Enhancing Provider's Profit on Cloud Market Infrastructure

Thesis Submitted in Partial Fulfillment of the Requirement for the Award of the Degree of

Master of Technology

in

INFORMATION SYSTEMS

SUBMITTED BY

JITENDRA SONI

(2K11/ISY/10)

UNDER THE GUIDANCE OF

N.S. RAGHAVA

ASSOCIATE PROFESSOR

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

BAWANA ROAD, DELHI-110042

(2011-2013)

I hereby certify that work which is being presented in this thesis entitled **"Enhancing Provider's Profit on Cloud Market Infrastructure"** in the partial fulfillment of the requirements for the award of the **Master of Technology** in **Information Systems** and submitted in the **Department of Information Technology** of **Delhi Technological University, New Delhi** is an authentic record of my own work carried out during the period from August 2012 to June 2013 under the supervision of **N.S. Raghava**, Associate Professor in the department of Information Technology.

The matter presented in the thesis has not been submitted by for the award of any other Degree of this or any other University/Institute.

Date: (Jitendra Soni)

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

Date:

N.S. Raghava
Associate Professor
Department of Information Technology
Delhi Technological University, Delhi

It is great pleasure to have the opportunity to extend my heartiest felt gratitude to everybody who helped me throughout the course of this thesis.

It is distinct pleasure to express my deep sense of guidance, encouragement and learned supervisor N.S**. Raghava,** ***Associate Professor,*** for his invaluable guidance, encouragement and patient review. He kept on boosting me with time, to put an extra ounce of effort to realize this work. With his continuous inspiration only, it becomes possible to complete this dissertation.

I would also like to take this opportunity my sincere regards to all the faculty members of the department for their support and encouragement.

I am grateful to my parents for their moral support all the time; they have been always around to cheer me up, in the odd times of this work. I am also thankful to my batchmatesfor their unconditional support and motivation during this work. It would be injustice if I do not mention the names of **Pradeep Kumar** and **Kislay Srivastava** – my two batchmates. Both of them helped me a lot to complete this thesis.

JitendraSoni
Roll No.: 2K11/ISY/10
Dept. of Information Technology
Delhi Technological University

Cloud computing is a computational model based on consumption of resources, offered by internet and consume its resources under requirement. A service-level agreement (SLA) is a part of a service contract where a service is formally defined. The consumption of resources is public, private, or both and has SLA that normalizes them. The infrastructure and platform services of the cloud are becoming increasingly popular all over the world but resource scheduling and allocation of multiple virtual machines on clouds still a difficult task. Most of companies are determined to reduce their computing cost through the means of virtualization. This demand of reducing the computing cost has led to the innovation of cloud computing. Cloud computing is the sum of Software as a Service (SaaS), Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). As the demand of resources from user varies with time, load on cloud generally remains low in normal hours and demand of resource remains high during peak hours. Single cloud provider may not have the resources to fulfill user's requests at the peak hour and on the contrary there may be some providers who are having under-utilized resources. To root out this limitation concept of cloud federation was invented. With this paradigm, a new algorithm that deals with the resource management in this environment is needed. Cloud federation allows outsourcing at peak time i.e. underutilized providers can give their resources on rent basis to other IaaS providers. When number of user request goes to high, single cloud provider does not work efficiently so need to research in new resource allocation algorithm which shares the common resource of multiple clouds to satisfy the customer requirement.

Cloud federation offers better computing through improved utilization and reduced administration and infrastructure costs. Federation allows a provider to dynamically outsource resources to other providers in response to demand variations. It also allows a provider that has underused resources to rent part of them to other providers.

Resource allocation and scheduling also impact in federated clouds, where resources can be bought from other members of cloud federation.

Optimization of these issues can bring the advantage of improving the energy savings and load balancing in large datacenters. When many of the cloud providers join the federation, our policies to improve the cloud providers profit shows that in which conditions cloud providers getting for profit. In this thesis use four conditions (allocation with in provider, insourcing, outsourcing, and both (insourcing and outsourcing)) and find out the results that give valuable information to cloud service provider.

This approach tries to improve the cloud providers' profit. I have used the CloudSim to find-out the results, results show that how cloud federation help to improve their profit. In this thesis enhancement for cloud provider profit is discussed and an algorithm for federated cloud environment has also been elaborated, which focuses to improve the cloud provider's profit.

LIST OF FIGURES

Fig. No	Title	Pg. No
2.1	Virtualization	5
2.2	Hosted Architecture	9
2.3	Bare-Metal (Hypervisor) Architecture	10
2.4	Para-virtualization	12
3.1	Architecture of Cloud computing	18
3.2	Cloud service models	37
3.3	Cloud federation	41
4.1	Flowchart of Proposed Algorithm	57
5.1	Layered Architecture of Cloudsim	64
5.2	Simulation Result for Profit and Revenue	70
5.3	Simulation Result for Profit and Revenue	72

List of Tables

5.1	Simulation Results for Profit and Revenue	70
5.2	Simulation Results for Profit and Revenue with special values of matrix	71

TABLE OF CONTENT

Title ***Page No.***

CERTIFICATE …….... …. (ii)

ACKNOWLEDGEMENT ……. … (iii)

ABSTRACT …… (iv)

LIST OF FIGURES ……... (vi)

Chapter 1

Introduction and Statement of the Problem

1.1 Introduction …… 1

1.2 Motivation ……..... 2

1.3 Identify the problem……... 3

1.4 Thesis Organization ……..…… 3

Chapter 2

Background and Literature Review

2.1 Virtualization Technology ……..……. 5

2.1.1 Virtual Machines ……. 6

2.2.2 Virtual Machine Monitor…….. 6

2.1.3 Virtual Machines Benefits ……..… 7

2.1.4 Aim of VM Scheduling …… ….. 7

2.2 Virtualization Approaches…….…….. 8

2.2.1 Hosted Architecture…….…….…… 8

2.2.2 Bare-Metel (Hypervisor) Architecture …….…….. 10

2.3 Para-virtualization ……..…….…….. 11

2.4 Advantages of Virtualization ……..…… 12

Chapter 3

Cloud Computing

3.1 Introduction ……. 14

3.2 Theoretical background……. ……..…….. 16

3.3 Components ……..…….. 20

3.3.1 Clients……...……..… 20

3.3.2 Datacenters ……..……..... 21

3.3.3 Distributed servers ……...…..…… 21

3.4 Characteristics……..……..… 21

3.5 Cloud computing versus Other computing technology ……..…….. 24

3.6 Issues with cloud computing ……..….. 26

3.6.1 Security, privacy and human rights issues ……..... 26

3.6.2 Economic development issues…….. 27

3.7 Cloud Computing Risks…… ... 28

3.8 Service Models…… …….. … 35

3.9 Dynamic Provisioning…… ……. 37

3.10 Federated Cloud Providers …… …….. 39

3.10.1 Cloud Coordinator (CC) …… 42

3.10.2 Cloud Broker (CB) ……. 44

3.10.3 Cloud Exchange(CE) …….. 45

3.11 Service-Level Agreement (SLA) ……. 46

3.12 Literature Review ……..… 47

3.13 Research Gaps ……..… 50

Chapter 4

Practical part

4.1 VM Provisioning ……..……...….. 52

4.2 Framework for VM Provisioning in federated cloud Environment.... 52

4.3 Benefits ……..……..... 54

4.3 Approaches of the federated cloud…….... 55

4.3.1 SpotCloud ……..…….. 55

4.3.2 OnApp……. 55

4.3.3 Tier 3……..…. 56

4.4 Flow-chart……..... 57

4.5 VM Provisioning based on Profit in Federated cloud Environment.. 58

4.5.1 Allocation within the provider……..... 58

4.5.2 Outsourcing to federated Clouds…….. 59

4.5.3 Insourcing from federate Clouds……... 60

4.5.4 Both insourcing and outsourcing in federation …….. 61

Chapter 5

Results and Discussion

5.1 CloudSim Architecture ……..…….. . 63

5.1.1 Modeling the Cloud ……..…. 65

5.1.2 Modeling the Cloud Market……... 66

5.1.3 Modeling the Network Behavior…….. 67

5.1.4 Modeling a Federation of clouds…….. ….. 68

5.2 Results for VM Provisioning based on Profit in Federated cloud Environment

…….. 69

Chapter 6

Conclusion and Scope for Future Work

6.1 Conclusion ……..… 73

Reference…….. 74

CHAPTER 1

Introduction and Statement of the Problem

- *Introduction*
- *Motivation*
- *Identify the problem*
- *Thesis Organization*

1.1 Introduction

Cloud has provided a major contribution in the field of industries, academics, and research. With increased use of cloud computing, many problems also comes under concern such as resource limit, better management of resources (resource allocation), virtual machine instance creation, VM destruction and etc. Nowadays almost all web applications starting from e-commerce to social networking have started to shift their business on cloud. But, in cloud computing technology new problems arise such as job scheduling, security, VM provisioning, load balancing, resource management, virtualization, etc. By this rapid growing use of cloud computing single cloud service provider may not be able to handle all the requests by all alone. Similarly there may be a situation in which cloud provider have many of its resources underutilized. To root out the above limitations, concept of cloud federation was invented. With this paradigm, a new algorithm that deals with the resource management in this environment is needed. The objective of this thesis is to provide a set of policies for the federate cloud environment with the aim to improve the cloud service provider's profit.

The key cloud platforms in cloud computing domain includes Google AppEngine [1], Microsoft Azure [2], Amazon Web Services [3], and GoGrid, offer many type of services for monitoring, managing and provisioning of resources and application services. All the above stated giant cloud providers are actually composed of multiple small cloud providers. These clouds enable collaborative work and sharing of computational and storage resources. All such collaborations are known as cloud federation.

Cloud computing has provided contribution to various sciences and engineering disciplines which implies that clouds mostly work in isolation and thus they do not use the resource of other

in order to fulfill any user request. Current clouds are like “islands” without resource sharing among them. This condition does not fit with the original vision of cloud computing, which imagined to be a single global infrastructure. To address this issue, there should be interoperability among several clouds providers to grant easy access from one cloud to another.

However this interoperability effort enables the sharing of resources between clouds still remains a challenging issue. It can be observed that interconnectivity and allowing resource sharing between clouds offers big opportunities and benefits. The proposed algorithm discusses the operations on different parameters that can be applied to manage the resource utilization and thus increasing the provider’s profit in cloud federation environment. There are four different scenarios for resource utilization in cloud federation environment. Present work provides the results for all the scenarios and concludes the implementation of the proposed algorithm in order to attain the maximum profit.

1.2 Motivation

Cloud has provided a major contribution in the field of industries, academics, and research. With increased use of cloud computing many problem arise, resource limit, better management of resources (resource allocation), VM instance creation, VM destruction. Now a day’s almost all web applications starting from e-commerce to social networking have been started to shift their business on cloud. By this rapid growing use of cloud computing single cloud service provider is not able handle all the requests alone. Suppose provider contain huge number of resources, the another problem arise is underutilization of resource. To root out this limitation concept of cloud federation was invented, with this new paradigm it requires new algorithm that deals with the resource management in this environment. In cloud federation, different providers running

services that have complementary resource requirements over time can mutually collaborate to share their respective resources in order to fulfill each one's demand.

1.3 Identification of the problem

The objective of the dissertation is to design a system that tackles the issues of virtual machine (VM) provisioning policies in federated cloud environment that improve the providers profit among the cloud Providers. The problem could be further subdivided into following parts:-

- To handle the issue of virtual machine provisioning policies in federated cloud environment that provide batter profit than single cloud service provider.
- Design the equations, on basis of it, profit is calculated and compare in different scenarios.

1.4 Thesis Organization

This thesis report comprises of six chapters including this chapter that introduces, the topic and states the problem. The rest of the dissertation is organized as follows.

Chapter 2 deals with the concepts of virtual machine provision in cloud computing. The virtualization in federated cloud environment is also discussed.

Chapter 3 introduces the background of cloud computing and concept of federated cloud computing, VM migration and a brief literature review of related work including research gaps.

Chapter 4 describes the framework designed for simulating profit based VM provisioning policies in federated environment and also gives the implementation details of the proposed framework and the details of experiments performed.

Chapter 5 discusses the advantages of our proposed VM provisioning techniques, validation of the framework, comparison of the results with previously existing single cloud provider techniques.

Chapter 6 concludes the dissertation work and further suggestions for future work is mentioned.

CHAPTER 2

Background and Literature Review

- ***Virtualization Technology***
- ***Virtualization Approaches***
- ***Para-virtualization***
- ***Advantages of Virtualization***

2.1 Virtualization Technology

Virtualization is software technology which uses a physical resource such as a server and divides it up into virtual resources called virtual machines (VM's). Today's computer hardware was designed to run a single operating system and a single application, leaves the resource utilization of most machines vastly underutilized. To increase resource utilization virtualization technology is used. Virtualization enables you to run multiple virtual machines on a single physical machine, with each virtual machine sharing the resources of that one physical computer across multiple environments. Different virtual machines can run different operating systems and multiple applications on the same physical computer. VMware, Microsoft Hyper-V, Virtual Iron, and Oxen are used in virtualization.

