without focusing on allocating VMs to minimize energy consumption for the Cloud computing platform.

Table 3.1: Heterogeneous Workload Consolidation Techniques

Technique (Year)	Power Saving Technique	System Resources	Workload Type	Findings
(i) Multifaceted Resource Management for Dealing with Heterogeneous Workloads in Virtualized Data Centers (2010)	VM consolidation	CPU	(i) HPC jobs (ii) Web based applications	(i) Provides 15% benefit to the provider. (ii) Reduces power.
(ii) A Dynamic MapReduce Scheduler for Heterogeneous Workloads (2009)	VM consolidation	CPU, Memory	(i) CPU bound (100%utilization) (ii) CPU bound without shuffle) (iii) I/O bound	Improve Hadoop throughput by 30%.
(iii) Managing SLAs of Heterogeneous Workloads using Dynamic Application Placement (2008)	DVFS, VM consolidation	CPU	(i) Transactional Applications (ii) Long Running Jobs	Maximized performance of heterogeneous Workloads

(iv) Phoenix Cloud: Consolidating Heterogeneous Workloads of Large Organizations on Cloud Computing Platforms (2008)	Server Consolidation	CPU	(i) Parallel Batch Jobs (ii) Web Service	Increased no. of completed jobs
(v) Server virtualization in autonomic management of heterogeneous workloads (2007)	Server Consolidation, VM consolidation	CPU	(i) Web Application, (ii) Non-interactive workloads	(i) Heterogeneous workloads can be collocated on any server machine. (ii) Used high level performance goals for resource allocation.

### 3.2 Problem Statement

Heterogeneous workloads can be consolidated in a better way by consolidating the virtual machines. Virtualization technologies, on which Cloud computing environments heavily rely on, provide the ability to transfer VMs between physical nodes using live or offline migration. This enables the technique of dynamic consolidation of VMs to a minimal number of nodes according to current resource requirements. As a result, the idle nodes can be switched off or put to a power saving mode (e.g. sleep, hibernate) to reduce total energy consumption by the data center.

The problem is to determine what kind of applications can be allocated to a single host that will provide the most efficient overall usage of the resources. Current approaches to energy efficient consolidation of workloads in data centers do not investigate the problem of combining different types of workload. These approaches usually focus on one particular workload type or do not consider different kinds of applications assuming uniform workload.

The aim of this thesis work is to consolidate the heterogeneous workloads in an efficient way so that resource utilization should be maximized and energy consumption of the data center should be minimized which will result in the less number of carbon footprints.

# Proposed Technique for Heterogeneous Workloads

---

This chapter discusses about how the problem stated in previous chapter can be solved with the help of PDL of the proposed Technique, Data Flow Diagrams and UML Diagrams.

### 4.1 Design of Solution

The solution to the problem (Heterogeneous Workload Consolidation Technique) has been designed through PDL of the proposed Technique, Data Flow Diagrams and UML Diagrams (use case, activity, sequence and deployment diagrams). Following section presents the design of the solution using PDL of the proposed Technique, Data Flow Diagrams and UML Diagrams.

#### 4.1.1 PDL of the Proposed Technique

Figure 4.1 shows the layered architecture of Cloud computing. PaaS Layer includes the Heterogeneous Workload Consolidation technique to calculate the energy consumption of the data center and also gives the information about the SLA violation, as the

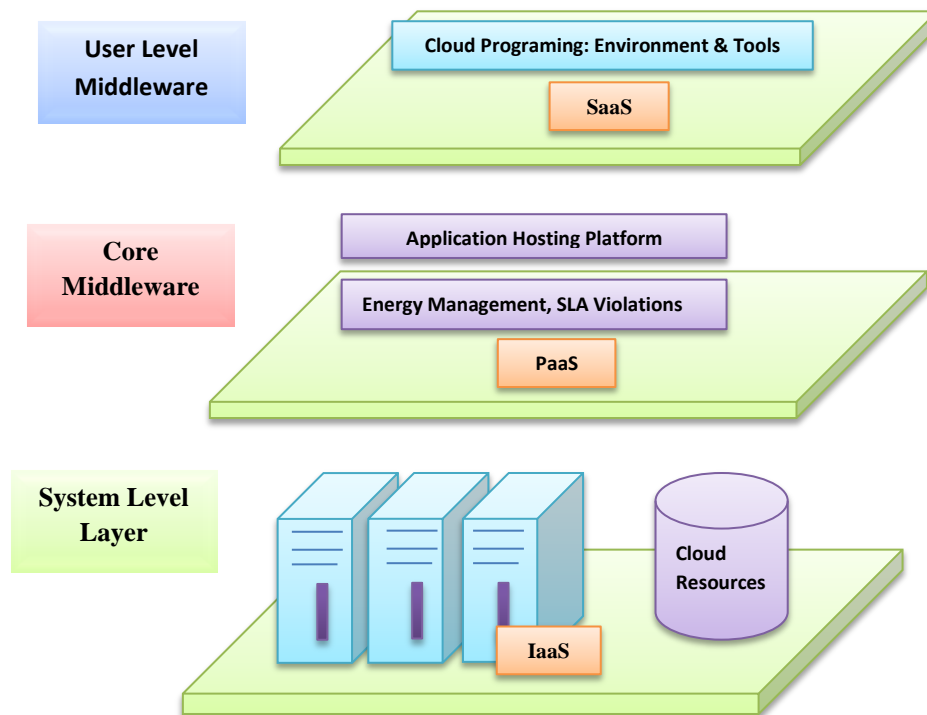


Figure 4.1: Layered Architecture of Cloud Computing

allocation policies are implemented on PaaS layer which are followed by the IaaS layer.

Below is the proposed algorithm to calculate the power consumption of the the data center by the Heterogeneous Workload Consolidation. The algorithm takes hostlist as an input and gives migration list of the VMs and power consumed by the data center. First the algorithm sorts the list of VMs in a decreasing order of CPU utilization and then it repeatedly looks for the best VM to migrate from the host and also calculate the power consumption of the data center. The algorithm stops by giving the migration list of the VM's and the power consumed by the data center.

```
1  Input: hostlist and Output: migration list, power
2  Foreach host in hostlist Do
3  Vmlist  $\leftarrow$  host.getVmlist( )
4  Vmlist.sortDecreasingUtilization( )
5  hostUtil  $\leftarrow$  host.getUtil( )
6  bestUtil  $\leftarrow$  MAX
7  minPower  $\leftarrow$  MAX
8  While hostUtil > UP_Threshold Do
9      Foreach Vm in Vmlist Do
10         if host has enough resource for Vm Then
11             Power  $\leftarrow$  estimate Power( host, Vm)
12         if Vm.getUtil( ) > hostUtil - UP_Threshold Then
13             t  $\leftarrow$  Vm.getUtil( ) - hostUtil + UP_Threshold
14             if t < betUtil Then
15                 bestUtil  $\leftarrow$  t
16                 bestVm  $\leftarrow$  Vm
17             else
18                 if bestUtil = MAX Then
19                     bestVm  $\leftarrow$  Vm
20                 Break
21     hostUtil  $\leftarrow$  hostUtil - bestVm.getUtil( )
```

```

22 migrationlist.add (bestVm)
23 Vmlist.remove(bestVm)
24 return migration list, Power

```

### 4.1.2 Data Flow Diagram

Data Flow Diagrams (DFD) of Heterogeneous Workload Consolidation Technique implemented by designing a Workload Consolidation CloudPortal(WC CloudPortal), designed for thesis is shown in Figure 4.2, which shows the context level (level 0) DFD of the WC CloudPortal with three entities- new user (want to register), member (registered user) and administrator.

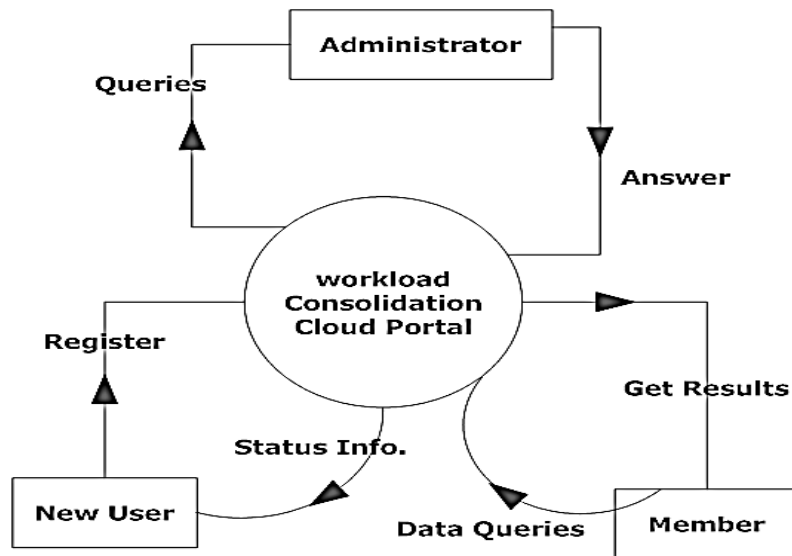


Figure 4.2: Context Level DFD

### 4.1.3 UML Diagram

Each UML diagram is designed to let developers and customers view a software system from a different perspective and in varying degrees of abstraction.

- Use Case Diagram

Figure 4.3 shows the use case diagram for WC CloudPortal. It shows that member of the WC cloud Portal can ask the query to administrator, after doing the login member can check the energy consumption of both homogeneous and heterogeneous workloads and view the graphs. Administrator can accept or reject the request of a new user and answers the query asked by the member.

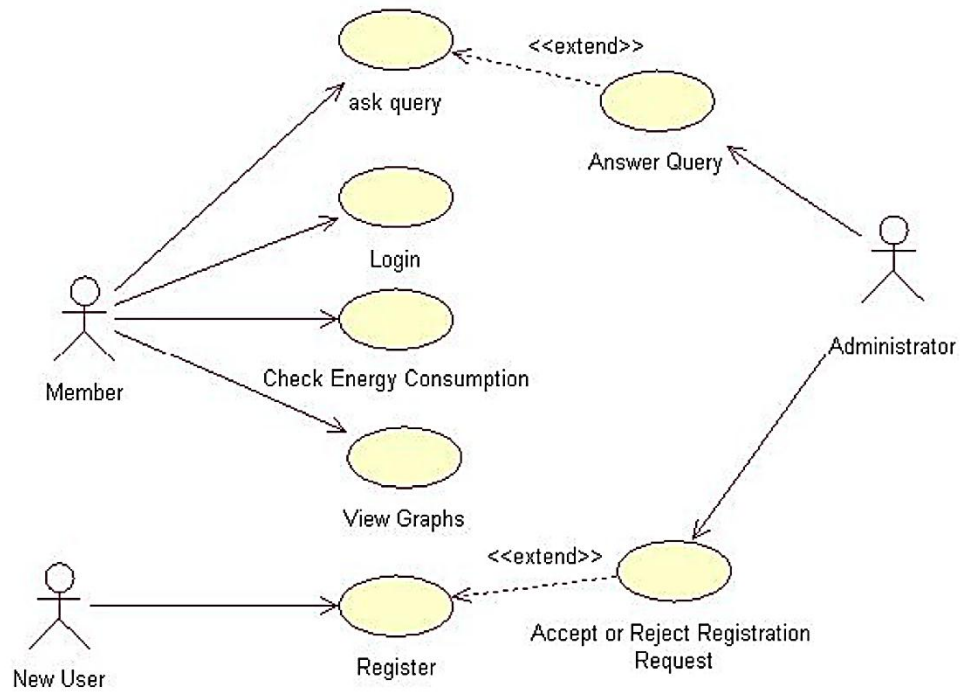


Figure 4.3: Use Case Diagram for WC CloudPortal

- Sequence Diagram

Figure 4.4 And Figure 4.5 shows sequence diagrams for WC CloudPortal for different objects- new user, member and administrator. Figure 4.4 shows the interaction between user and administrator. Figure 4.5 shows the interaction of member with server and administrator that how the member logs in to the system and checks the energy consumed, view graphs, etc.