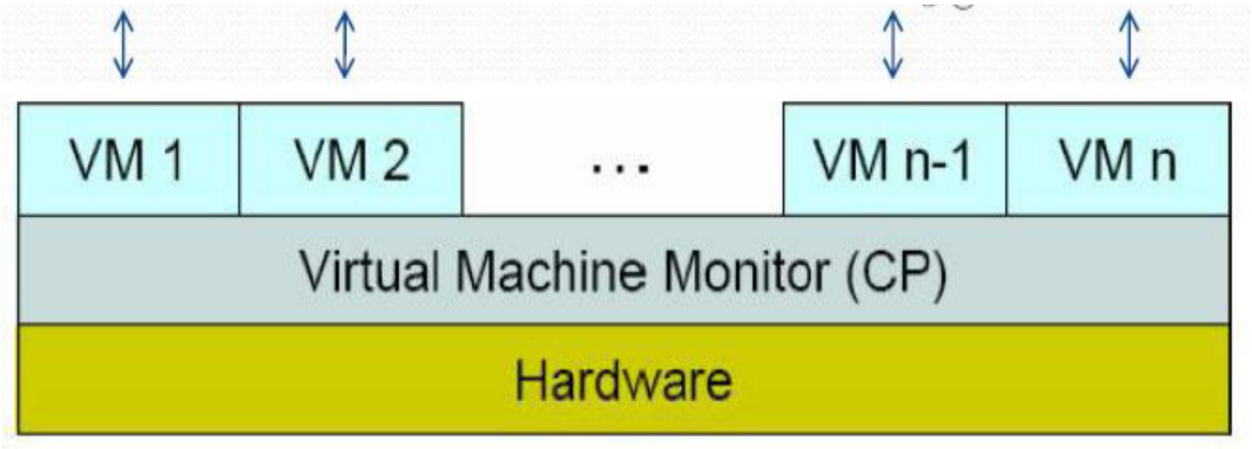

Figure 2.1:- Virtualization

Figure 2.1 illustrates the structure of a virtualization. In virtualization structure use two main components which is define below:

1. Virtual Machine (VM)
2. Virtual Machine Monitor

2.1.1 Virtual Machine(VM)

A self-contained operating environment that behaves as if it is a separate computer or a virtual machine (VM) is a software implementation of a computing environment in which an operating system or program can be installed and run. A virtual machine work almost in similar way as a physical computer and contains its own virtual (I.e., software-based) CPU, RAM hard disk and network interface card. Multiple VMs each running their own operating system (called guest operating system) are frequently used in server consolidation, where different services that used to run on individual machines to avoid interference are instead run in separate VMs on the same physical machine. An operating system cannot identify the difference between a virtual machine and a physical machine, nor can applications and about the other computers on a network. Even the virtual machine thinks it is a “real” computer. A virtual machine is composed entirely of software and contains no hardware components.

2.1.2 Virtual Machine Monitor

A virtual machine monitor is a host program that allows a single computer to support multiple, identical execution environments. All the users see their systems as self-contained computers isolated from other users, even though every user is served by the same machine. In this context,

a virtual machine is an operating system (OS) that is managed by an underlying control program. For example, IBM's VM/ESA can control multiple virtual machines on an IBM S/390 system. The key is that virtual machine monitors generally provide this service transparently; the OS above has little clue that it is not actually controlling the hardware of the machine. The key method that virtual machine monitors use to do so is to extend the notion of limited direct execution; by setting up the hardware to enable the virtual machine monitor to interpose on key events (such as traps), the virtual machine monitor can completely control how machine resources are allocated while preserving the illusion that the OS requires. VM benefits and Scheduling is described in next two sections.

2.1.3 Virtual Machines Benefits

VMware virtual machines possess four key characteristics that benefit the user:

- **Compatibility-** Virtual machines are compatible with all standard computers.
- **Isolation-** Virtual machines are isolated from each other as if physically separated.
- **Encapsulation-** Virtual machines encapsulate a complete computing environment.
- **Hardware independence-** Virtual machines run independently of underlying hardware.

2.1.4 Aim of VM Scheduling

Like any other processing unit, VMs need to be scheduled on the cloud in order to:

- Maximize utilization
- Do the job faster
- Consume less energy

2.2 Virtualization Approaches

While virtualization has been a part of the IT landscape for decades, it is only recently (in 1998) that VMware delivered the benefits of virtualization to industry-standard x86-based platforms, which now form the majority of desktop, laptop and server shipments. A key benefit of virtualization is the ability to run multiple operating systems on a single physical system and share the underlying hardware resources – known as *partitioning*. Today, virtualization can apply to a range of system layers, including hardware-level virtualization, operating system level virtualization, and high-level language virtual machines. Hardware-level virtualization was pioneered on IBM mainframes in the 1970s, and then more recently Unix/RISC system vendors began with hardware-based partitioning capabilities before moving on to software-based partitioning. For Unix/RISC and industry-standard x86 systems, the two approaches typically used with software-based partitioning are hosted and hypervisor architectures (Shown in Figure 2.2 and 2.3).

2.2.1 Hosted Architecture

VMware Workstation virtualizes I/O devices using a novel design called the Hosted Architecture. The primary feature of this design is that it takes advantage of a pre-existing operating system for I/O device support and still achieves near native performance for CPU-intensive workloads. Figure 2.2 illustrates the structure of a virtual machine in the hosted architecture.

In this architecture, the CPU virtualization is handled by the VMM. A guest application or operating system performing pure computation runs just like a traditional mainframe-style virtual machine system. However, whenever the guest performs an I/O operation, the VMM will

intercept it and switch to the host world rather than accessing the native hardware directly. Once in the host world, the VMApp will perform the I/O on behalf of the virtual machine through appropriate system calls. For example, an attempt by the guest to fetch sectors from its disk will become a read () issued to the host for the corresponding data. The VMM also yields control to the host OS upon receiving a hardware interrupt. The hardware interrupt is reasserted in the host world so that the host OS will process the interrupt as if it came directly from hardware. The hosted architecture is a powerful way for a PC-based virtual machine monitor to cope with the vast array of available hardware. One of the primary purposes of an operating system is to present applications with an abstraction of the hardware that allows hardware-independent code to access the underlying devices. For example, a program to play audio CD-ROMs will work on both IDE and SCSI CD-ROM drives because operating systems provide an abstract CD-ROM interface. VMware Workstation takes advantage of this generality to run on whole classes of hardware without itself needing special device drivers for each possible device.

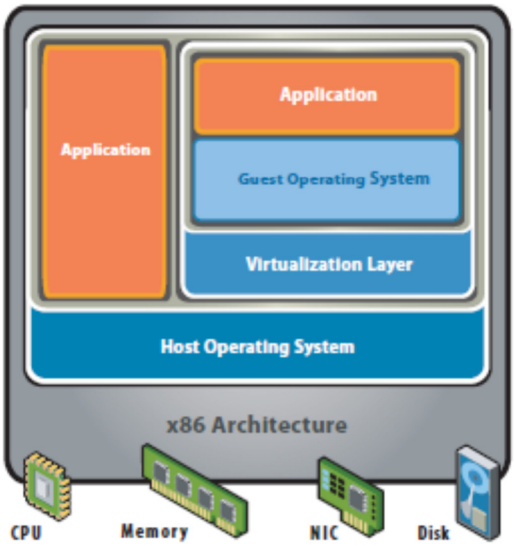

Figure 2.2:-Hosted Architecture

2.2.2 Bare-Metal (Hypervisor) Architecture

A bare-metal virtualization hypervisor does not require admins to install a server operating system first. Bare-metal virtualization means the hypervisor has direct access to hardware resources, which results in better performance, scalability and stability. One disadvantage of a bare-metal virtualization hypervisor, however, is that hardware support is typically more limited, because the hypervisor usually has limited device drivers built into it.

Bare-metal virtualization is well suited for enterprise data centers, because it usually comes with advanced features for resource management, high availability and security. Admins can centrally manage this kind of virtualization hypervisor, which is critical when you have many hosts in your virtual infrastructure. The most popular bare-metal virtualization hypervisors are:

- VMware ESX and ESXi
- Microsoft Hyper-V
- Citrix Systems XenServer
-

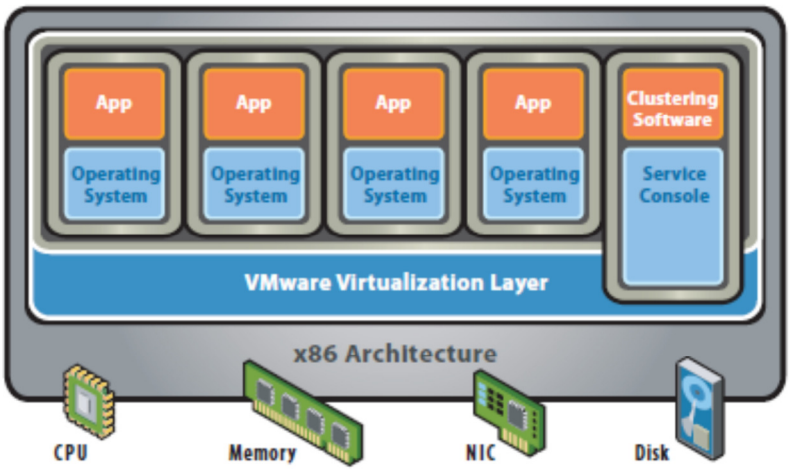

Figure 2.3:- Bare-Metal (Hypervisor) Architecture

2.3 Para-virtualization

Although virtualization is rapidly becoming mainstream technology, the concept has attracted a huge amount of interest, and enhancements continue to be investigated. One of these is Para-virtualization, whereby operating system compatibility is traded off against performance for certain CPU-bound applications running on systems without virtualization hardware assist (Shown in Figure 2.4). The Para-virtualized model offers potential performance benefits when a guest operating system or application is 'aware' that it is running within a virtualized environment, and has been modified to exploit this. One potential downside of this approach is that such modified guests cannot ever be migrated back to run on physical hardware. In addition to requiring modified guest operating systems, Para-virtualization leverages a hypervisor for the underlying technology. In the case of Linux distributions, this approach requires extensive changes to an operating system kernel so that it can coexist with the hypervisor. Accordingly, mainstream Linux distributions (such as Red Hat or SUSE) cannot be run in a Para-virtualized mode without some level of modification. Likewise, Microsoft has suggested that a future version of the Windows operating system will be developed that can coexist with a new hypervisor offering from Microsoft.

Yet Para-virtualization is not an entirely new concept. For example, VMware has employed it by making available as an option enhanced device drivers (packaged as VMware Tools) that increase the efficiency of guest operating systems. Furthermore, if and when Para-virtualization optimizations are eventually built into commercial enterprise Linux distributions, VMware's hypervisor will support those, as it does all mainstream operating systems.

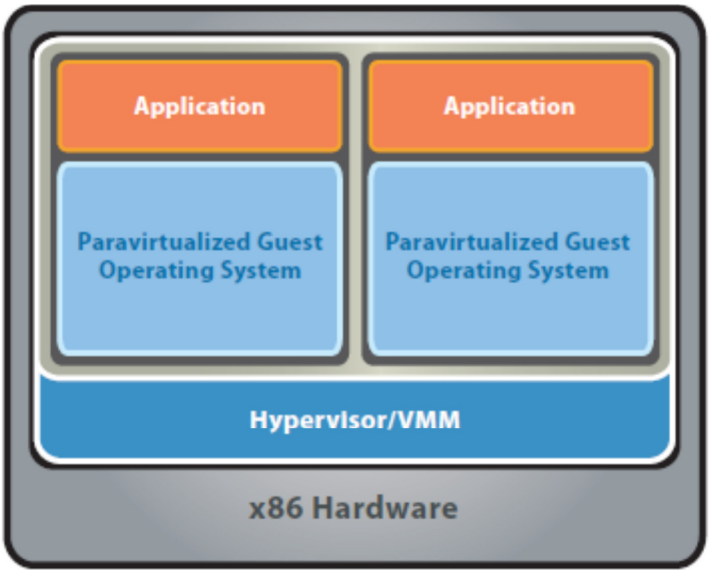

Figure 2.4:- Para-virtualization

2.4 Advantages of Virtualization

- **The most mature, proven, and comprehensive platform**: - VMware vSphere is fifth- generation virtualization—many years ahead of any alternative. It delivers higher reliability, more advanced capabilities, and greater performance than competing solutions.
- **High application availability: -** Purchased separately, high-availability infrastructure remains complex and expensive. But VMware integrates robust availability and fault tolerance right in to our platform to protect all your virtualized applications.
- **Simple, streamlined management: -** VMware lets you administer both your virtual and physical environments from a "single pane of glass" console right on your web browser. Time-saving features such as auto-deploy, dynamic patching,

and live VM migration reduce routine tasks from hours to minutes. Management becomes much faster and easier, boosting productivity without adding to your headcount.

- **Higher reliability and performance: -** Our platform blends CPU and memory innovations with a compact, purpose-built hypervisor that eliminates the frequent patching, maintenance, and I/O bottlenecks of other platforms. The net result is best-in-class reliability and consistently higher performance for heavy workloads, 2-to-1 and 3-to-1 performance advantages over our nearest competitors.
- **Superior security: -** VMware hypervisor is far thinner than any rival, consuming just 144 MB compared with others' 3-to-10 GB disk profile. Our small hypervisor footprint presents a tiny, well-guarded attack surface to external threats, for airtight security and much lower intrusion risk.
- **Greater savings: -** VMware trumps other virtualization solutions by providing 50 to 70 percent higher VM density per host—elevating per-server utilization rates from 15% to as high as 80%. You can run many more applications on much less hardware than with other platforms, for significantly greater savings in capital and operating costs.

CHAPTER 3

Cloud Computing

- *Introduction*
- *Theoretical background*
- *Components*
- *Characteristics*
- *Cloud computing vs. Other computing technology*
- *Issues with cloud computing*
- *Cloud Computing Risks*
- *Service Models*
- *Dynamic Provisioning*
- *Federated Cloud Providers*
- *Service-Level Agreement (SLA)*
- *Literature Review*
- *Research Gaps*

3.1 Introduction

Individual users in general do not have problem with posting their personal data on the internet. Actually, sometimes they even want their personal data, such as pictures, videos and blogs etc., to be shared. Companies, on the other hand are much more cautious as they seek assurance that their private data will be secured, even if allowed to access through internet by others.

People have a variety of thoughts about the cloud computing hence; it can be defined in many ways, some of them are discussed here. Several authors have proposed a definition that is centered on security, pay-per-use utility model and virtualization. According to Armbrust et al. [16] "Cloud computing can be defined as the applications delivered as services over the internet and or the hardware and system software's in the data centers that can provide as resources". According to Garner, Cloud computing is a way of computing where services are offered across the internet using several layers and models of abstraction. This definition generally match to recent developments about the cloud, where both software applications and hardware infrastructures are moved from private domain to third party data centers and enables accessibility through the internet. Buyya et al. [6] define a cloud as 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. This definition puts cloud computing into a market oriented perspective and stresses the economic nature of this phenomenon. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure provider by means of customized service level agreements.

L. Vaquero defined clouds as a large pool of services which is easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization.

One of useful thing that comes to mind because of the above definition of cloud given by Armbrust [16] and Buyya [6] is about the cloud federation, which is defined at the later stage. Cloud provides the ability to deliver both infrastructure and software as services that are consumed on a pay-per-use-basis. Past trends, like Grid computing, were limited to a certain class of users, or specific kinds of IT resources that were mostly shared on a collaborator basis. It is the approach of universal and encompasses the entire computing stack. It provides services to the mass, ranging from the e-commerce to enterprises outsourcing their entire IT infrastructure to other data centers. Service Level Agreements (SLAs), which include QoS requirements, are set up between customers and cloud providers. An SLA specifies the details of the service to be provided in terms of metrics agreed upon by all parties, and penalties for violating the expectations. SLAs act as a warranty for users, who can more comfortably move their business to the cloud. As a result, enterprises can cut down maintenance and administrative costs by renting their IT infrastructure from cloud vendors. Similarly, end-users influence the cloud not only for accessing their personal data from everywhere, but also for carrying out activities without buying expensive software and hardware.

3.2 Theoretical background

Cloud computing is a relatively new term that describes a concept which builds up naturally from previous advancements in computing technology. The concept is basically about offering computing as a service accessible on-demand over a network to cater different applications.

Service providers have to ensure that they can be flexible with their service delivery while keeping the users isolated from the underlying infrastructure (Yeo, Broberg, Venugopal, Buyya, &Brandic, 2009). As the definition of cloud computing suggests, services of cloud computing are offered through virtualized resources, which are referred to as virtual machines (VM). A VM abstracts a physical machine with the use of software and offers great flexibility to users together with other advantages. Main attributes of VM are software compatibility, isolation, encapsulation and performance.

There are three scenarios of cloud computing services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS can offer a full virtual infrastructure including storage and processing capacity which is easily accessible over the internet. PaaS offers an environment with development tools to add, change and host application software. The last scenario of SaaS offers the users access to software for which they do not need to own licenses.

Before the cloud computing one should know how computing can be done by the digital computers. The first practical microchip was developed in the late 1950s, and as the chip technology developed generations of different computers came into existence. As the computers could do more complex calculations, people started developing programs for different applications such as technical, socio and science, business applications. The first routine office

computer job is done in 1951 by the LEO computer, which was designed to deal with the overnight production requirements, payroll, inventory and other administrative tasks for J. Lyons & Co., a catering and food manufacturing company. This could be considered the first integrated management information system.

Once this has been developed people wanted to send the information from one computer to another. By this networking of computer has come into picture which roots back to the year 1969 when the creators of ARPANET (Advanced Research Projects Agency Network) envisioned the spread of "computer utilities" which would service individuals in homes and offices. The creation of the internet was the primary milestone in achieving this goal. The emergence of Web 2.0 and the widespread of broadband internet the way for cloud computing which combines features of its predecessors. As the technology advanced, so do the different paradigms promising to deliver IT as services, the architecture structure of the cloud computing is as shown in Fig. 3.1. The predecessors of cloud computing are for example Grid computing, Peer-to-Peer computing, Services computing and Market-oriented computing.

Individual internet users are used to cloud computing services in the form of SaaS. The most recognized examples are e-mail services and other web applications like the ones from Google for example Social networks are also good examples of cloud computing. These applications are accepted well in general, and users do not have problems with storing their personal information in these clouds, which is proved by the number of active users of Gmail or facebook. Businesses on the other hand have serious security concerns when it comes to using cloud services. Most of these concerns are related to the diminishing information system boundaries, while others are related to the availability of the service. This creates various business process risks which need to be tackled.

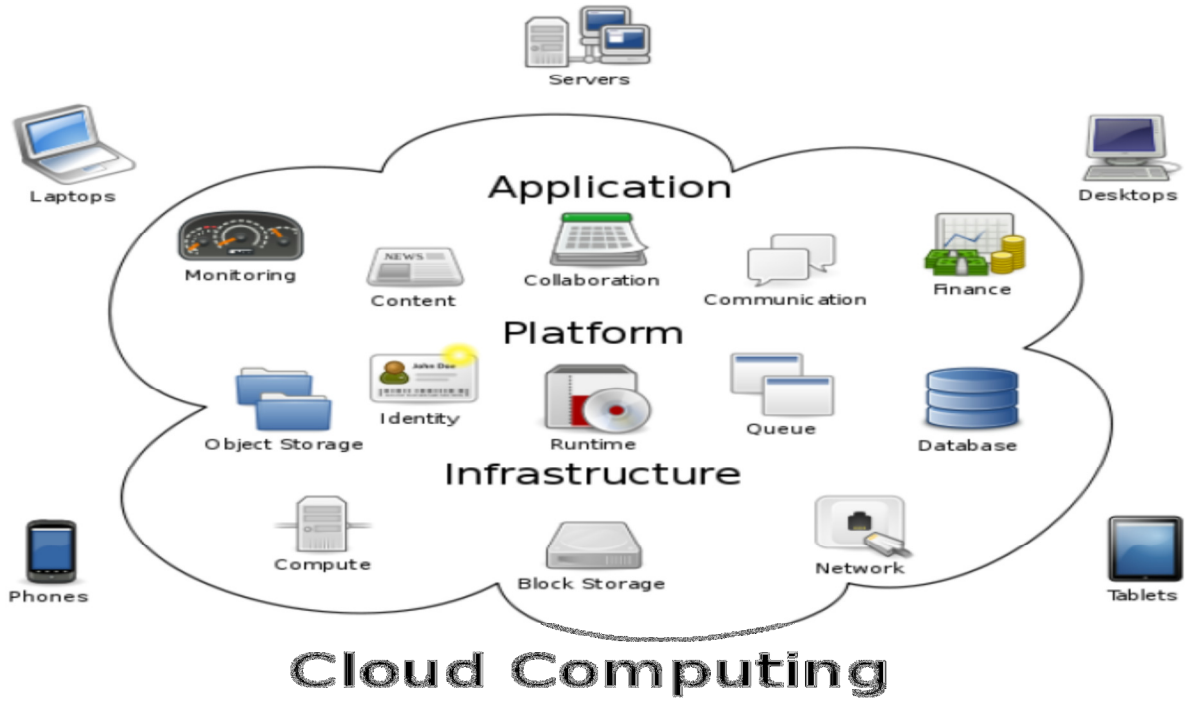

Figure 3.1:- Architecture of Cloud computing

Security is usually referred to as a combination of three factors: confidentiality, integrity and availability. What is more, cloud computing raises three more factors, namely compliance, policy and risk (Hobson, 2009). For companies, sending confidential data over the internet and storing them in a cloud which is shared with other users is a problem. They need a guarantee that no one else will have access to their data and that integrity will be maintained. Since it is usually possible to access the cloud only by internet, availability is vulnerable in case of a connection outage. Companies are required to have alternative connections to the internet at their disposal at all times to ensure availability of access to the services. There are also problems with compliance requirements. For example, companies that handle their customers' credit card numbers need to fulfill with the PCI standard. There are also sector-specific regulations such as HIPPA for

healthcare or FISMA and ITAR for organizations in the public sector. Naturally, there can be a mixture of regulations that need to be satisfied. Ensuring that a company's services will be managed correctly when they do not own the infrastructure of the data center is therefore a challenge. A company's data storage policy must not only address these compliance issues, but there are also many legal ones. The physical location of the data center that holds the data may be important. Local laws may threaten their security, since laws like the US Patriot Act allow the government to have virtually limitless access to any kind of information. On the other hand, the EU is in favor of much more strict protection of privacy. The last security factor of cloud computing is the risk of the service provider to stop doing business. In that case, the users need to know if it would be possible to transfer all their data and how long would this take. It is clear that even if the service provider can give reasonable assurance for confidentiality, integrity and availability, due to compliance, data storage policy and risk of service discontinuity, not all data is suitable for the cloud. An IS boundary also known as a perimeter contains all of the information assets in an IS (Raval&Fichadia, 2007). The problem of controlling an IS boundary is not new, although it is still a challenge. Cloud computing cascaded within the group of distributed systems. This means that instead of inventing completely new security solutions, one can learn from previous studies of this kind of system, even though cloud computing is a relatively new concept. Distributed computing technologies follow a similar pattern of interaction, where disparate and sometimes heterogeneous systems interact with one another over a common communication platform (Belapurkar, Chakrabarti, Ponnapalli, Varadarajan, Padmanabhuni, &Sundarrajan, 2009). Since the security measures at the system boundary cannot bring sufficient assurance, it is necessary to go deeper into the IS. The company makes sure that the data on the cloud is secured as well as in the internal network. Consequently, one has to

address security at the data level. Data flowing in a network needs to have incorporated information on how it should be treated. Companies have various policies for handling this kind of security and the most common solutions are encryption, access controls, rights management and auditing. The policies do not approach all the data in the same way, and define different levels of confidentiality. It would be too costly and inconvenient to treat all data as most confidential due to the tradeoff between security, usability and costs (Raval&Fichadia, 2007). In that case, using cloud services would be impossible, moreover the IS itself would be probably isolated from the internet completely. In the next section different components in cloud computing is discussed as they are the ones which can store, communicate and distribute the required data.

3.3 Components

There are three major components of cloud computing –

- Clients
- Datacenters
- Distributed servers

3.3.1 Clients

Clients are the end users of cloud services. They can be further divided into thick, thin and mobile clients. The most common are thick clients which are normal standalone computers connected to the network, while thin clients are computers that only have internal memory, because they access all the data from the servers remotely. Mobile clients can be laptops, PDAs or various types of Smartphone's. It is necessary to differentiate between the types of clients, because they propose various security challenges.

3.3.2 Datacenters

Datacenters are large rooms with physical machines that provide storage and processing power. They house the applications that are accessed as services. Because of the decreasing prices of computers, datacenters of cloud computing providers are usually composed of hundreds to thousands of commoditized computers with specialized processors that have hardware support for virtualization.

3.3.3 Distributed servers

The servers to which the clients connect appear as if they are located in one datacenter, but in reality they are usually geographically distributed over various locations. Generally, the clients do not need to care about the complexity of the infrastructure that hosts their services. This is usually hidden from the client, but in some cases the geographical location of the server is important.

3.4 Characteristics

- **On demand self-services**: Computer services such as email, applications, network or server service can be provided without requiring human interaction with each service provider. Cloud service providers providing on demand self-services include Amazon Web Services (AWS), Microsoft, Google, IBM and Salesforce.com. New York Times and NASDAQ are examples of companies using AWS (NIST).
- **Broad network access**: Cloud Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms such as mobile phones, laptops and PDAs.

- **Resource pooling**: The provider's computing resources are pooled together to server multiple consumers using multiple-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. The resources include among others storage, processing, memory, network bandwidth, virtual machines and email services.
- **Rapid elasticity**: Cloud services can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
- **Measured service**: Cloud computing resource usage can be measured, controlled, and reported providing transparency for both the provider and consumer of the utilized service. Cloud computing services use a metering capability which enables to control and optimize resource use. This implies that just like air time, electricity or municipality water IT services are charged per usage metrics – **pay per use**. The more you utilize the higher the bill. Just as utility companies sell power to subscribers, and telephone companies sell voice and data services, IT services such as network security management, data center hosting or even departmental billing can now be easily delivered as a contractual service.
- **Empowerment:** In Cloud computing end-users of computing resources has its own control on available resource; in traditional computing system control is in hand of a centralized IT service.