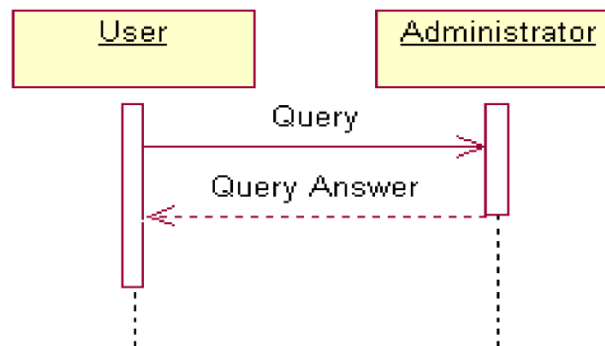


Figure 4.4: Sequence Diagram of user and administrator

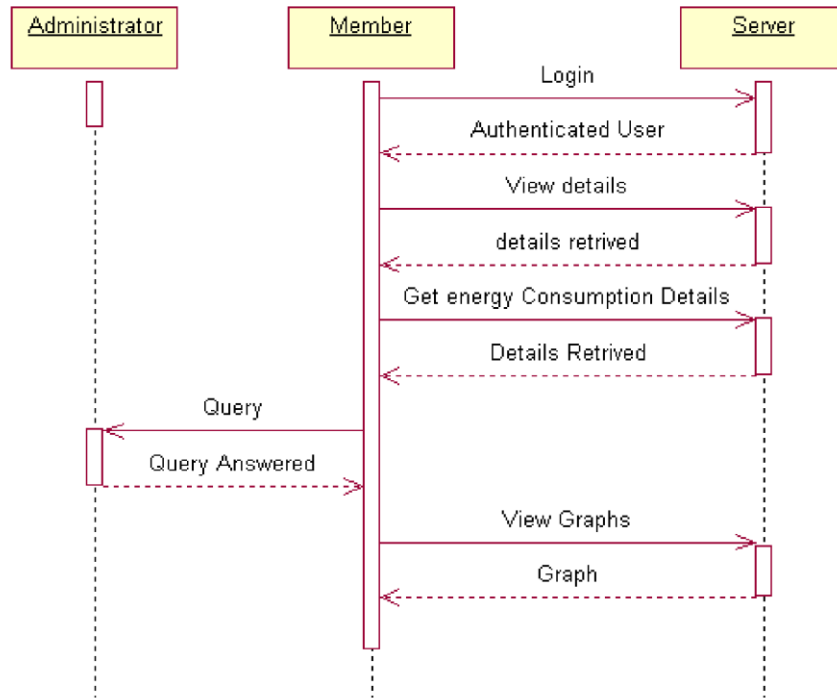


Figure 4.5: Sequence Diagram of member with server and administrator.

- Activity Diagram

Figure 4.6 and Figure 4.7 shows the activity diagram of various functions of WC CloudPortal. Figure 4.6 shows the different activities of the user and administrator for query processing. Figure 4.7 shows the different detail activities of the user to check the energy consumption by the workloads, to view the graphs and to ask the query to the administrator.

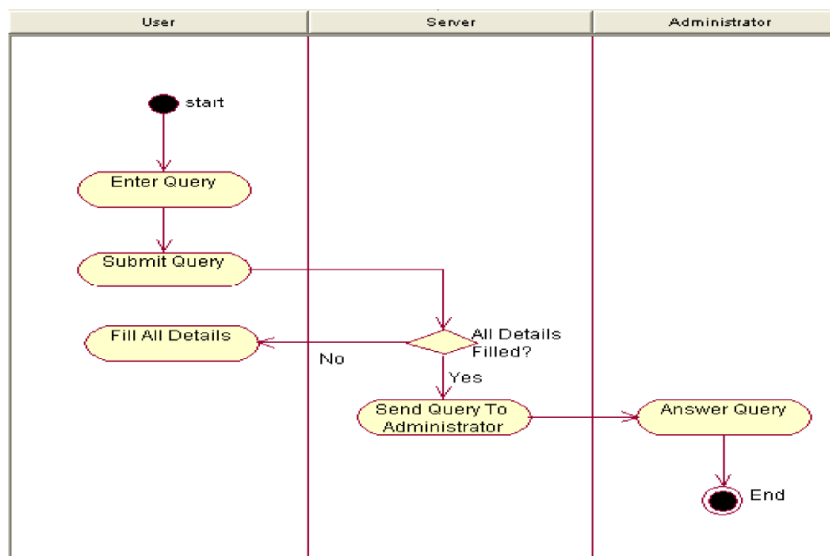


Figure 4.6: Activity Diagram for query Processing

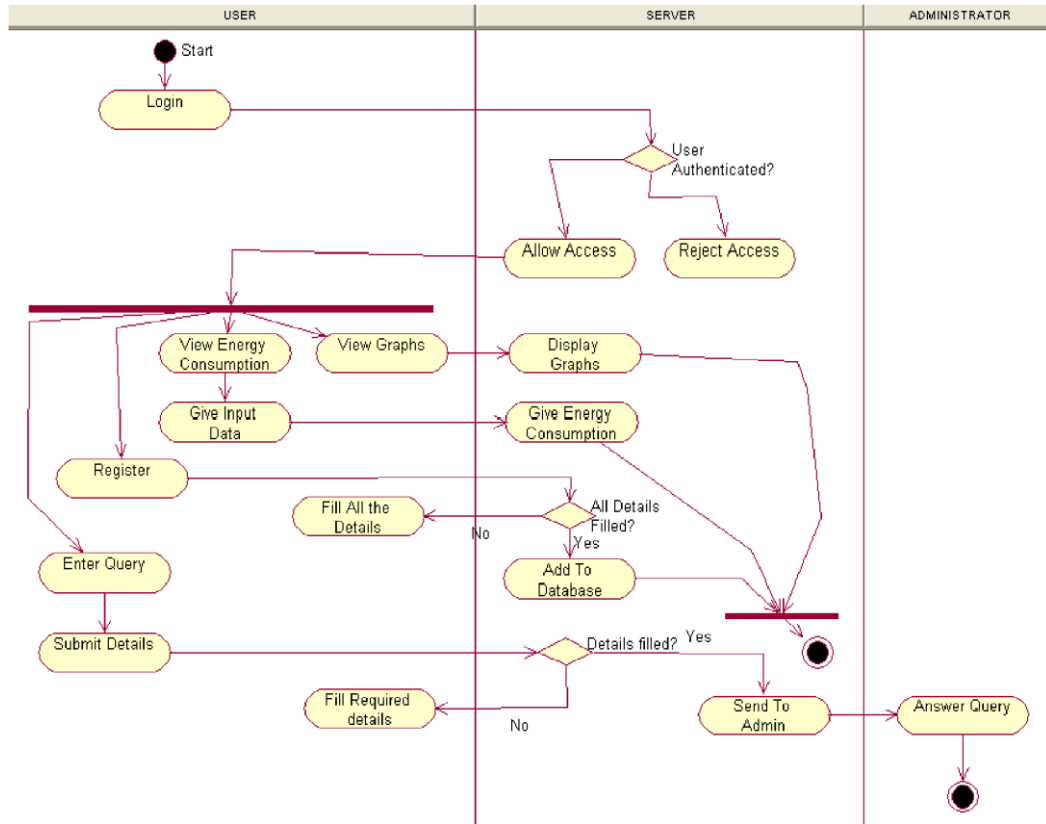


Figure 4.7: Activity Diagram showing activities after a member login.

- Deployment Diagram

Deployment Diagram models the physical deployment of artifacts on nodes. Deployment diagram depicting hardware components where Workload Consolidation CloudPortal deployed is shown in Figure 4.8.

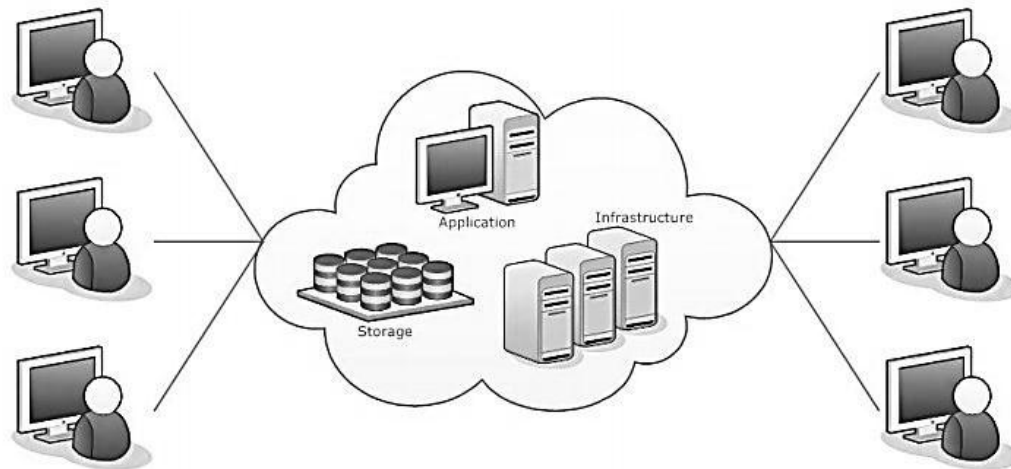


Figure 4.8: Deployment Diagram

## Chapter 5

### Experimental Results

---

This chapter focuses on tools for setting Cloud environment, implementation of heterogeneous workload consolidation technique on CloudSim Toolkit and experimental results of this approach.

#### **5.1 Tools for setting Cloud Environment**

Cloud applications have different composition, configuration and deployment requirements. So, tool required to implement the workload consolidation technique on Cloud is described below.