- **Application programming interface** (API): Cloud computing provide accessibility to software that enables machines to interact with cloud software in the same way the user interface facilitates interaction between humans and computers.
- **Cost**: Cloud computing use pay as you go model, so user does not need to pay whole money to buy the costly resource, maintain datacenters. In old computing scenario user has to maintain data center with large number of powerful and is claimed to be reduced and in a public cloud delivery model capital expenditure is converted to operational expenditure.
- **Device and location independence:** Cloud computing enable users to access systems using a web browser regardless of their location or what device they are using like PC and mobile phone. As infrastructure is typically provided by a third-party and accessed via the Internet, users can access resource from anywhere at any time.
- **Virtualization:** This technology allows servers and storage devices to be shared and utilization is increased. Applications can be easily migrated from one physical server to another in order to manage the load balancing among the datacenter.
- **Reliability:** It is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.
- **Performance:** In Cloud performance is monitored and consistent and loosely coupled architectures are constructed using web services as the system interface.
- **Security:** Cloud security is one the challenging issue, could improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is

often as good as or better than other traditional systems, in part because providers are able to devote resources to solving security issues that many customers cannot afford.

- **Maintenance:** Cloud computing applications are easier to manage, because they do not need to be installed on each user's computer and can be accessed from different places.

3.5 Cloud computing versus Other computing technology

Cloud computing involves the use of services on the internet rather than local computers while grid computing involves sharing of tasks over multiple computers. Resources of multiple computers are shared in grid computing which greatly helps in improving the flexibility and power of the network whereas this not the case with cloud computing. Applications like spreadsheets, presentations, email and word processors are part of cloud computing whereas in grid computing, data storage or complex calculations are done.

Cloud computing is a technology that delivers many kinds of resources as services, mainly over the internet, while distributed computing is the concept of using a distributed system consisting of many self-governed nodes to solve a very large problem (that is usually difficult to be solved by a single computer). Cloud computing is basically a sales and distribution model for various types of resources over the internet, while distributed computing can be identified as a type of computing, which uses a group of machines to work as a single unit to solve a large scale problem. Distributed computing achieves this by breaking the problem up to simpler tasks, and assigning these tasks to individual nodes.

This difference between cloud computing and utility computing is substantial; since it reflects a difference in the way computing is approached. Utility computing relies on standard computing practices, often utilizing traditional programming styles in a well-established business context.

Cloud computing, on the other hand, involves creating an entirely distinctive virtual computing environment that empowers programmers and developers in new ways. Even normal business computing tasks can look drastically different through these two computing styles. One such example is in customer relationship management (CRM). This routine task involves the storage and use of client information, including contact details, contract specifics, and other related content. Through utility computing, businesses can easily maintain a traditional approach to CRM, and even companies that lack resources to invest heavily in infrastructure and software can still have a booming CRM program. This can be especially powerful for up and coming businesses, which may lack the capital needed to develop their own infrastructure but still need a way to maintain their thriving clientele base. Through cloud computing, CRM can look radically different. While there is still less upfront cost through cloud computing, the approach to maintaining CRM changes drastically. The way the information is filed and accessed is enhanced through cloud computing, making the process faster and more accessible overall. All of this can be accomplished without any specific understanding of the technology that supports this interface, which allows for all attention to be diverted to the CRM processes themselves.

Cloud computing is a technology that delivers many kinds of resources as services, mainly over the internet, while cluster computing focuses on improved performance and availability of a service by interconnecting a collection of stand-alone machines to form a single integrated computing resource. Clusters are mainly used for load balancing and providing high availability, whereas cloud computing focuses on providing services such as software, platforms, etc. But one important thing to note is that cloud computing is build based on a server cluster.

3.6 Issues with cloud computing

Computing on the cloud requires vigilance about security, manageability, standards, governance, and compliance. Some issues with cloud computing illustrate below:

3.6.1 Security and privacy issues

When using cloud-based services, one is entrusting their data to a third-party for storage and security. Can one assume that a cloud-based company will protect and secure ones data (back it up, check for data errors, defend against security breaches) if one is using their services at a very low cost? Or often for free? Once data is entrusted to a cloud-based service, which third-parties do they share the information with? Cloud-sourcing involves the use of many services, and many cloud based services provide services to each other, and thus cloud-based products may have to share your information with third parties if they are involved in processing or transferring of your information. They may share your information with advertisers as well, as many do to help cover the costs. Of course each cloud-based service has its own terms and conditions, or service level agreement, that the user agrees to (often without reading), and is often updated. People need to be aware of terms and conditions as well as to keep up with updates. Terms and conditions between user and company cannot alone protect the privacy and security of the user's information. Security can be breached, infrastructure can be damaged, and a company can become bankrupt, often leaving users without recourse. Furthermore, terms and conditions or service level agreements, may be unfair, as well as illegal in some countries, and can of course easily be broken. As well as policy, there needs to be standards and best practices adhered to for storing, encrypting and securing data securely. Some corporations have policies about this, and

some co-regulations about the protection of private information. But in many countries the storing of personal information by companies is not regulated.

The information stored by cloud services is subject to the legal, regulatory and policy environments of the country of domicile of the cloud service, as well as the country in which the server infrastructure is based. This is complicated by the fact that some data in transit may also be regulated.

As more and more information is stored in the cloud these issues become pertinent, and cloud computing will continue to offer challenges to national policy and regulation as well as to internet governance, on how best to resolve privacy and security issues.

3.6.2 Economic development issues

Many promoters of cloud computing seem to predict the increasing irrelevance of local IT, whether this is in-house IT or IT contractors. How can cloud computing be strategically used in Africa in such a way that simultaneously takes advantage of services even as building IT capacity? It was asked during the plenary in Vilnius whether there was “a risk of market dominance by the most powerful I.T. Companies?” And can such a possible dominance “further amplify inequalities of wealth distribution in the world?” If governments and corporations increasingly use cloud services, many of which are controlled by corporations in the global north that house much of their infrastructure there, how to ensure that the move to cloud computing does not reduce local IT infrastructure.

3.7 Cloud Computing Risks

The previous section has provided a good overview of what cloud computing. Some issues and risks have been mentioned there, but this section will go more into detail about the risks that cloud computing brings along as stated in previous research. The risks can be divided into two areas; these areas are privacy related risks and (data) security related risks are discussed below:

- **Privacy and confidentiality risk**

  A research for the world privacy forum came up with a whole list of findings in the area of cloud computing. It covers most of general risks that cloud computing brings regarding privacy and confidentiality.

  The users and clients of cloud computing are dependent on their cloud provider when it comes to their privacy or confidentiality. The provider of the cloud computing services determines what policies are held. Imagine that these providers also have the ability to make changes in their policies. It could completely change the privacy for clients. (For example when the data inserted by the cloud users is protected in the preliminary made up policy being used). Changing policies which will allow insight in this data for third parties could be a serious risk depending on the importance of data that is being used. Another example is that cloud providers could extract information from different organizations in the cloud. They could visualize information that could be by any means revealing. It could also detect information that is commercially valuable for them. What stays important is that most cloud users (clients) are usually not aware of the complete policy and thus do not know very well what risks they are exposed to when entering their data into the cloud.

This brings us to the next point where the problem lies in that cloud users share their information with the cloud provider. On itself this is not the problem, but there could (and are) laws in some specific cases that state that certain information is not to be shared with third parties. In this case the third party would be the cloud provider. There are a lot of examples to think about; privacy laws containing specific rules about sharing a client's personal information, such as phone number and address. When an organization uses cloud computing and they put the clients information in the software that is hosted in the cloud, they are actually sharing the clients information with a third party. These laws and regulations will decline the effect of using cloud computing. Organizations will still need software running on their own servers in order to keep the information which is legally bounded to a certain set of rules.

When there are no laws about sharing certain information with third parties such as cloud computing providers, another problem arises. Sensitive information shared in the cloud might get controlled by weak privacy protection. When this data is stored in your own datacenters you can determine how you want to protect this data and also you are the only organization that is able to access this data. You can choose when and whether or not you want to share this information with certain organization. When all this information is in the cloud, they decide upon over the data privacy. Other organizations could extract the information from the cloud provider more easily then when this data would be stored in your own datacenters. An example is a DNA database. This database could store peoples DNA in order to find a certain cure for a disease. When this database is stored on an own datacenter a hospital or research institute can determine whether they want to share it with a police department for example. The police

department could be looking for a fugitive and a DNA database would be handy. Though, when the research institute does not want to provide the database information because they promised their costumers confidentiality, the police department would not get access. On the other side, if this information would be in a cloud, and the cloud provider is not aware of the importance of the data and the confidentiality as promised by the research institute, they would provide this information more easily to a police department. The laws differ in countries, so important is to know where the data stays in the cloud. In essence the information in the cloud is stored on a machine that is provided by a certain organization. The laws that apply for the information on this machine depend on the location where it is stored. So for example when you have a service contract with your cloud provider, but your data is (partially) stored in another country with other laws, there are different regulations concerning privacy in this case. Authorities could pressure the cloud provider more easily into handing over the information in the cloud. When the data would always be in the same location (country) this problem would not exist.

The different locations of data storage bring another problem to mind. Different locations (countries) provide different regulations. When a cloud provider moves the data of a user along different countries that are in the cloud, the legislation of the data also changes. This means again that it is difficult to guarantee a certain degree of privacy about the data. For example when a client enters the cloud with their data and with the help of a service contract determines the privacy. This service contract is then bounded to the legislation of that particular country. When the cloud is rearranging their data and the clients data gets moved to another country with other jurisdiction you could

get the same problems as described before. Local authorities could pressure the cloud providers in other jurisdictions to provide the information of the cloud.

Even though that as an organization you can have a good service contract with a cloud provider there are laws that still override these contracts. Cloud computing provides services for all kind of organizations and people, so it is inevitable that there will be transfer of information concerning such 'crimes'. The law 'against terrorism' in most countries then obligates the cloud provider to pass this information to the authorities in order to prevent terrorism in this case. The user records could be obligated to be accessible for authorities when they assume there is critical information stored in the cloud. The laws do not change as rapidly as the technologies do. It is commonly known that changes in the law are made very slow. This will result in grey areas with for example cloud computing. Does data needs to be publicly accessible for authorities or not? Such questions provide a certain amount of risk. Governments and authorities could pressure cloud providers to provide cloud information because there is nothing concrete stated in the law about this matter.

Cloud providers should be very familiar with the current laws and regulations in the countries where they provide cloud computing. The privacy and confidentiality risks need to be mentioned in policies and contracts. This would result for user in making more accurate decision about where they will store critical or confidential information.

- **Security risks**: Research has been done in the area of cloud computing concerning more technical risks. There are several risks to be found in this area. These forms of risk have to do with hacking (technical), or attacks from people with malicious intends.

- **Attacks:** The web in general is haunted by attacks on XML signatures. XML is a web based language and as cloud computing could also be web based, they are exposed to this problem. These forms of hack are usually used to obtain data without having the rights to access them. A rightful user does a request for a certain piece of data or information, and the hacker intercepts this request. He then uses the `sign` of the rightful user in order to obtain the data he wants. He sends his request to the cloud, and because the cloud recognizes him as a valid user it responds with the requested data. This is a dangerous risk because the hacker can act as if he is a legitimate user of the cloud.

  An example would be a credit card company that is using the cloud. When an employee of the bank requests for a client's personal data, this request is send to the cloud. The hacker intercepts this request and uses the ´sign´ of this employee. The employee verifies himself by logging into the system with a password and username combination. This provides the sign for that particular employee. Now when the hacker intercepts this request, he can act as being the employee. He is then able to change the request for its personal use. Client's personal data could be changed into client's credit card numbers. The results would speak for it. The cloud is a kind of remote server like Dreamweaver, Zentyal, Rackspace etc.. Cloud can be connected with the client PC´s laptops PDA etc. and use it for input and output. Also important is that it verifies 'us' as users. It gives us thus authorization to access certain areas in the cloud. All these devices themselves do not connect to the cloud; in essence it is the browser on these machines that establishes the actual connections. Most users have Internet Explorer[1], Firefox[2] or Chrome[3] installed and use this to connect to the cloud. This link is another security risk and needs protection. Browser security is something that depends on the provider of the browsers.

[1]http://*www.microsoft.com/Internet-Explorer_9*

[2]http://www.mozilla.org/firefox

[3]http://www.google.nl/chrome

The browsers are used to navigate to the cloud, but also to navigate to other websites. These browsers have to read scripts that are used on the websites. It is important that browsers can detect the difference between malicious scripts that could be made for controlling browser information. For this issue there is help of a firewall. This security measure is depended on the browser used. i.e. Firefox, Chrome or Internet Explorer.

Besides risk in the access tools for the cloud, the cloud itself is also exposed to risks. These risks must be coped with by the provider of cloud services. The first risk described for the cloud is called an Injection attack. These forms of attacks try to implement an own coded malware applications into the existing cloud. The goals of this malware could differ from obtaining only data to completely control applications in the cloud. The software is brought into the cloud, and the cloud is fooled to believe that that software is provided by the cloud user. Whenever cloud users then use the systems, they will be redirected to the malware instead of their real software.

Another important form of risk for a cloud is a so called flooding attack. In general, flooding attacks are to be seen as a huge amount of requests for a service. A “hacker” sends many request to a server which the cloud hosts. These entire requests are in fake fact and have the goal to get the cloud offline. It tries to make so many requests for a particular service that the server cannot cope with the amount of request so it goes down.

Flooding attacks cause both direct and indirect denial of service. Logically when a cloud finds a lot of requests for a particular server, it accounts additional computing power to that service in order to handle all the requests. This is the general idea of cloud computing. However in the real situation, this would only be in advantage of a “hacker”. The hacker now only needs to focus his flooding attack on a single server in order for the cloud to account all the computing power to that service. This is the so called direct denial of service because the hacker focuses on a service and wants to get that particular service down.

On the other side there is indirect denial of service. This then affects other services when an attacker means to hack a particular service down in the direct denial of service. These effects depend on the computing power the hacker has access to. If he tries to cause downtime for a particular service (which is hosted on a server) it could cause downtime for other services too. The servers account all their computing power to all the requests that are being made for one specific service, and thus this causes that there is no rest of computing power to access other applications in the cloud on that particular server. Though it depends on the infrastructure of the cloud, how bad the side effects are. For example the cloud could export the service to another server when it notices that a particular server is not able anymore to cope with all the requests. This will cause even more downtime on other services than before.

3.8 Service Models

Cloud computing providers offer their services according to several fundamental models: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The cloud service models are as shown in Fig. 3.2 which is discussed below:

- **Software as a Service (SaaS):** SaaS is a very popular service in which suppliers deliver web applications to its users. The pioneer of providing SaaS was a company called Salesforce.com. Another common example of a SaaS provider is Google with its email and office tools like word processor, spreadsheet or calendar. It could be any type of application and users can even create their own, hosting them on the provider's servers. They work well in the internet environment, as they are specially designed for it. Microsoft has taken a different approach, by creating a service which they call Software plus Service. These services are commercially known as services with the words Live or Online attached to them. They enable a seamless integration of their desktop applications with cloud services. This allows for the comfort of having all the functionality of desktop software and the option to store files on-site, together with easy and convenient team collaboration functionalities. The newest Office 2010 natively supports some of these functions.
- **Platform as a Service (PaaS)**: In order to avoid confusion of this service with Saas, it is good to imagine it as a cloud OS. The providers of the service enable its users to install their applications on a platform, which can provide any operating system or even emulate various types of hardware. With SaaS, the providers enable access to their own

applications which they host on their servers, but with PaaS the users have greater possibilities for deployment of their own software which does not have to be specialized for the web browser. The processing power is scalable, so they do not have to worry about the number of servers they would have to use in-house, especially if the required power is only seasonal, for example when closing a financial year.

- **Infrastructure as a Service (IaaS):** This service can provide the functionalities of a whole infrastructure including storage, any platform and any number of desktops. Therefore, it is possible to find another name for this service, which is everything as a Service. The current leaders in cloud computing are Google and Amazon (Velte, Velte, & Elsenpeter, 2010). There can be also variations of IaaS, where one is referred to as Hardware as a Service. In contrast to SaaS or PaaS which provide software, Hardware as a Service only provides for virtual hardware on which a user can install anything. In theory, it would be sufficient for a company to have only thin clients in-house and host everything in the cloud. This is not realistic at this moment though, due to the fact that the pricing of cloud services may not be beneficial for all sorts of businesses, but especially due to many security concerns which is discuss throughout this thesis.

  Cloud computing is not a service that can suit every type of organization and requires a thorough consideration of many factors. In order to mitigate some of the security concerns, some companies decide to create so called private clouds. This means that they are the only company occupying it and they probably run the datacenter in-house. This creates a more controlled secure environment, but it comes at a cost of losing some major

benefits of public clouds, for example the lack of need to invest heavily into the infrastructure.

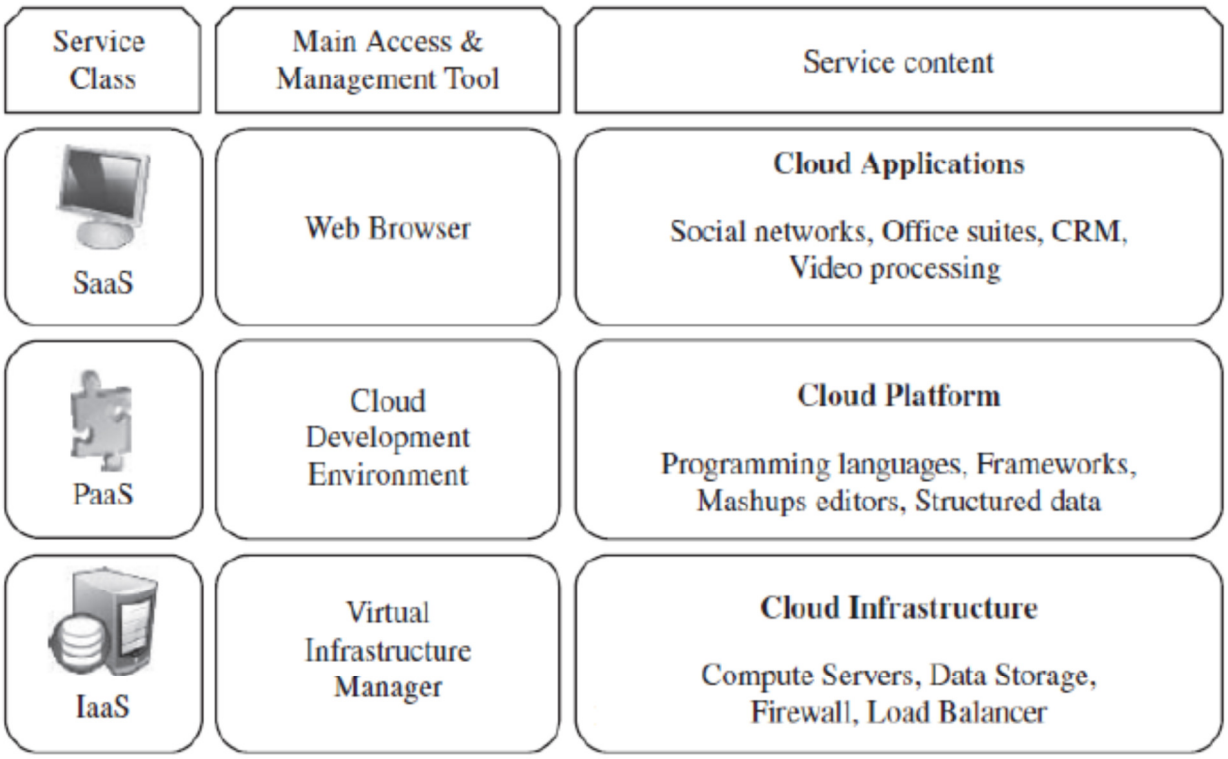

Figure 3.2:- Cloud service models

3.9 Dynamic Provisioning

Dynamic provisioning refers to the ability of dynamically acquiring new resources and integrating them into the existing infrastructure and software system. Resources can be of different nature: hardware or software. In the most common cases resources are virtual machines of software services and they identify two very well-known segments in the cloud computing market: Infrastructure-as-a-Service and Software-as-a-Service. Dynamic provisioning in Aneka is mostly focused on provisioning and controlling the lifetime of virtual nodes where to deploy the Aneka daemon and the container software. Hence, it specifically refers to the provisioning of

hardware in the form of virtual machines whether they are provided by an IaaS provider such as Amazon EC2 or GoGrid or a virtual machine manager such as Xen Server.

A computing system, such as web application or platform for high performance computing such as Aneka, is normally configured with a reasonable installed base of computing nodes that in most of the cases are sufficient to address the performance and application requirements of associated to the usual workload. There could be cases in which the installed base is not enough to cover the demand of users. These cases are normally identified by peak load conditions or very specific and stringent Quality of Service requirements for a subset of applications or users. Here is where dynamic provisioning can constitute an effective solution: by leveraging a dynamic base, commonly composed by a pool of virtual machines, the capacity and the throughput of the system can be temporarily increased in order to address the specific needs of users.

To make this happen, a collection of system components coordinate their activities:

- ***Scheduling and management*** facilities identify the need of additional resources and ultimately decide to provision them.

- ***Load balancing and monitoring*** facilities currently monitor the system load and the status of current installed base (physical and virtual nodes under management).

- ***Budget management*** facilities provide information about whether the system can afford additional resources given the requirements of users, their budget, and the system policies.

• ***Dynamic provisioning*** facilities actually deal with the technical issues of issuing the provisioning request and keep under control the lease of the managed virtual instances.

It is expected that the virtual nodes will be properly configured with the required software stack and virtual hardware to be integrated in the existing system. Whereas the virtual hardware properties can be generally defined while issuing the provisioning request, the software stack installed in the virtual machine can be statically configured in the template used to create virtual instances or dynamically deployed by the system once the node has become available. Obviously, this is just reference architecture but it is general enough to cover a huge variety of cases in which dynamic provisioning are used to enhance the performance and the throughput of the system. For example there could be cases in which the system only relies on a local virtualization facility that is freely available; in this case the provisioning process will be simpler and not conditioned by any consideration about the budget. But, since dynamic provisioning mostly relies on Infrastructure-as-a-Service providers that make available resources on a pay-as-you-go model, the common case includes an evaluation of the economical aspect of provisioning resources and whether it is of advantage to leverage them.

3.10 Federated Cloud Providers

Cloud federation comprises services from different providers aggregated in a single pool supporting three basic interoperability features - resource migration, resource redundancy and combination of complementary resources resp. services. Migration allows the relocation of resources, such as virtual machine images, data items, source code, etc. from one service domain to another domain. While redundancy allows concurrent usage of similar service features in

different domains, combination of complementary resources and services allows combining different types to aggregated services. Service disaggregation is closely linked to Cloud Federation as federation eases and advocates the modularization of services in order to provide a more efficient and flexible overall system. Two basic dimensions of cloud federation are identifying: horizontal, and vertical. While horizontal federation takes place on one level of the Cloud Stack e.g., the application stack, vertical federation spans multiple levels. In the following which is focus on horizontal federation; aspects of vertical federation are out of the scope of this publication. It is not possible for single Cloud infrastructure provider will be able to establish their data centers at all locations around the world. As a result it creates a problem to Cloud application service providers in delivering QoS expectations for all their client/customer. Hence, customer wants to use the services of multiple cloud infrastructure service providers. Multiple cloud infrastructure service providers can provide can help in meeting their specific consumer needs. This type of needs frequently arises in organizations with global operations and applications such as media hosting, Internet service, and web applications. This make compulsory to build mechanisms for federation of cloud infrastructure service providers for virtualized provisioning of resource from multiple cloud service providers. There are many issues occur while creating such federation of cloud service providers.

To fulfill these requirements, next generation cloud service providers should be able to:-

(i) Dynamically scalable or resize their provisioning capability based on random change in input workload demands by leasing available computational and storage capabilities from other cloud service providers.

(ii) Operate as parts of a market driven resource leasing federation, where application service providers such as Salesforce.com host their services based on negotiated Service Level Agreement (SLA) contracts driven by competitive market prices.

(iii) Deliver on demand, reliable, cost-effective, and QoS aware services based on virtualization technologies while ensuring high QoS standards and minimizing service costs. They need to be able to utilize market-based utility models as the basis for provisioning of virtualized software services and federated hardware infrastructure among users with heterogeneous applications and QoS targets.

Components of cloud federation are explaining in next three sections and the Cloud federation model is as shown in Fig. 3.3.