- **CloudSim**

CloudSim is an extensible simulation toolkit that enables modeling and simulation of Cloud computing systems and application provisioning environments. The CloudSim toolkit supports both system and behavior modeling of Cloud system components such as data centers, virtual machines (VMs) and resource provisioning policies. It implements generic application provisioning techniques that can be extended with ease and limited effort. Currently, it supports modeling and simulation of Cloud computing environments consisting of both single and internetworked Clouds (federation of Clouds). Moreover, it exposes custom interfaces for implementing policies and provisioning techniques for allocation of VMs under inter-networked Cloud computing scenarios. CloudSim offers the following novel features [20]:

- Support for modeling and simulation of large-scale Cloud computing environments, including data centers, on a single physical computing node.
- A self-contained platform for modeling Clouds, service brokers, provisioning, and allocation policies.
- Support for simulation of network connections among the simulated system elements.
- Facility for simulation of federated Cloud environment that internetworks resources from both private and public domains, a feature critical for research studies related to Cloud-Bursts and automatic application scaling.

- Availability of a virtualization engine that aids in the creation and management of multiple, independent, and co-hosted virtualized services on a data center node.
- Flexibility to switch between space-shared and time-shared allocation of processing cores to virtualized services.
- CloudSim Architecture

Figure 5.1 shows the multi-layered design of the CloudSim software framework and its architectural components. The CloudSim simulation layer provides support for modeling and simulation of virtualized Cloud-based data center environments including dedicated management interfaces for VMs, memory, storage, and bandwidth.

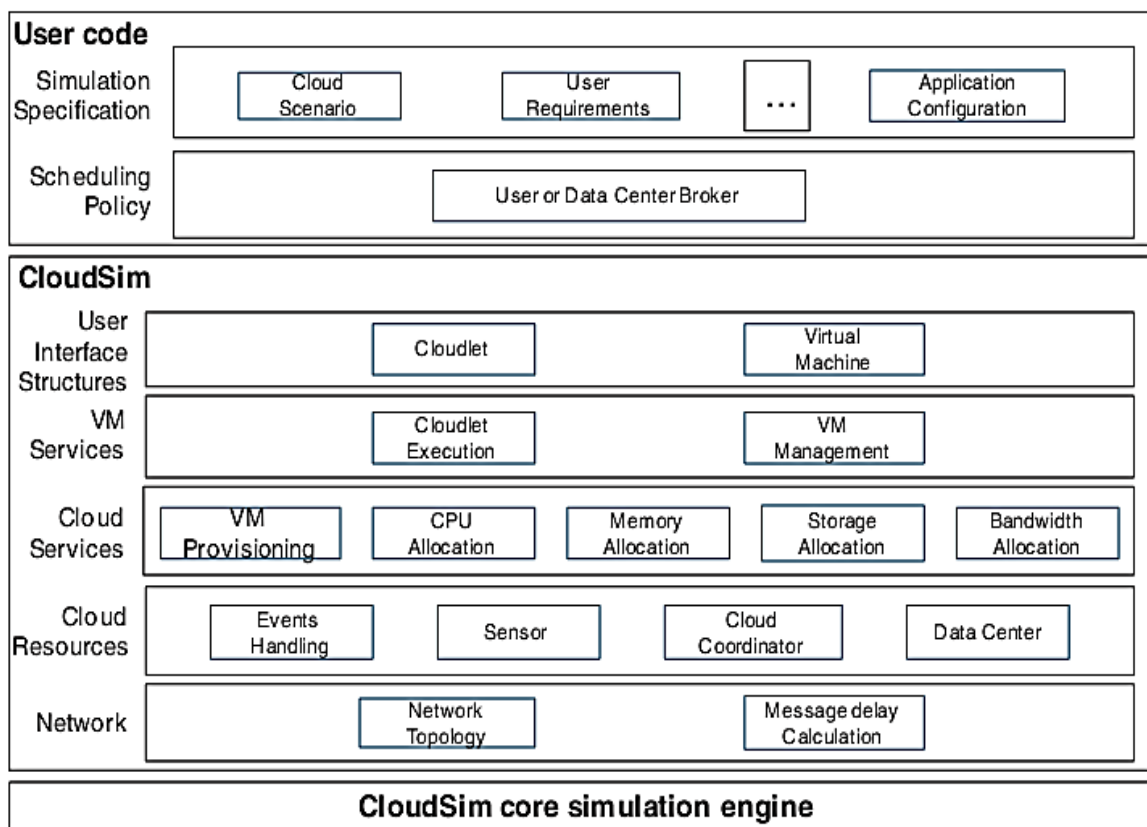


Figure 5.1: Layered CloudSim Architecture [20].

- Eclipse

Eclipse is a multi-language software development environment comprising an integrated development environment (IDE) and an extensible plug-in system [35]. It is written mostly in Java. It can be used to develop applications in Java and, by means of various plug-ins, other programming languages including Ada, C, C++, etc. The Eclipse SDK (which includes the Java development tools) is meant for Java developers. Users can

extend its abilities by installing plug-ins written for the Eclipse Platform, such as development toolkits for other programming languages, and can write and contribute their own plug-in modules. The Eclipse SDK includes the Eclipse Java development tools (JDT), offering an IDE with a built-in incremental Java compiler and a full model of the Java source files.

- Microsoft Access

Microsoft Office Access, previously known as Microsoft Access, is a database management system from Microsoft that combines the relational Microsoft Jet Database Engine with a graphical user interface and software-development tools [36]. MS Access stores data in its own format based on the Access Jet Database Engine. It can also import or link directly to data stored in other applications and databases. Access is supported by Visual Basic for Applications, an object-oriented programming language that can reference a variety of objects including DAO (Data Access Objects), ActiveX Data Objects, and many other ActiveX components.

## 5.2 Implementation of the Proposed Technique

Workload Consolidation Cloud Portal has been developed to implement the heterogeneous workload consolidation technique in CloudSim Toolkit by using Eclipse. Its basic functionalities are shown as snapshots below. Figure 5.2 shows the home page of the portal.

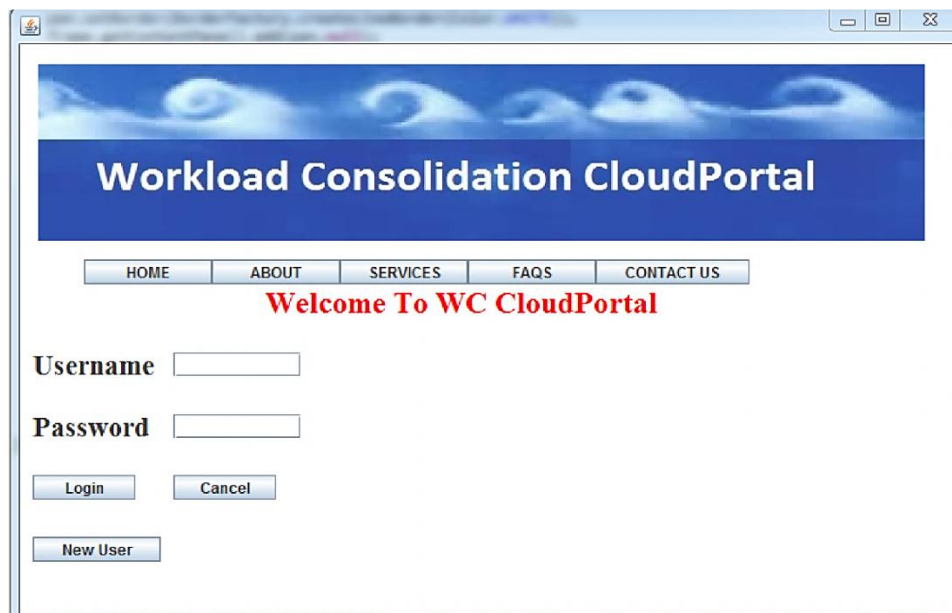


Figure 5.2: Home Page of the Portal

Figure 5.3 shows the login page where user is trying to access with wrong user name or password. Figure 5.4-5.7 shows the pages for registration to WC Cloud Portal. Figure 5.4 shows the registration page. Figure 5.5 shows the registration page with full details entered by the new user. On clicking the submit button the user gets the message of successful registration.