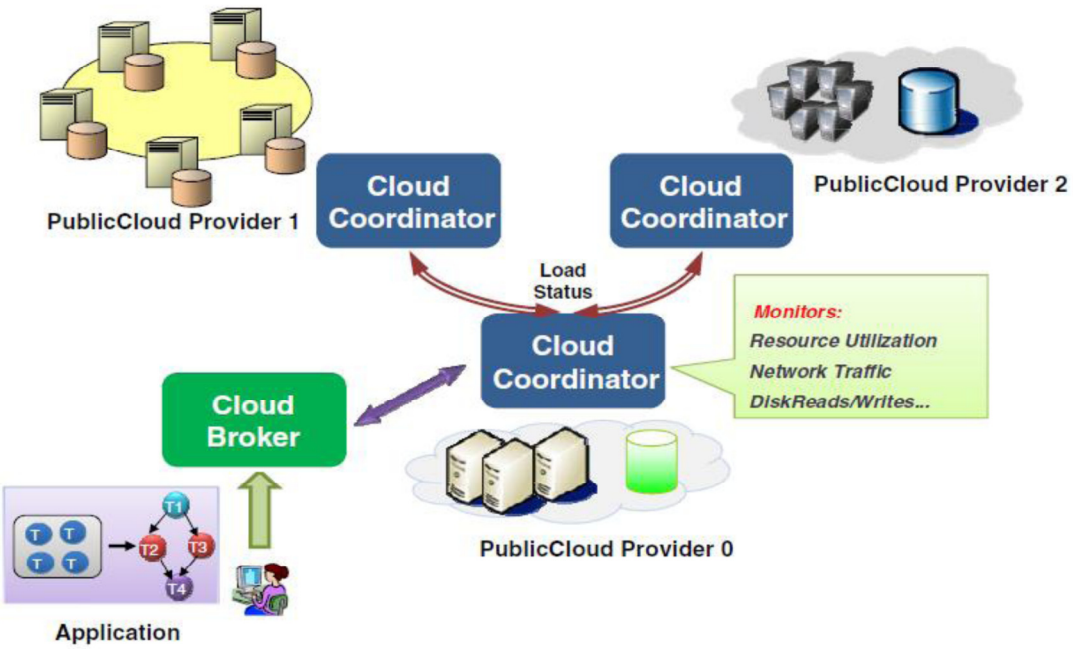

Figure 3.3:- Cloud federation

3.10.1 Cloud Coordinator (CC)

The cloud coordinator service is responsible for the management of domain specific enterprise clouds and their membership to the overall federation driven by market-based trading and negotiation protocols. It provides a programming, man-agreement, and deployment environment for applications in a federation of Clouds. The Cloud Coordinator exports the services of a cloud to the federation by implementing basic functionalities for resource management such as scheduling, al-location, (workload and performance) models, market enabling, virtualization, dynamic sensing/monitoring, discovery, and application composition as discussed below:-

- **Scheduling and Allocation:** This component allocates virtual machines to the cloud nodes based on user's QoS targets and the clouds energy management goals.
- **Market and Policy Engine:** The SLA module stores the service terms and conditions that are being supported by the cloud to each respective cloud broker on a per user basis. Based on these terms and conditions, the Pricing module can determine how service requests are charged based on the available supply and required demand of computing resources within the cloud. Cloud customers can normally associate two or more conflicting QoS targets with their application services. In such cases, it is necessary to trade off one or more QoS targets to find a superior solution.
- **Application Composition engine:** This component of the cloud coordinator encompasses a set of features intended to help application developers create and deploy applications, including the ability for on demand interaction with a database backend such as SQL Data services provided by Microsoft Azure, an application server such as Internet Information Server (IIS) enabled with secure ASP.Net scripting engine to host

web applications, and a SOAP driven Web services API for programmatic access along with combination and integration with other applications and data.

- **Virtualization:** VMs support flexible and utility driven configurations that control the share of processing power they can consume based on the time criticality of the underlying application. However, the current approaches to VM-based cloud computing are limited to rather inflexible configurations within a cloud. This limitation can be solved by developing mechanisms for transparent migration of VMs across service boundaries with the aim of minimizing cost of service deli-very (e.g., by migrating to a cloud located in a region where the energy cost is low) and while still meeting the SLAs. This limitation has to be addressed in order to support utility driven, power-aware migration of VMs across service domains.
- **Sensor**: Sensor infrastructure will monitor the power consumption, heat dissipation, and utilization of computing nodes in a virtualized cloud environment. Sensor Web provides a middleware infrastructure and programming model for creating, accessing, and utilizing tiny sensor devices that are deployed within a cloud. The cloud coordinator service makes use of Sensor Web services for dynamic sensing of cloud nodes and surrounding temperature. The output data re-ported by sensors are feedback to the coordinator's Virtualization and Scheduling components, to optimize the placement, migration, and allocation of VMs in the cloud. Such sensor-based real time monitoring of the cloud operating environment aids in avoiding server breakdown and achieving optimal throughput out of the available computing and storage nodes.
- **Discovery and Monitoring:** In order to dynamically perform scheduling, re-source allocation, and VM migration to meet SLAs in a federated network, it is mandatory that

up-to-date information related to Cloud's availability, pricing and SLA rules are made available to the outside domains via the Cloud Exchange. This component of Cloud Coordinator is solely responsible for interacting with the Cloud Exchange through remote messaging.

3.10.2 Cloud Broker (CB)

The Cloud Broker entity helps users to find out appropriate Cloud service providers through the Cloud Exchange and consult with Cloud Coordinators entity for a scheduling of resources that meets QoS needs of users. Cloud Broker has following components is given below:-

- **User Interface:** This provides the access linkage between a user application interface and the broker. The Application Interpreter translates the execution requirements of a user application which include what is to be executed, the description of task inputs including remote data files (if required), the information about task outputs (if present), and the desired QoS.
- **Core Services:** They enable the main functionality of the broker. The Service Negotiator bargains for Cloud services from the Cloud Exchange. The Scheduler determines the most appropriate Cloud services for the user application based on its application and service requirements. The Service Monitor maintains the status of Cloud services by periodically checking the availability of known Cloud services and discovering new services that are available.
- **Execution Interface:** This provides execution support for the user application. The Job Dispatcher creates the necessary broker agent and requests data files (if any) to be dispatched with the user application to the remote Clouded resources for execution. The

Job Monitor observes the execution status of the job so that the results of the job are returned to the user upon job completion.

- **Persistence:** This maintains the state of the User Interface, Core Services, and Execution Interface in a database. This facilitates recovery when the broker fails and assists in user-level accounting.

3.10.3 Cloud Exchange (CE)

As a market maker, the Cloud Exchange acts as an information registry those stores the Cloud's current usage costs and demand patterns. Cloud Coordinators periodically update their availability, pricing, and SLA policies with the Cloud Exchange. Cloud Brokers query the registry to learn information about existing SLA offers and resource availability of member Clouds in the federation. Furthermore, it provides match-making services that map user requests to suitable service providers. As a market maker, the Cloud Exchange provides directory, dynamic bidding based service clearance, and payment management services as discussed below

- **Directory:** The market directory allows the global Cloud Exchange participants to locate providers or consumers with the appropriate bids/offers. Cloud providers can publish the available supply of resources and their offered prices. Cloud consumers can then search for suitable providers and submit their bids for required resources. Standard interfaces need to be provided so that both providers and consumers can access resource information from one another readily and seamlessly.
- **Auctioneer:** Auctioneers periodically clear bids and asks received from the global CEx participants. Auctioneers are third party controllers that do not represent any providers or

consumers. Since the auctioneers are in total control of the entire trading process, they need to be trusted by participants.

- **Bank:** The banking system enforces the financial transactions pertaining to agreements between the global Cloud Exchange participants. The banks are also independent and not controlled by any providers and consumers; thus facilitating impartiality and trust among all Cloud market participants that the financial transactions are conducted correctly without any bias. This should be realized by integrating with online payment management services, such as PayPal, with Clouds providing accounting services.

3.11 Service-Level Agreement (SLA)

A service-level agreement (SLA) is a part of a service contract where a service is formally defined. In practice, the term *SLA* is sometimes used to refer to the contracted delivery time (of the service or performance). In the term of cloud and customer a service-level agreement (SLA) is an association between a Cloud service provider and a customer that specifies, usually in measurable terms, what services the Cloud service provider will furnish. Many Cloud service providers (CSP) s provide their customers with an SLA. More recently, IT departments in major enterprises have adopted the idea of writing a service level agreement so that services for their customers (users in other departments within the enterprise) can be measured, justified, and perhaps compared with those of outsourcing Cloud providers. The SLA will typically have a technical definition in terms of mean time between failures (MTBF),mean time to repair or mean time to recovery (MTTR); various data rates; throughput; jitter; or similar measurable details.

Service level agreements are also defined at different levels:-

- **Customer-based SLA:** A customer-based SLA is an agreement with a specific customer or customer group covering all the IT services they use.
- **Service-based SLA:** An agreement for all customers using the services being delivered by the service provider.
- **Multilevel SLA:** The SLA is split into the different levels, each addressing different set of customers for the same services, in the same SLA.

- **Corporate-level SLA:** Covering all the generic service level management issues appropriate to every customer throughout the organization
- **Customer-level SLA:** covering all service level management issues relevant to the particular customer group, regardless of the services being used.
- **Service-level SLA:** covering all service level management issue relevant to the specific services, in relation to this specific customer group.

3.12 Literature Review

Several thesis of cloud computing have been studies, these gives us the idea about fundamentals of cloud computing like, architecture of cloud, deployment models, service models and essential characteristics of Cloud, job scheduling, resource allocation, security, virtualization. One important characteristics of cloud that is unlimited resources user can request any number of resources at any time depending on the needs, single cloud service provider is not able to guarantee this characteristics. To solve this problem, several platforms has been studies for cloud

federation; with different motivations for parties to join it, many fundamental problems related to federation still remain unanswered. One of these problems is deciding when providers should outsource their local requests to other member of the federation and at what price they should provide resources to the federation. The outsourcing problem is not considered only in the context of federated clouds; it was also provide as a way of increasing scalability of applications in hybrid Clouds. The idea of federation systems was already present in the Grid. For instance, other authors use federation to get more resources in a distributed Grid environment. The application of federation in the Cloud was initially proposed within the Reservoir project. In particular a Rochwerger describes the difficulty to merge different providers with different APIs and features. They do not present any model to decide when to move tasks to a federated provider based on economic criteria.

Goiri presents a profit-driven policy to take decisions related to outsourcing or selling idling resources. But, they did not take into account all kinds of VMs, like on-demand, reserved and spot VMs and it is special type of low priority one. It has been terminated when high priority request comes to provider. Current public Cloud service providers like Amazon EC2, usually use fixed pricing policies for the infrastructure services they provide. While this pricing models are not appropriate for federated environments as a policy to be applied between its participants, because it does not reflects current market price of resources due to dynamism in supply and demand. Dynamic pricing of resources, is not taken into consideration of current work, and has been a subject of other studies. Hence, in this work, a policy based on the provider utilization, is applied by federated providers to dynamically value resources. The subject of leveraging spot VMs has recently attracted considerable attention; Yi gives approach to minimize the costs of computations using Amazon EC2s spot instances for resource provisioning. This thesis considers

techniques for increasing customers' benefit by using concept of VMs. A study considers the application of market-oriented mechanisms in federated environment. These methods provide fairness and benefits for providers involved in the federation. The Cloud allows a service provider to virtualize its resources and dynamically provision them as unified computing resources based on a Service Level Agreement (SLA) established through negotiation between the service provider and the consumer. In order to be profitable, service providers tend to share their resources among multiple concurrent services owned by different customers, but at the same time, they must guarantee that each service has always enough resources to meet the agreed performance goals. This requires of complex resource management mechanisms that could dynamically manage the provider's resources in the most cost-effective way (e.g. maximizing their utilization or reducing their power consumption), while satisfying the QoS agreed with the customers. Former resource management approaches for Cloud providers hindered their market potential by considering a limited amount of resources. In these approaches, if a service provider had not enough local resources to fulfill its customers' requirements, it should start denying the acceptance of new customers or canceling low priority services that were already running on the system. This has further implications than just losing the revenue from some services, because it also implies a loss of reputation and therefore a loss of future customers. This problem can be overcome via Cloud federation. Different providers running services that have complementary resource requirements over time can mutually collaborate to share their respective resources in order to fulfill each one's demand. For instance, a provider could outsource resources to other providers when its workload cannot be attended with its local resources. In this way, the provider would obtain higher profit because it can attend more customers. Of course, the expected revenue from these customers should be higher than the cost of outsourcing the additional

resources in order to be worth doing it. Similarly, a provider that has underused resources could rent part of them to other providers. For this situation prefer insourcing of the resources. In this way, the provider would improve its benefit, better exploit its resources, and compensate the cost of maintaining them. Again, the expected benefit from renting its resources should be higher than the cost of maintaining them operative.

3.13 Research Gaps

In Cloud computing environment, there are two things which is focuses on, one is client/user and other is Cloud service provider. Currently, there are numbers of algorithms available, out of which some of them focuses on client and others on Cloud service provider.

The main goals of the algorithm, which focuses on the clients, are –

- Reduction of the cost of user's job execution at the Cloud provider side.
- Allocate resources in such a manner that job should be completed within a given time span. And few methods focus on both cost and time.
- Allocation of VM such that QoS (quality of services) for a user should be meets.

And the goals of the algorithms which focus on Cloud Providers are-

- Improve the profit of Cloud provider,
- Minimize the number of physical nodes (servers), and power.

All the above mentioned algorithms work within the single Cloud provider. But in current scenario, where the number of users and its requirements are increasing much rapidly, it is not possible to handle all the clients' requirements by a single Cloud provider. So in that case,

introduce the concept of Cloud federation which more than one Cloud providers share its resources to fulfill the user's requirements.

Currently, no algorithm is available to improve the profit of the cloud providers. So, an algorithm are proposed which is focus on increasing the availability of resources to customer and increase the profit of cloud service providers by using the concept of a separate component which will contain the information about the available resources in all cloud providers. And in my algorithm, when a user request comes to a cloud provider, then this cloud provider first searches its own available resources. If it doesn't find resource, then it will contact to other cloud provider by using the shared component between them and allocate the resource form other cloud provide to virtual machine.