Figure 5.3: Login Error: user name or password do not matches

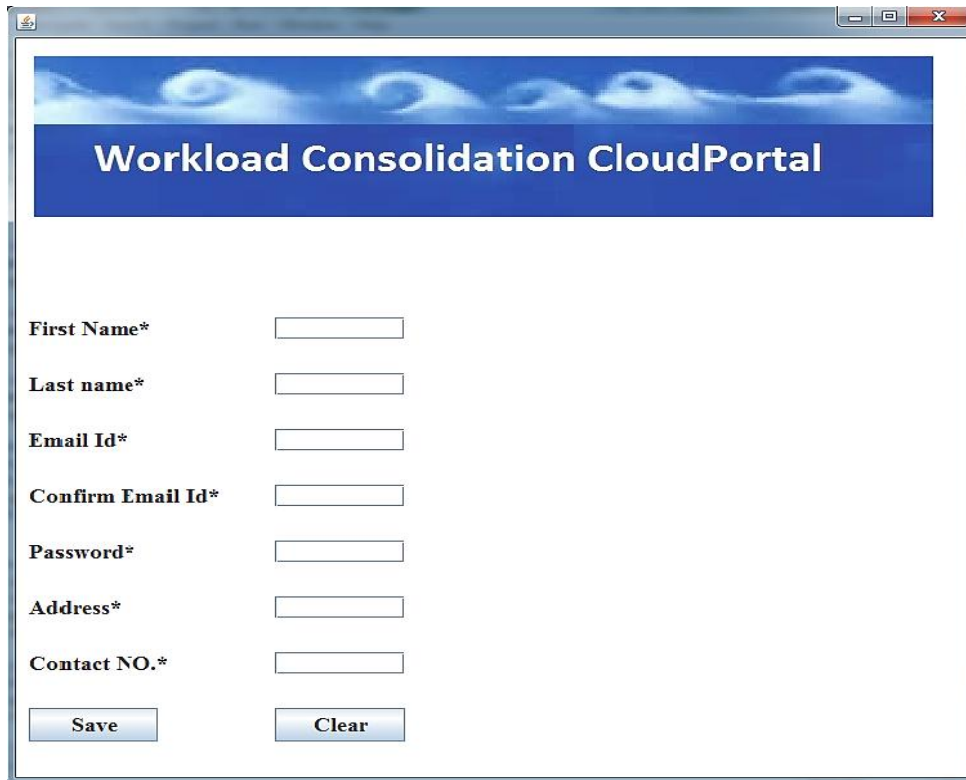


Figure 5.4: Registration Page for New User

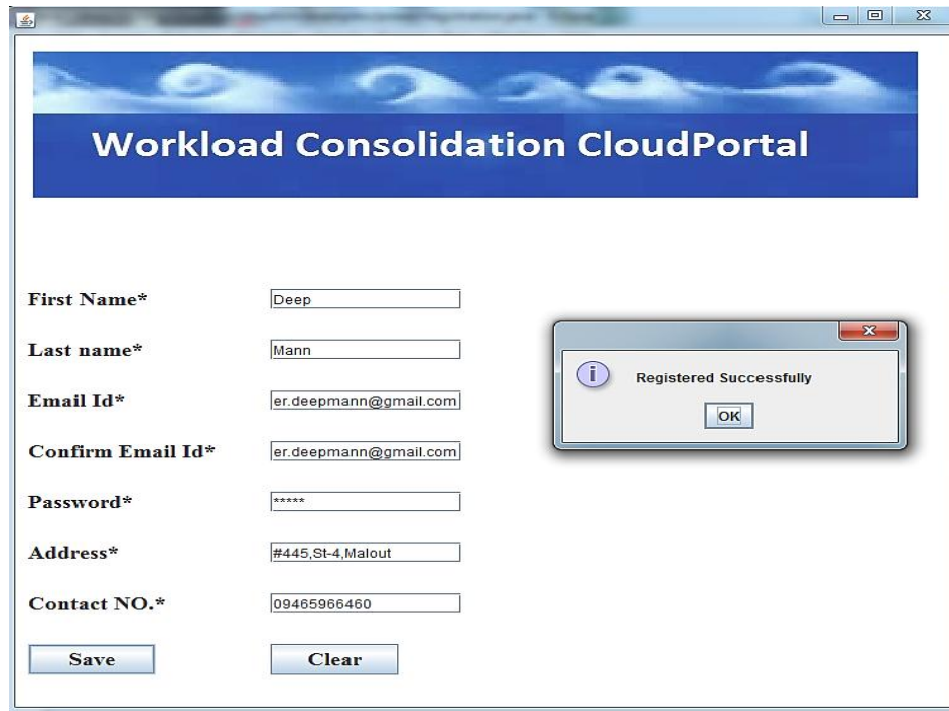


Figure 5.5: Registration done by the New User

Figure 5.6 shows the incomplete registration done by the user and Figure 5.7 shows the registration page error where user entered the wrong email address in the textbox to confirm the email id of the user.

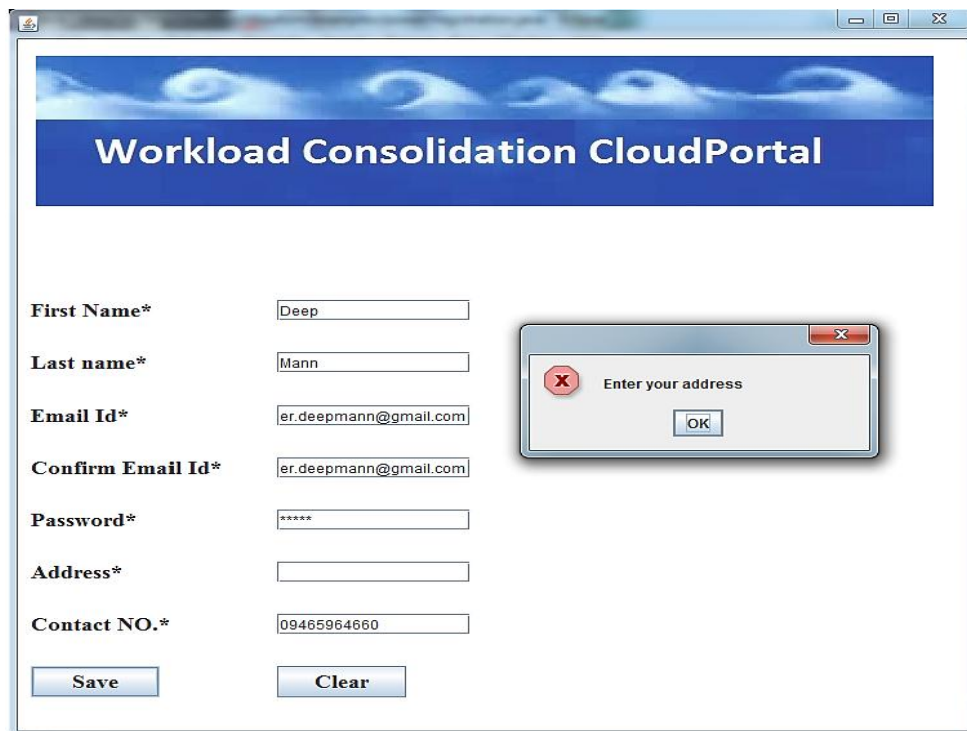


Figure 5.6: Registration Error: Required field left empty.

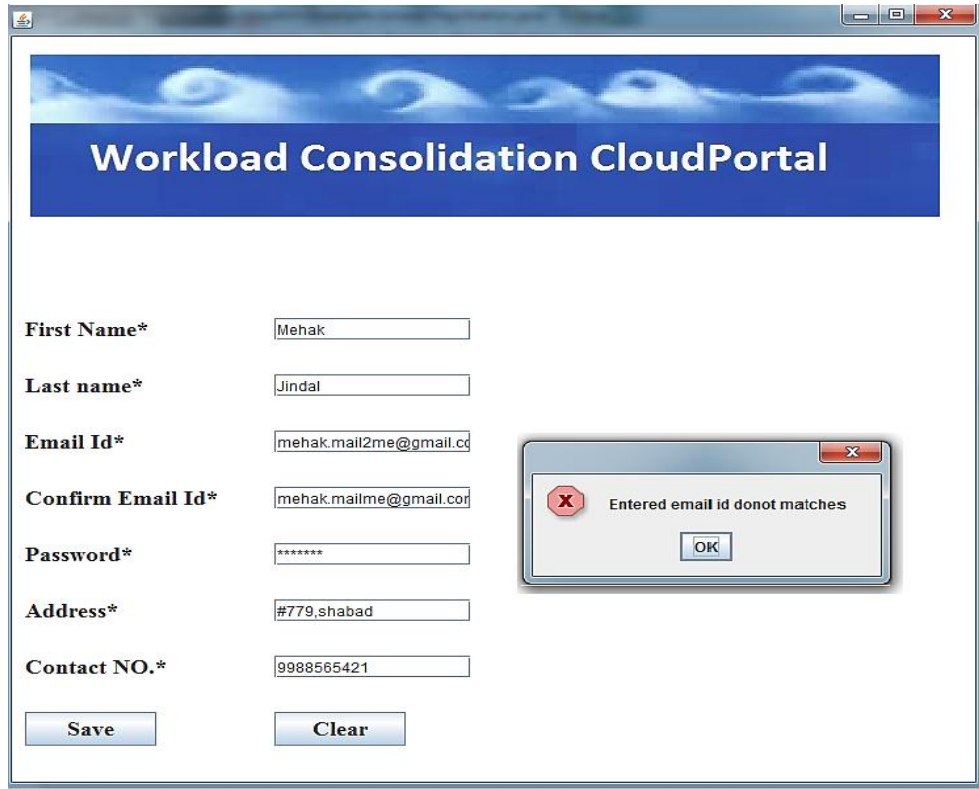


Figure 5.7: Registration Error: Email id do not matches.

Figure 5.8-5.9 shows the query pages for the users. Figure 5.8 shows the query entered by the user. On clicking submit button user gets the message that Query is saved and replied soon. Figure 5.9 shows the query error page i.e. required field left empty.

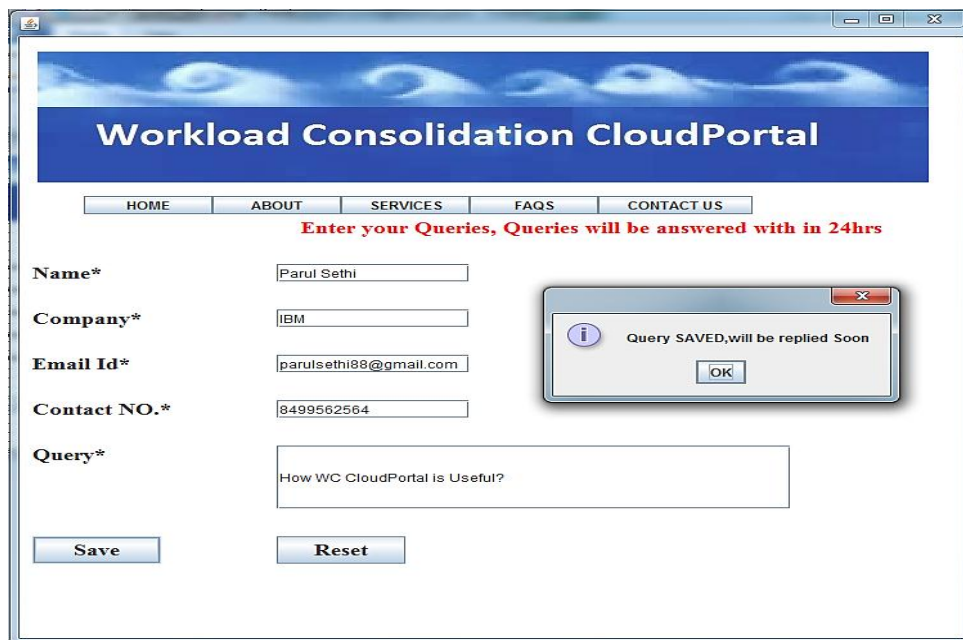


Figure 5.8: Query entered by the user.

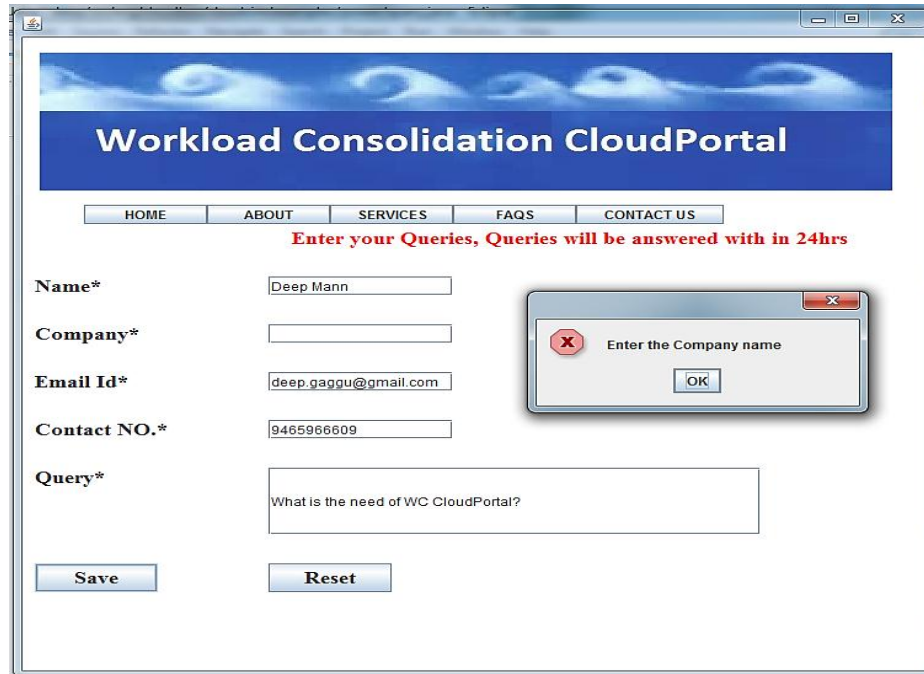


Figure 5.9: Missing required field in Query page.