CHAPTER 4

Practical part

- *VM Provisioning*
- *Framework for VM Provisioning in federated cloud Environment*
- *Benefits*
- *Approaches of the federated cloud*
- *Flow-chart*
- *VM Provisioning based on Profit in Federated cloud Environment*

4.1 VM Provisioning

Provisioning can be defined as a "high-level management of computing, network, and storage resources that allow them to effectively provide and deliver services to customers". The diversity and flexibility of the functionalities suddenly scale-up and scale-down computing systems modeled by federated Cloud computing model, combined with the magnitudes and uncertainties of its components workload, compute servers, services, cause difficult problems in effective provisioning and delivery of application services. In particular, finding efficient solutions for following challenges is critical to exploiting the potential of federated Cloud infrastructures. The process of provisioning in Clouds is a complex undertaking, as it requires the application provisioned to compute the best software and hardware configuration to ensure that QoS targets of application services are achieved, while maximizing the overall system efficiency and utilization. Achieving QoS targets is important for meeting SLAs agreed with end-users and justifying the investment in Cloud based deployments. However, this process is further complicated by the uncertain behavior of virtualized IT resources and network elements.

4.2 Framework for VM Provisioning in federated Cloud Environment

Cloud and Grid computing build the pillars of todays modern scientific compute environments. Batch computing has traditionally supported high performance computing centers to better utilize their compute resources with the goal to satisfy the many concurrent users with sophisticated batch policies utilizing a number of well managed compute resources. Grid Computing and its predecessor Meta-computing elevated this goal by not only introducing the utilization of multiple queues accessible to the users, but by establishing virtual organizations that share resources among the organizational users. This includes storage and compute

resources and exposes the functionality that users need as services. Recently, it has been identified that these models are too restrictive, as many researchers and groups tend to develop and deploy their own software stacks on computational resources to build the specific environment required for their experiments. Cloud computing provides here a good solution as it introduces a level of abstraction that lets the advanced scientific community assemble their own images with their own software stacks and deploy them on large numbers of computational resources in clouds. Since a number of Infrastructures as a Service (IaaS) exist, our experience tells us the importance of offering a variety of them to satisfy the various user community demands. It is also important to provide users with the capabilities of staging their own software stack. Recently a number of test-beds have been created that allow the provisioning of software stacks by users.

Federation can be defined as a set of organizations that cooperate among each other respecting pre-established rules of trust in order to authenticate users and to promote sharing of resources. In this sense, a federation in cloud allows a provider to outsource resources to other providers according to its clients' demands. It is also possible to rent part of its resources to other providers in order to promote a more suitable utilization of the shared resources. This sharing of resources requires a monitor component that queries a domain about the availability of resources. Federated infrastructures must have components to intercept requests, and intermediary components to negotiate the better allocation to attend the cost and performance expectations of the cloud service providers. These components must be configurable in order to reduce the cost of resources rented from trust partners. In order to evaluate these issues and also perform experiments in the real cloudSim simulator.

4.3 Benefits

The short answer to this is that federated cloud can benefit both providers and end users. Federated cloud is a cloud service model that connects all of these local infrastructure providers to each other, creating a global marketplace that enables each provider to buy and sell capacity on demand. As a provider, you now have instant access to global infrastructure that was previously unattainable. If you have a customer that suddenly needs resources that you do not have available in your infrastructure, just buy the capacity needed from the marketplace. If another customer lives in New York but needs to accelerate their website or application in Hong Kong, you can simply purchase available resources from a federated cloud provider and use the infrastructure that is already there.

Even smaller service providers can offer a global network with federated cloud without spending anything to build new infrastructure. Federated cloud can provide data centers that have spare capacity a simple and immediate way to monetize that capacity by submitting it to the marketplace for other providers to purchase and utilize.

Federated cloud ultimately benefits end users the most. Currently, end users have only a handful of “global” cloud providers to choose from and they must adhere to the pricing and support those companies imposed. However, federated cloud can now offer those same end users the ability to pick from a whole new range of providers. They can choose to work with a local provider who is part of a federated cloud network and with whom they already have an established relationship and in which the pricing, support and expertise fit their budget. Although they are going with a local provider, they still have instant access to as much global resources as they need in addition

to the local resources they utilize. One other benefit to end users it that there is no longer a need to manage multiple providers and. They can now have a true single provider.

4.3 Approaches of the federated cloud

In this thesis, cover three companies that have staked a position in the federated-cloud market and highlight the potential benefits or drawbacks of their partnership opportunities.

Although these three aren't the only players in federated cloud, each one offers a different business model that potentially blazes a trail for cloud federation. Despite their differences, all three share two characteristics: Each has created a proprietary software platform upon which service providers can build cloud services, and each has assumed the role of the broker, bringing together suppliers and consumers of cloud resources while providing requisite usagetracking and billing.

4.3.1 Spot Cloud: Federation through a cloud services broker's marketplace

Learn how Spot Cloud uses a federated cloud to bring together buyers and sellers of commodity Infrastructure as a Service (IaaS) resources, as well as how this ecosystem can benefit participating providers.

4.3.2 OnApp: With federated cloud, CDN services possible for providers

Find out how OnApp is enabling cloud providers to federate their resources and use each other’s excess capacity to offer global content delivery network (CDN) services that rival Amazon’s Cloud Front CDN offering.

4.3.3 Tier 3: Cloud federation enables providers to expand service footprint

Understand how Tier 3 licenses and centrally manages its platform to federate cloud providers 'resources, enabling them to expand their geographic reach and scalability for cloud computing services.

4.4 Flow-chart

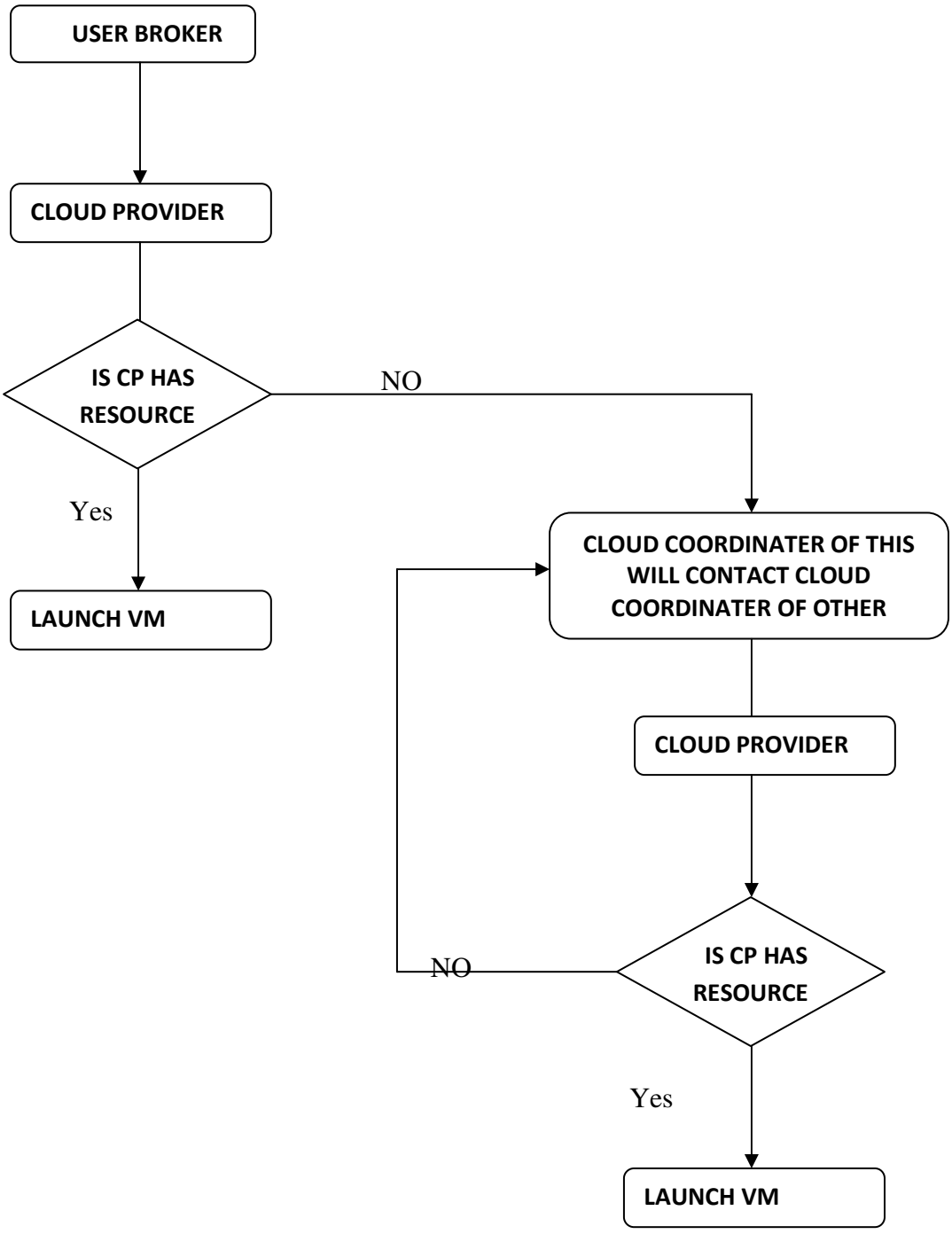

Figure 4.1:- Flowchart of Proposed Algorithm

4.5 VM Provisioning based on Profit in Federated Cloud Environment

In this Thesis, deal with new algorithms which focus on increasing the availability of resources to customers and increasing the profit of cloud service providers. In this algorithm when a user request come on a specified cloud provider it first search its own data center if it's resource to fulfill the user request then allocate the resource otherwise it's contact to other cloud provider by using the shared component between them.

Let us consider the following cases:-

4.5.1 Allocation within the provider

In this policy cloud provider has sufficient resources to satisfy the entire user request or on the other hand cloud provider doesn't sell or purchase its resource to other cloud provider. In this case revenue generated only by selling its resource to the customer. If it's all resource are busy it is reject next user request until it's resource until free. Cost can be define as a fixed cost to manage all the datacenters .This policy is considered as a base policy in order to allow verification of maximum profit a provider can make without the federation.

$$\text{Profit (t)} = \text{Revenue (t)} - \text{Cost (t)} \quad (1)$$

$$\text{R (t)} = (\text{VM\_price}) \,/\, \text{hour} * \text{total no. of VM} * \text{T} \quad (2)$$

Where R represents Revenue and VM is virtual machine

And Cost (t) = no. of VM required to complete request * manage cost per hour * T (3)

Where Revenue (t) and Cost (t) are the revenue and cost at particular time t, respectively.

4.5.2 Outsourcing to federated Clouds

Outsourcing of resources to federated providers can be preferable to over provisioning the data center when demand varies over time. In addition, it also allows the provider to insource its resources to other providers if these are not being used. The decision of using these capabilities is based on their economic viability. Introduces an equation that evaluates whether outsourcing resources to an external provider is profitable or not. When the provider has underutilized resources, it has two options either to shut down or sell (insource) the extra resources (servers). It is not desirable for cloud provider to shut down its resources because providers already have spent a lot of money to establish the datacenters. It is a market rule, that when any organization comes into market, it never thinks about going back, even if it sells its services at lower prices. Here total revenue becomes sum of the revenue earned from the requests that come to provider and revenue earned by renting its resources to other providers. In this case cost will be increase.

$$\text{Cost\_total (t)} = \text{Cost (t)} + \text{Cost\_out (t)} \quad (4)$$

$$\text{Where Cost\_out(t)} = \sum_{i=0}^{VMout} \text{Priceout} * \text{X} \quad (5)$$

Profit (t), Revenue (t) and Cost (t) can be calculated equation (1) (2) (3) respectively.

4.5.3 Insourcing from federate Clouds

As presented before, in a federated cloud the provider can offer its free resources to other providers. In this case, the total cost for the provider does not change. This means that there are not additional costs if the provider rents its free resources. Consider the case when requests from user come to particular cloud provider, provider does not have sufficient resources to fulfill the requests i.e. cloud provider (datacenter) is fully loaded. Cloud provider has two options, either to reject the requests or insource the resource from other cloud provider on the basis of agreements. If cloud provider rejects the requests, it not only loses the money but market value of that service provider also decreases. So in the case of insourcing total revenue becomes the revenue of provider itself and revenue earned by first outsourcing and, then selling the resources. But, the total revenue is expected to increase. For calculating Revenuei (t):

$$\text{Revenuei(t)} = \text{VM\_free} * \text{VM\_price / hour} * \text{T} * \text{Y} \quad (6)$$

$$\text{Where VM\_free} = (\text{no of virtual machine} - \text{utilization})$$

$$\text{Revenuet(t)} = \text{Revenue(t)} + \text{Revenuei(t)} \quad (7)$$

Profit (t), Revenue (t) and Cost (t) can be calculated equation (1) (2) (3) respectively.

Here X in equation (5) and Y in equation (6) show that a cloud provider rents its resources to other providers at fewer prices than a normal user and value of X and Y depends on the contract matrix among the service providers.

4.5.4 Both insourcing and outsourcing in federation

In this case cloud providers are allowed to insourcing and outsourcing of resource to other cloud provider. This case revenue generated becomes the sum of revenue generated by provider's own resources, revenue generated from outsourcing and revenue generated from insourcing as discussed above. And cost is same as in case of insourcing.

Cost_total (t), Revenuet (t), can be calculated equation (4) and (7) respectively.