Figure 5.10-5.13 shows the pages to calculate the energy consumption by homogeneous workloads. Figure 5.10 shows the page to give the inputs to calculate the energy consumed by homogeneous workloads. Figure 5.11 shows the energy consumed and other parameters by homogeneous workloads. Figure 5.12 shows the energy consumed when threshold (input parameter) is set to 50%. Figure 5.13 shows the wrong value of inputs entered by the user.

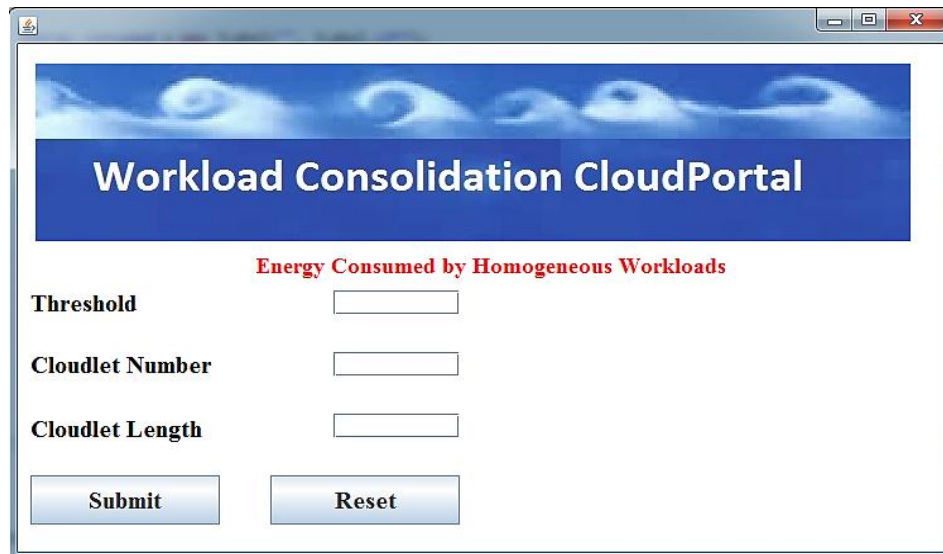


Figure 5.10: page to give the inputs to calculate the energy consumed by homogeneous workloads.

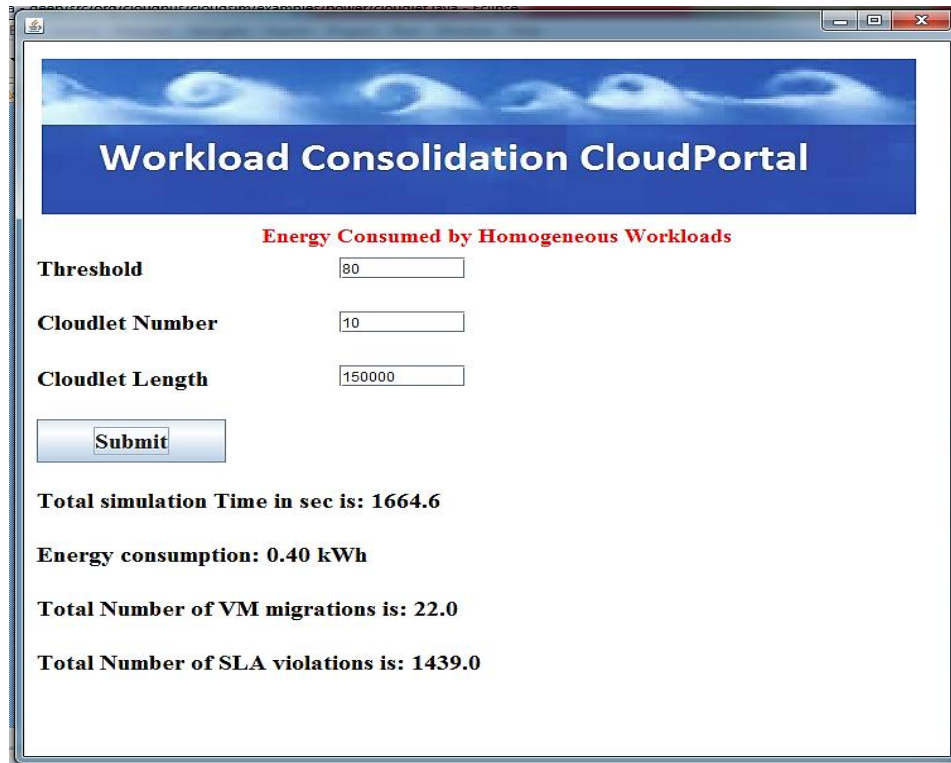


Figure 5.11: Shows the energy consumed by homogeneous workloads.

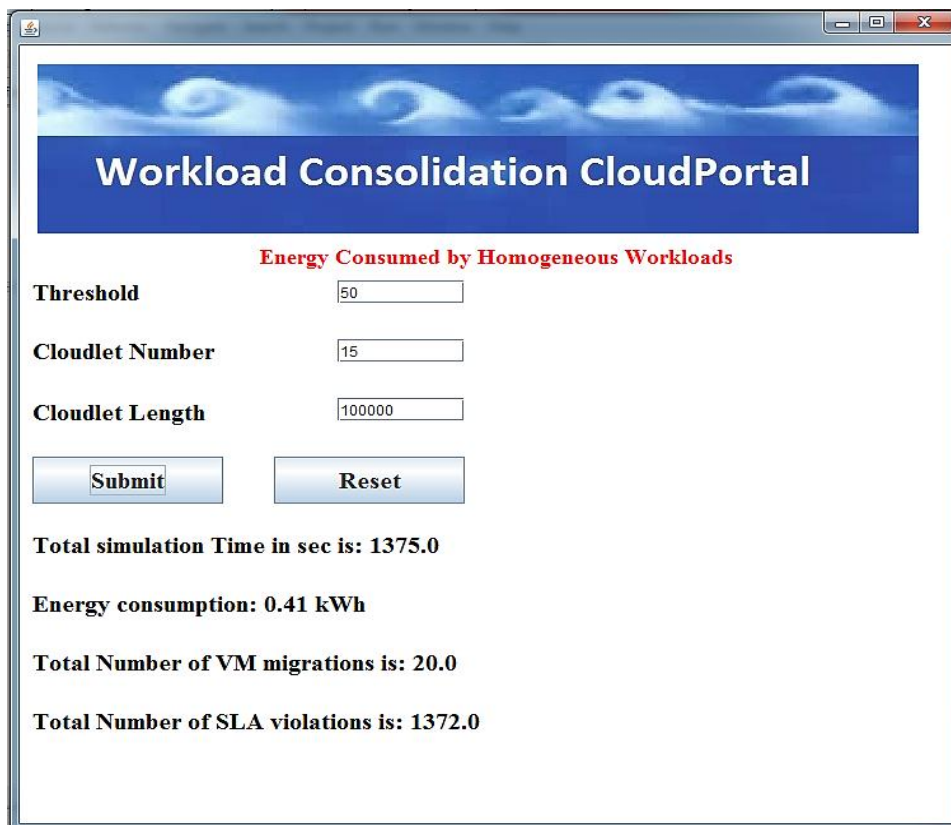


Figure 5.12 shows the energy consumed when threshold is set to 50%.

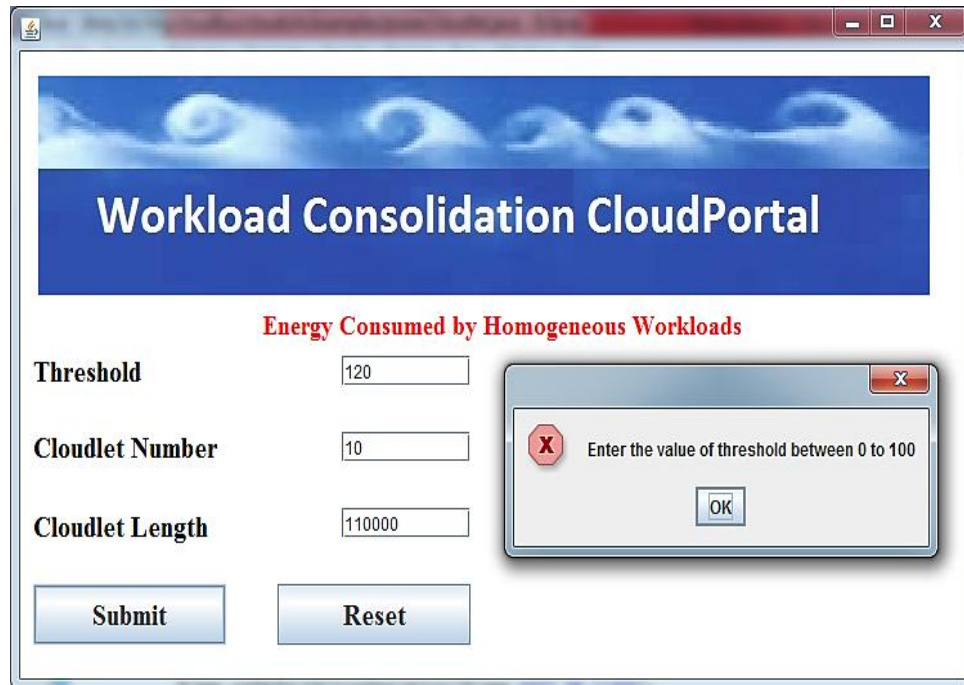


Figure 5.13 shows the wrong value of inputs entered by the user.

Figure 5.14-5.15 shows the pages to check the energy consumed by heterogeneous workloads by varying the length of the workloads (cloudlets). Figure 5.14 shows the energy consumed by heterogeneous workloads when the variability in length is 1000. Figure 5.15 shows the energy consumed by heterogeneous workloads when the variability in length is 5000.

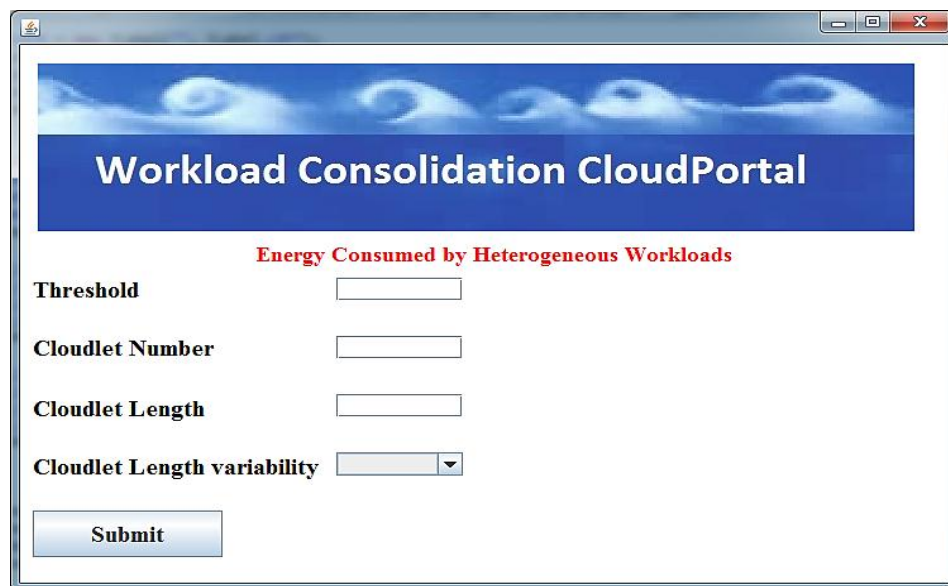


Figure 5.14: page to give the inputs to calculate the energy consumed by homogeneous workloads.

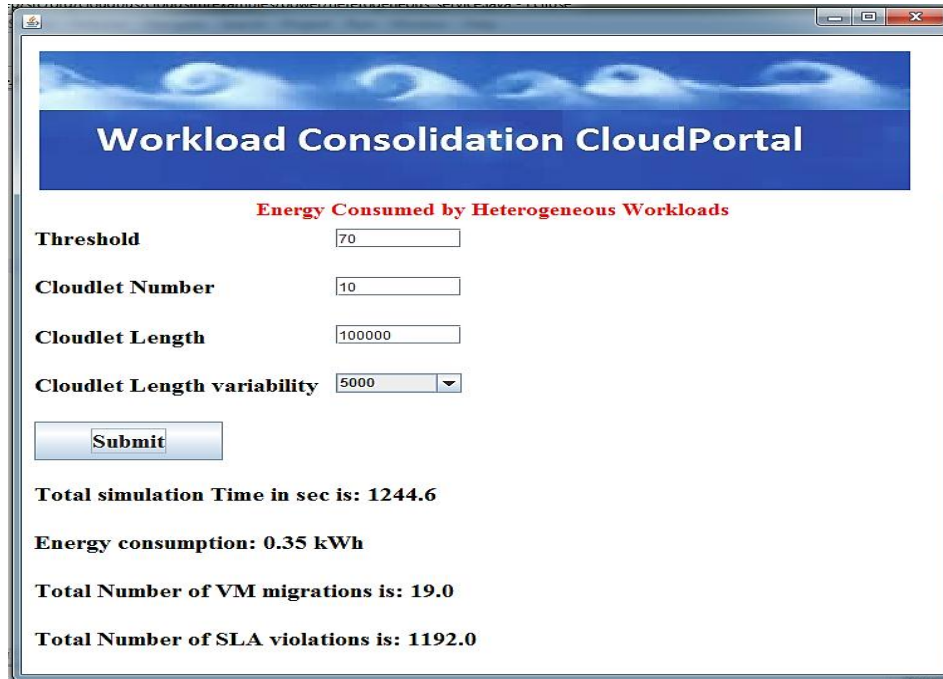


Figure 5.15: Shows the energy consumed by heterogeneous workloads when variability in length is 5000. Figure 5.16-5.17 shows the graphs of energy consumed by heterogeneous workloads. Figure 5.16 shows the graph when variability in the length of heterogeneous workloads is 1000. Figure 5.17 shows the graph when variability in the length of heterogeneous workloads is 5000.

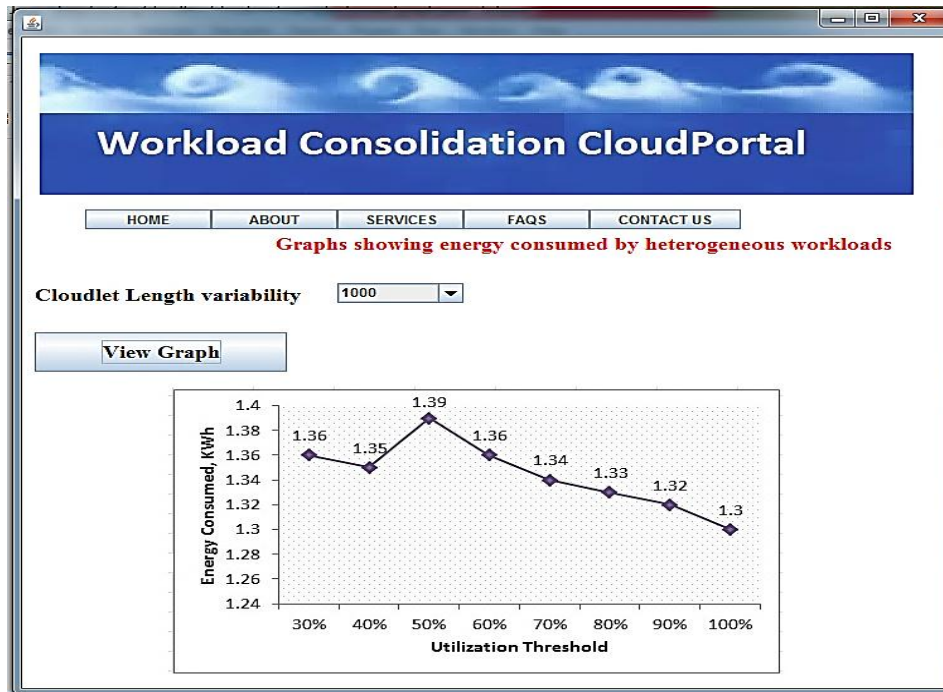


Figure 5.16 Shows the graph when variability in heterogeneous workloads is 1000.

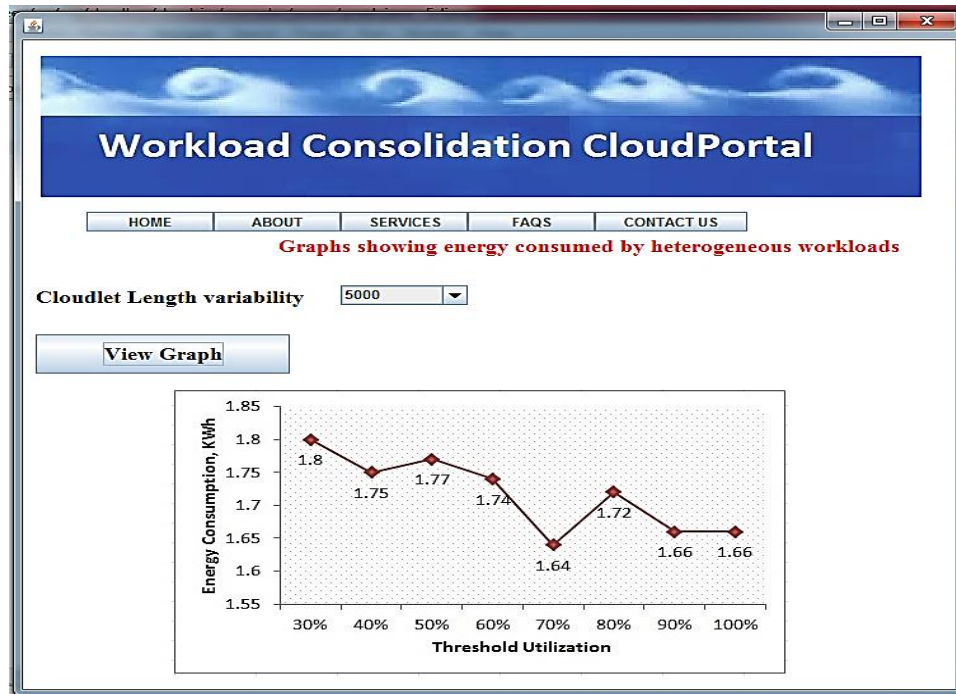


Figure 5.17 Shows the graph when variability in heterogeneous workloads is 5000.

### 5.3 Experimentation Work

- Performance Metrics

In order to get the optimal points where energy consumption in KWh of the data center is minimum we have chosen two metrics. First performance metric gives the total energy consumption of the data center by the resources due to the execution of heterogeneous workloads and the second performance metric gives the percentage SLA violations. The SLA violations occur when the virtual machine does not get the required amount of Million Instructions Per Second (MIPS).

Reference analysis has been done based on CloudSim Toolkit [13]. CloudSim provides an efficient way of performing the experiments with heterogeneous workloads and helps in calculating the energy consumption of the data center, VM migrations, SLA violations, etc.

100 physical nodes are taken to simulate data center, each having one CPU core performance with 1000, 2000, or 3000 Million Instructions Per Second, 1TB of storage and 8GB of RAM. All VM requires one CPU core with 250, 500, 750 or 1000 MIPS, 1 GB of storage and 128 MB of RAM. Each VM executes a heterogeneous workload. Total number of VMs taken is 100.

Each application is submitted to the virtual machine for its execution as per the requirement. We have taken the SPECpower benchmark, stated in the fourth quarter of 2010, that the average power consumption at 100% utilization for servers consuming less than 1000 Watt was approximately 259 W [21], to compare the energy efficiency at different threshold utilizations. The heterogeneity of the applications is based on the size (number of MIPS) of the applications. We have divided the applications into six cases as defined below.

- Case 1: The heterogeneity between the applications is 1000 MIPS. The results of the simulation as shown in Figure 5.18 indicate that the energy consumption as well as the SLA violations increases or decreases at different threshold utilizations. The optimal utilization of the host should be 90% as shown in the figures indicated. At this point energy consumption is 1.32KWh and SLA violations are minimum i.e. 5.37%, giving the minimum energy consumption with respect to the SLA violations at this threshold.

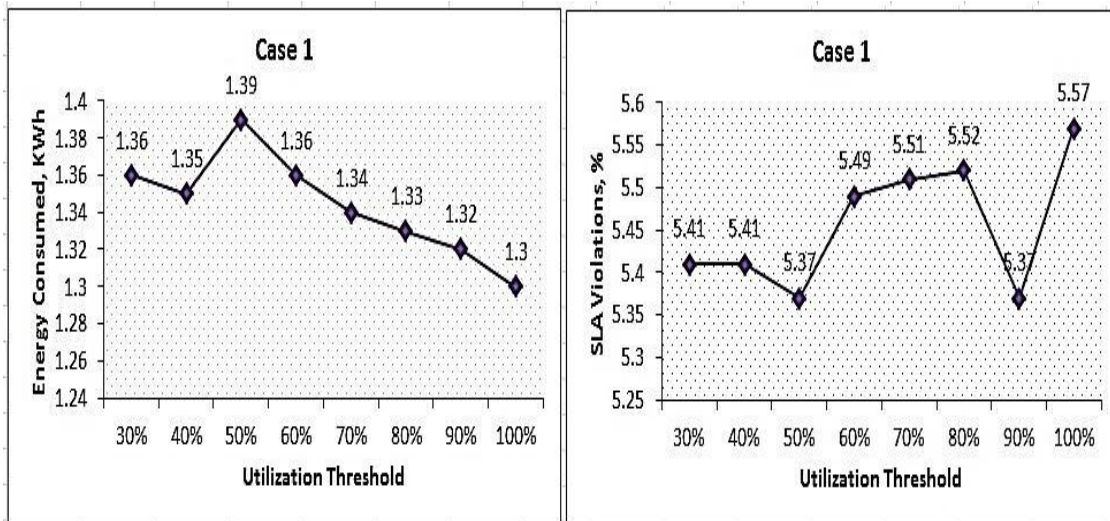


Figure 5.18: Showing energy consumption and SLA violation of case 1.

- Case 2: The heterogeneity between the applications is 2500 MIPS, as shown in Figure 5.19. In this case optimal utilization of the host should be 80% as at this point energy consumption is minimum, 1.4 KWh and SLA violations are also minimum 5.56%.

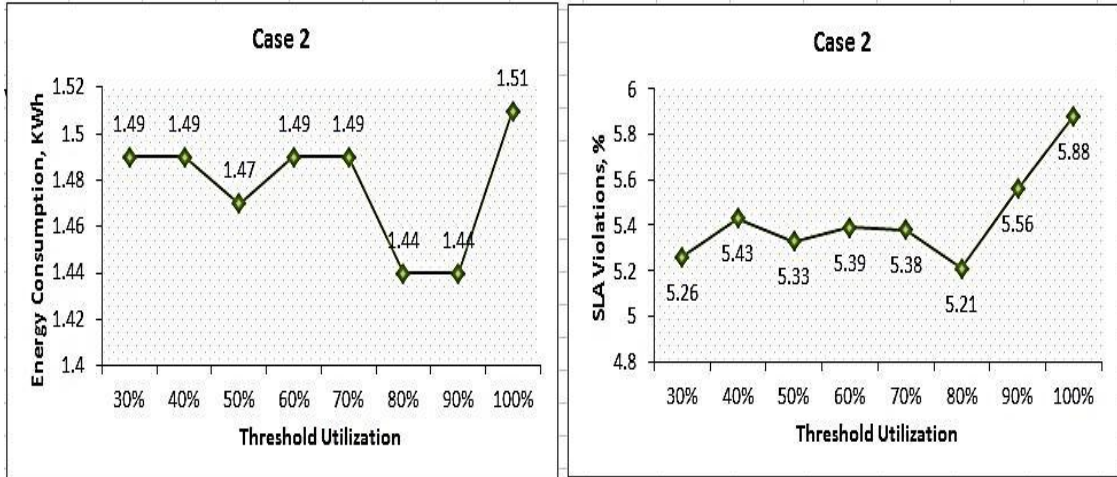


Figure 5.19: Showing energy consumption and SLA violation of case 2.

- Case 3: The heterogeneity between the applications is 5000MIPS, as shown in Figure 5.20. The energy consumption is increasing with the increase of heterogeneity of workloads. In this case optimal utilization should be 70% as at this point energy consumption is 1.64KWh and SLA violations are 4.82%.

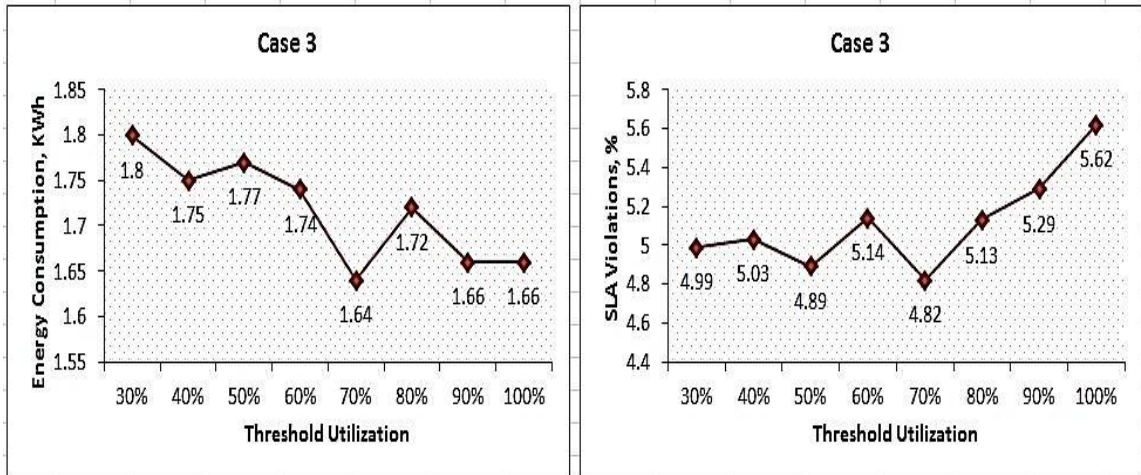


Figure 5.20: Showing energy consumption and SLA violation of case 3.

- Case 4: The heterogeneity between the applications is 7000 MIPS, as shown in Figure 5.21. In this case energy consumption decreases with the increase of utilization and the optimal utilization should be 80% as at this point both energy consumption and SLA violations are minimum, i.e. 1.91KWh and 4.64% respectively.

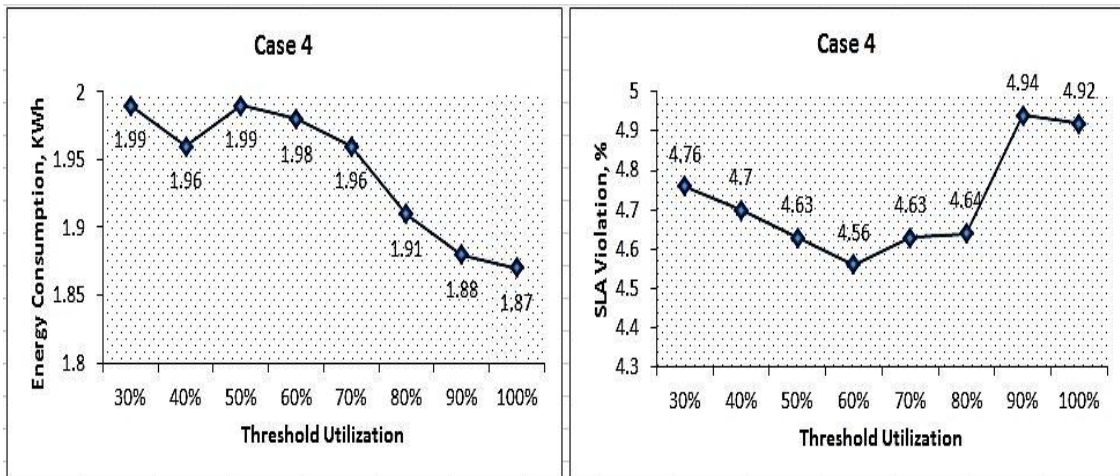


Figure 5.21: Showing energy consumption and SLA violation of case 4.

- Case 5: The heterogeneity between the applications is 20000 MIPS, as shown in Figure 5.22. In this case optimal utilization should be 70% as at this point energy consumption is minimum, i.e. 3.03 KWh and respective SLA violations are 4.02%.

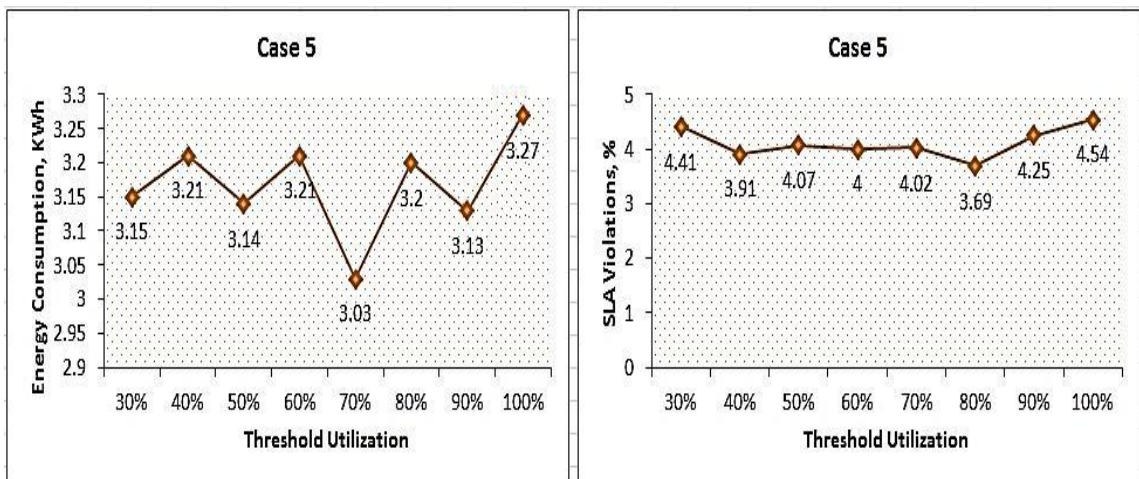


Figure 5.22: Showing energy consumption and SLA violation of case 5.

- Case 6: The heterogeneity between the applications is 50000 MIPS, as shown in Figure 5.23. In this case optimal utilization should be 70% as at this point energy consumption is minimum, i.e. 5.95 KWh as shown in graph and respective SLA violations are 3.44 %.

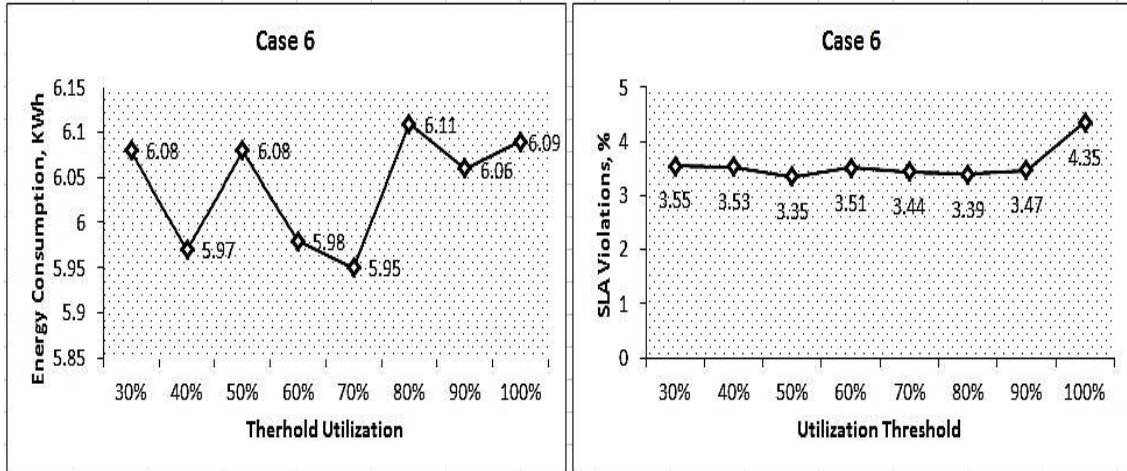


Figure 5.23: Showing energy consumption and SLA violation of case 6.

### 5.4 Results and Conclusions

It can be concluded from the results that energy consumption of the data center increases with the increasing heterogeneity of the workloads. The SLA violations too increase with the decrease in the energy consumption. The allocation of the virtual machines help to save the energy consumption as the workload will be allocated to the virtual machine having less utilization. So we can set the threshold utilization of the node according to the variability of the workloads as shown in the table 5.1.

Table 5.1: Value of Threshold according to the heterogeneity between the applications

Heterogeneity between Applications (MIPS)	Threshold Utilization (%)	Energy Consumption (KWh)	SLA Violations (%)
1000	90	1.32	5.37
2500	80	1.4	5.56
5000	70	1.64	4.82
7000	80	1.91	4.64
20000	70	3.03	4.02
50000	70	5.95	3.44

## Chapter 6

### Conclusion and Future Scope

---

This chapter discusses the conclusions of the work presented in this thesis. This chapter ends with a discussion of the future direction which can be taken further.

#### **6.1 Conclusion**

This thesis gives the introduction of Cloud computing and background of various workload consolidation techniques to manage both homogeneous and heterogeneous workloads. In this work a heterogeneous workload consolidation technique to manage the energy consumption of the heterogeneous workloads has been proposed. A workload consolidation Cloud portal has been designed and developed in this thesis work to implement the heterogeneous workload consolidation technique. Technique has been developed in java, deployed on CloudSim toolkit and Experimental results have been gathered.

#### **6.2 Thesis Contribution**

- a) In this thesis existing workload consolidation techniques have been analyzed and compared according to their features.
- b) A heterogeneous workload consolidation technique has been designed and the design of this technique has been presented through Data Flow Diagrams and UML Diagrams.
- c) The designed has been implemented and deployed on CloudSim toolkit by using Eclipse as a run time environment.
- d) Experimental results have been gathered.

#### **6.3 Future Scope**

- a) This work shows the energy consumption and SLA violations of the heterogeneous workloads. Other processing element like number of CPU's required by a cloudlet can also be considered to further increase the efficiency of workload consolidation technique.
- b) In future, dynamic resource allocation technique can also be used to manage the heterogeneous workloads and can be validated by working on Cloud environment and energy can be calculated by using Joule meter.

## References

---

- [1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing”, University of California at Berkeley, Electrical Engineering and Computer Sciences, Technical Report No. UCB/EECS-2009-28, February 10, 2009.
- [2] S. Zhang, S. Zhang, X. Chen and X. Huo, “The Comparison Between Cloud Computing and Grid Computing”, International Conference on Computer Application and System Modeling, 2010.
- [3] R. Buyya, C. Yeo, and S. Venugopal, “Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities”, in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-OB, IEEE CS Press. Los Alamitos, CA. USA), 2008.
- [4] P. Mell, and T. Grance, “The NIST Definition of Cloud computing,” National Institute of Standards and Technology, 2009.
- [5] V. R. Srinivasa, N. K. R. Nageswara and E. K. Kumari, “Cloud Computing: An Overview,” Journal of Theoretical and Applied Information Technology, 2009.
- [6] D. C. Wyld, “Moving to the Cloud: An Introduction to Cloud Computing in Government,” IBM center for The Business of Government e- Government Series, 2009.
- [7] K. Jeffery, B. Neidecker-Lutz, “The Future of Cloud Computing. Opportunities for European Cloud Computing Beyond 2010”, Expert Group Report, European Commission, January 2010.
- [8] S. Garg and R. Buyya, “Green Cloud Computing and Environmental Sustainability”, Harnessing Green IT: Principles and Practices, S. Murugesan and G. Gangadharan (eds), Wiley Press, UK, in press, accepted on April 2, 2011.
- [9] T. Dillon, C. Wu, and E. Chang, “Cloud Computing: Issues and Challenges”, in 24th IEEE International Conference on Advanced Information Networking and Applications, page no. 27-33, 2010.

- [10] V. K. Reddy, B. T. Rao, L. S. S. Reddy and P.S. Kiran, "Research issues in Cloud Computing," *Global Journal Computer Science & Technology* Vol. 11, Issue no. 11, page no. 70-76, June 2011.
- [11] Q. Zhang, L. Cheng and R. Boutaba, "Cloud Computing: state-of-the-art and research challenges", *Journal of Internet Services and Applications*, Vol. 1, Issue no. 1, page no. 7-18, Springer, 2010.
- [12] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfield, "Live migration of virtual machines," In the Proceeding of NSDI, 2005.
- [13] J. Hamilton, "Cooperative expendable micro-slice servers (CEMS): low cost, low power servers for Internet-scale services," In the Proceeding of CIDR, 2009.
- [14] B. Li, "EnaCloud: an energy-saving application live placement approach for cloud computing environments," In the Proceeding of international conference on cloud computing, 2009.
- [15] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," in *Grid Computing Environments Workshop, 2008, GCE'08*, page no. 1-10, 2008.
- [16] J. Koomey, "Estimating total power consumption by servers in the US and the world," Final report, vol. 15, February 2007.
- [17] A. Berl, E. Gelenbe, M. D. Girolamo, G. Giuliani, H. D. Meer, M. Q. Dang and K. Pentikousis, "Energy-Efficient Cloud Computing," *The Computer Journal*, Vol. 53, Issue no. 7, page no. 1045-1051, 2009.
- [18] L. Lefevre and A. C. Orgerie, "Designing and evaluating an energy efficient Cloud," *The Journal of Supercomputing*, Springer, Vol. 51, Issue no. 3, page no. 352-373, March 2010.
- [19] R. Buyya, A. Beloglazov and J. Abawajy, "Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges," In: *International conference on parallel and distributed processing techniques and applications (PDPTA)*, Las Vegas, USA, 2010.

- [20] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya, “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms,” *Software: Practice and Experience*, Wiley Press, New York, USA, 2010.
- [21] A. Beloglazov, J. Abawajy and R. Buyya, “Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing”, *Future Generation Computer Systems*, ISSN: 0167-739X, Elsevier Science, Amsterdam, The Netherlands, (in press, accepted on April 28), 2011.
- [22] R. Nathuji and K. Schwan, “Virtualpower: Coordinated power management in virtualized enterprise systems,” *ACM SIGOPS Operating Systems Review*, Vol. 41, Issue no. 6, page no. 265–278, 2007.
- [23] K. H. Kim, A. Beloglazov and R. Buyya, “Power-aware provisioning of cloud resources for real-time services,” in *Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science*, Urbana Champaign, Illinois, USA, page no. 1–6, 2009.
- [24] D. Gmach, J. Rolia, L. Cherkasova and A. Kemper, “Resource pool management: Reactive versus proactive or let’s be friends,” *Computer Networks*, 2009.
- [25] A. Verma, P. Ahuja and A. Neogi, “pMapper: power and migration cost aware application placement in virtualized systems,” in *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware*. Springer-Verlag New York, page no. 243–264, 2008.
- [26] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy and G. Jiang, “Power and performance management of virtualized computing environments via lookahead control,” *Cluster Computing*, Vol. 12, Issue no. 1, page no. 1–15, 2009.
- [27] M. Sharifi, H. Salimi and M. Najafzadeh, “Power-Efficient Distributed Scheduling of Virtual Machines using Workload-Aware Consolidation Techniques,” *Journal of Supercomputing*, An International Journal of High-Performance Computer Design, Analysis and Use, ISSN 0920-8542, Springer, DOI 10.1007/s11227-011-0658-5, Published Online: 14 July 2011.

- [28] K. Jeffery and B. Neidecker-Lutz, "The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010," Expert Group Report, European Commission, January 2010.
- [29] D. Carrera, M. Steinder, I. Whalley, J. Torres and E. Ayguade, "Managing SLAs of heterogeneous workloads using dynamic application placement," HPDC '08 Proceedings of the 17th international symposium on High performance distributed computing, New York, USA, 2008.
- [30] J. Zhan, L. Wang, B. Tu, Y. Li, P. Wang, W. Zhou and D. Meng, "Phoenix Cloud: Consolidating Different Computing Loads on Shared Cluster System for Large Organization," Proceeding Workshop Cloud Computing and Its Application (CCA '08), 2008.
- [31] C. Tian, H. Zhou, Y. He and L. Zha, "A dynamic MapReduce scheduler for heterogeneous workloads," in Proceedings of the 2009 Eighth International Conference on Grid and Cooperative Computing, IEEE Computer Society, page no. 218-224, 2009.
- [32] M. Steinder, I. Whalley, D. Carrera, I. Gaweda and D. M. Chess, "Server virtualization in autonomic management of heterogeneous workloads," Integrated Network Management, page no. 139-148, 2007.
- [33] I. Goiri, J. Fito, F. Julia, R. Nou, J. L. Berral and J. Guitart, "Multifaceted resource management for dealing with heterogeneous workloads in virtualized data centers," In 11th ACM/IEEE international conference on grid computing (Grid 2010), Brussels, Belgium, page no. 25–32, Oct 2010.
- [34] Aasys, "Virtualization Basics," Vol. 6, Issue 9, September 2008.
- [35] "Eclipse," Available at: [http://en.wikipedia.org/wiki/Eclipse\\_\(software\)](http://en.wikipedia.org/wiki/Eclipse_(software)).
- [36] "Microsoft Access," Available at: [http://en.wikipedia.org/wiki/Microsoft\\_Access](http://en.wikipedia.org/wiki/Microsoft_Access).
- [37] "Xen User Manual," [http:// bits.xensource.com/Xen/docs/user.pdf](http://bits.xensource.com/Xen/docs/user.pdf)
- [38] L. Liang, H. Wang, X. Liu, X. Jin, W. He, Q. Wang and Y. Chen, "GreenCloud: A New Architecture for Green Data Center," Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session, 2009.

- [39] R. Mata-Toledo and P. Gupta, "Green data center: how green can we perform," Journal of Technology Research, Academic and Business Research Institute, Vol. 2, Issue no. 1, page no. 1-8, May 2010.
- [40] S. Kabiraj, V. Topkar and R.C. Walke, "Going Green: A Holistic Approach to Transform Business," International Journal of Managing Information Technology (IJMIT), Vol. 2, Issue no. 3, page no. 22-31, Aug 2010.
- [41] J. Baliga, R.W.A. Ayre, K. Hinton and R.S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage and Transport," Proceedings of the IEEE, Vol. 99, Issue no. 1, page no. 149-167, Jan 2011.

## List of Papers

---

### **Published**

- [1] D. Mann and I. Chana, "Workload Consolidation Techniques in Cloud Computing: A Systematic Review," International Conference on Recent Trends in Computing, Mechatronics and Communication, at Om Institute of Technology and Management, Hisar, page no. 434 - 438, ISBN 978-81-923446-0-7, February, 2012.

### **Communicated**

- [2] D. Mann and I. Chana, "Heterogeneous Workload Consolidation for Efficient Management of Data Centers in Cloud Computing," International Journal of Computer Applications, Vol. 49, July, 2012.