CHAPTER 5

Results and Discussion

- *CloudSim Architecture*
- *Results for VM Provisioning based on Profit in Federated cloud Environment*

The experiments presented in this thesis were developed using CloudSim discrete-event Cloud simulator. The number of providers is one of the issues during the simulation and estimated the effect of the policies considering dissimilar number of federation members. The pricing policy of one of the commonly known cloud service provider that is Amazon EC2 is used. While the numbers in this thesis was not be interpreted correctly. Actually, reference values for revenues, costs, and virtualization parameters are used for to demonstrate how our proposed equations can drive resource allocation decisions, though the particular values of these parameters will highly depend on the real provider's characteristics. To achieve more accuracy, each experiment is done 15 times by using different input values of workload and the average of the results is reported. Study the CloudSim concept to test all the equation, till date no software platform available to test any proposed approach that works in federated Cloud environment. The primary objective of this CloudSim is to provide a generalized and scalable simulation framework that enables seamless modeling, simulation, and experimentation of emerging Cloud computing setups and application services. By using CloudSim, researchers and industry-based developers can focus on specific system design issues that they want to investigate, without getting concerned about the low level details related to Cloud-based infrastructures and services.

Overview of CloudSim functionalities:

- Provide Support for modeling and simulation of large scale Cloud computing data centers.
- Provide support for modeling and simulation of virtualized server hosts, with customizable policies for provisioning host resources to virtual machines.
- Provide support for modeling and simulation of energy-aware computational resources.
- Provide support for modeling and simulation of federated clouds.

- Provide support for user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines.
- Provide support for dynamic insertion of simulation elements, stop and resume of simulation.

As it has been mentioned above that CloudSim provide support for simulation and modeling of cloud federation. So CloudSim as tool to find out the result of our proposed work have been choosing.

5.1 CloudSim Architecture

Figure 5.1 shows the layered design of the CloudSim software framework and its components. Initially CloudSim used SimJava as discrete event simulation engine that supports several core functionalities, such as queuing and processing of events, creation of Cloud system entities (services, host, data center, broker, and virtual machines), communication between components, and management of the simulation clock. However in the recent version, SimJava layer has been removed, so that Cloudsim allow some additional operations that are not supported by it.

The CloudSim simulation layer provides support for modeling and simulation of virtualized Cloud-based data center environments including dedicated management interfaces for virtual machines (VMs), memory, storage, and bandwidth. The fundamental issues such as provisioning of hosts to VMs, managing application execution, and monitoring dynamic system state are handled by this layer. A Cloud provider, who wants to study the efficiency of different policies in allocating its hosts to VMs (VM provisioning), would need to implement their strategies at

this layer. Such implementation can be done by programmatically extending the core VM provisioning functionality.

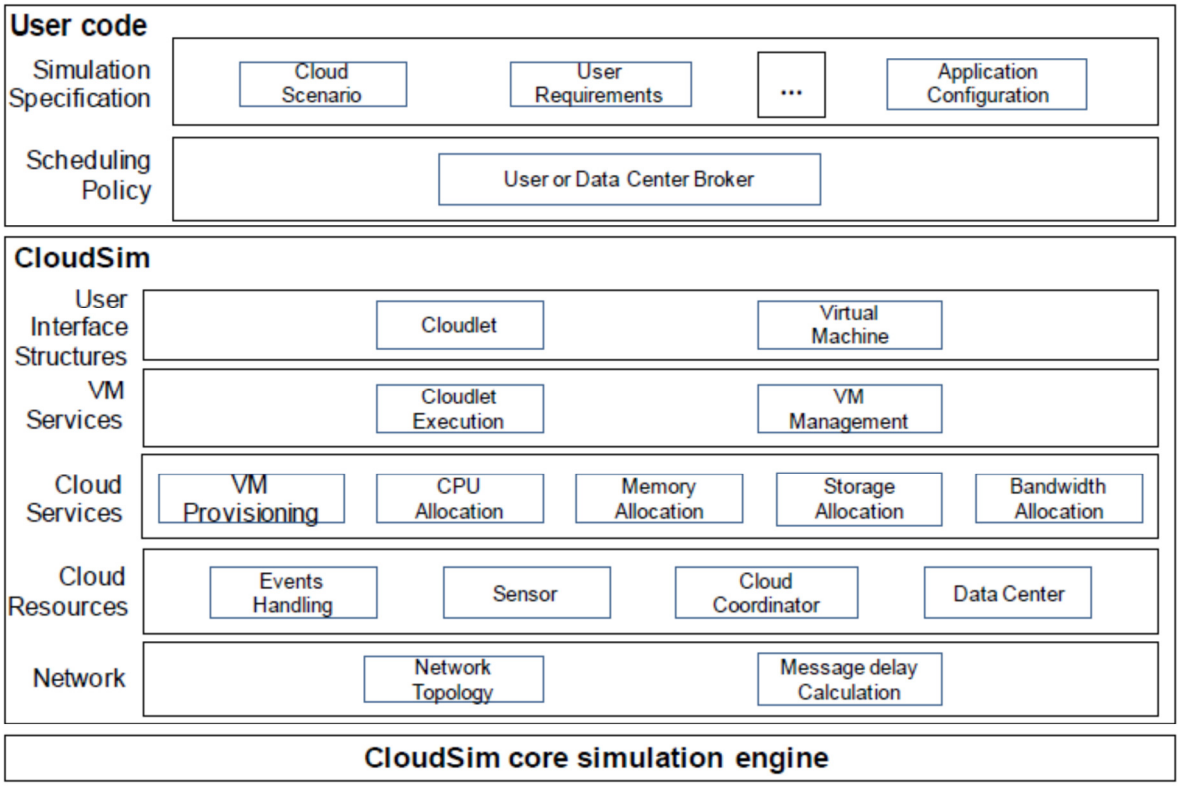

Figure 5.1:- Layered Architecture of CloudSim

There is a clear distinction at this layer related to provisioning of hosts to VMs. A Cloud host can be concurrently allocated to a set of VMs that execute applications based on SaaS provider's defined QoS levels. This layer also exposes functionalities that a Cloud application developer can extend to perform complex workload profiling and application performance study. The top-most layer in the CloudSim stack is the User Code that exposes basic entities for hosts (number of machines, their specification and so on), applications (number of tasks and their

requirements), VMs, number of users and their application types, and broker scheduling policies. A Cloud application developer can perform following activities: (i) generate a mix of workload request distributions, application configurations; (ii) model Cloud availability scenarios and perform robust tests based on the custom configurations and (iii) implement custom application provisioning techniques for clouds and their federation.

5.1.1 Modeling the Cloud

The VM provisioning policy wear the responsibility of whole life cycle operation management such as, provisioning of a host to a VM, VM creation, VM destruction, and VM migration. In same way, one or more application services can be provisioned within a single VM instance, called to as application service provisioning in the context of Cloud computing. To simulate infrastructure level cloud computing services to inherit the datacenter entity of CloudSim is needed. Host entities have been contained within the datacenter entity. The hosts are assigned to one or more VMs based on a VM allocation policy that should be defined by the Cloud service provider. In the context of CloudSim, an entity is an instance of a component. A Clouds component can be a class abstract or complete, or set of classes that represent one CloudSim model data center, host. In addition, service models or provisioning techniques that developers want to implement and perform tests with CloudSim without enforcing any limitation. If an application service is defined and modeled, it can be assigned to one or more pre-instantiated VMs through a service specific allocation policy. Allocation of application-specific VMs to Hosts in a Cloud-based data center is the responsibility of a Virtual Machine Allocation controller component called VM Allocation Policy. A Datacenter can manage several hosts that in turn manage VMs during their life cycles. Host is a CloudSim component that represents a

physical computing server in a Cloud: it is assigned a pre-configured processing capability expressed in millions of instructions per second (MIPS), memory, storage, and a provisioning policy for allocating processing cores to virtual machines. The Host component implements interfaces that support modeling and simulation of both single-core and multi-core nodes. VM allocation (provisioning) is the technique of instantiation of VM instances on hosts that most suitable in term of capabilities like storage, memory, configurations software environment, and requirements availability zone of the application services provider. CloudSim supports the formation of custom application service models that can be deployed within a VM instance and its users are required to extend the core Cloudlet object for implementing their application services.

5.1.2 Modeling the Cloud Market

Market is a crucial component of the Cloud computing. It is necessary for regulating Cloud resource trading and on-line negotiations in public Cloud computing model, where services are offered in a pay-as-you-go model. Hence, research studies that can accurately evaluate the cost-to-benefit ratio of emerging Cloud computing platforms are required. Furthermore, SaaS providers need transparent mechanisms to discover various Cloud providers' offerings IaaS, PaaS, SaaS, and their associated costs. Thus, modeling of costs and economic policies are important aspects to be considered. When designing a Cloud simulator. The Cloud market is modeled based on a multi-layered design. The first layer contains economic of features related to IaaS model such as cost per unit of memory, cost per unit of storage, and cost per unit of used bandwidth. Cloud customers (SaaS providers) have to pay for the costs of memory and storage when they create and instantiate VMs whereas the costs for network usage are only incurred in

event of data transfer. The second layer models the cost metrics related to SaaS model. Costs at this layer are directly applicable to the task units (application service requests) that are served by the application services.

5.1.3 Modeling the Network Behavior

Modeling complicated network topologies to connect simulated Cloud computing entities (hosts, storage, end-users) is an important consideration because latency messages directly affects the overall service satisfaction experience. An end user or a SaaS provider consumer who is not satisfied with the delivered QoS is likely to switch their Cloud provider hence it is very important requirement that Cloud system simulation frameworks provide facilities for modeling realistic networking topologies and models. Inter-networking of Cloud entities data centers, hosts, SaaS providers, and end-users in CloudSim is based on a conceptual networking abstraction. In this model, there are no actual entities available for simulating network entities, such as routers or switches. Instead, network latency that a message can experience on its path from one CloudSim entity host to another Cloud Broker is simulated based on the information stored in the latency matrix.

At any instance of time, the CloudSim environment maintains *m x n* size matrix for all CloudSim entities currently active in the simulation context. An entry *eij* in the matrix represents the delay that a message will undergo when it is being transferred from entity *i* to entity *j* over the network. Recall, that CloudSim is an event-based simulation, where different system models (entities) communicate via sending events. The event management engine of CloudSim utilizes the inter-entity network latency information for inducing delays in transmitting message to entities.

5.1.4 Modeling a Federation of Clouds

In order to federate or inter-network multiple clouds, there is a requirement for modeling a Cloud Coordinator entity. This entity is responsible not only for communicating with other data centers and end-users in the simulation environment, but also for monitoring and managing the internal state of a data center entity. The information received as part of the monitoring process, that is active throughout the whole duration of simulation, is utilized for making decisions related to inter-cloud provisioning. It can be observed that none of the software offering similar functionality to the Cloud Coordinator is offered by existing providers, such as Amazon, Azure, or Google App Engine presently. So, if a developer of a real-world Cloud system wants to federate services from multiple clouds, one of the challenging tasks is to develop a Cloud Coordinator component. By having such an entity to manage the federation of Cloud-based data centers, aspects related to communication and negotiation with foreign entities are isolated from the data center. Therefore, by providing such an entity among its objects, CloudSim helps Cloud developers in speeding up their application service performance testing.

The two fundamental aspects that must be handled when simulating a federation of clouds include communication and monitoring. The first aspect communication is handled by the data center through the standard event-based messaging process. The second aspect data center monitoring is carried out by the Cloud Coordinator. Every data center in CloudSim needs to instantiate this entry in order to make itself a part of Cloud federation. The Cloud Coordinator start the inter-cloud load transfer process based on the current load of the data center. The specific sets of events that affect the transfer load are implemented via a specific sensor entity. This entity provides the information about the utilization of resource within the datacenter. Each sensor entity implements a fixed parameter such as under provisioning, over provisioning, and

SLA violation, related to the data center. For enabling live monitoring of a data center host, a sensor that keeps track of the host status utilization, heating is attached with the Cloud Coordinator. At every monitoring step, the Cloud Coordinator requests the sensor. If a certain pre-configured threshold is achieved, the Cloud Coordinator starts the communication with its peers other Cloud Coordinators in the federation, for possible load-shredding. The negotiation protocol, load-shredding policy, and compensation mechanism can be easily extended to suit a particular research study.

5.2 Results for VM Provisioning based on Profit in Federated Cloud Environment

With the different value of contract matrix [Cij]={{a_{11},a_{12},a_{13},a_{14}} ,{a_{21},a_{22},a_{23},a_{24}} ,{a_{31},a_{32}, a_{33}, a_{34}},{a_{41},a_{42},a_{43},a_{44}} } experiments are performed and results are evaluated. Contract matrix (Cij) values depend on the agreement of shared resource between them. Matrix values (a_{ij}) helps to calculate revenue and cost of different cloud providers. The value in Cij means that revenue multiplied by Cij get to provider 'i', if the i^{th} provider uses the resources of j^{th} provider. For the first experiment the values of contract matrix given below:-

Contract [Cij]={ {1,0.3,0.8,0.6}, {0.7, 1,0.2,0.9}, {0.2,0.8, 1,0.5}, {0.4,0.1,0.5,1} }

	Revenue	Profit
Allocation within the provider	**0**	**0**
Insourcing	**41**	**37**
Outsourcing	**18**	**11**
Both insourcing and outsourcing	**60**	**56**

Table 5.1: Simulation Results for Profit and Revenue

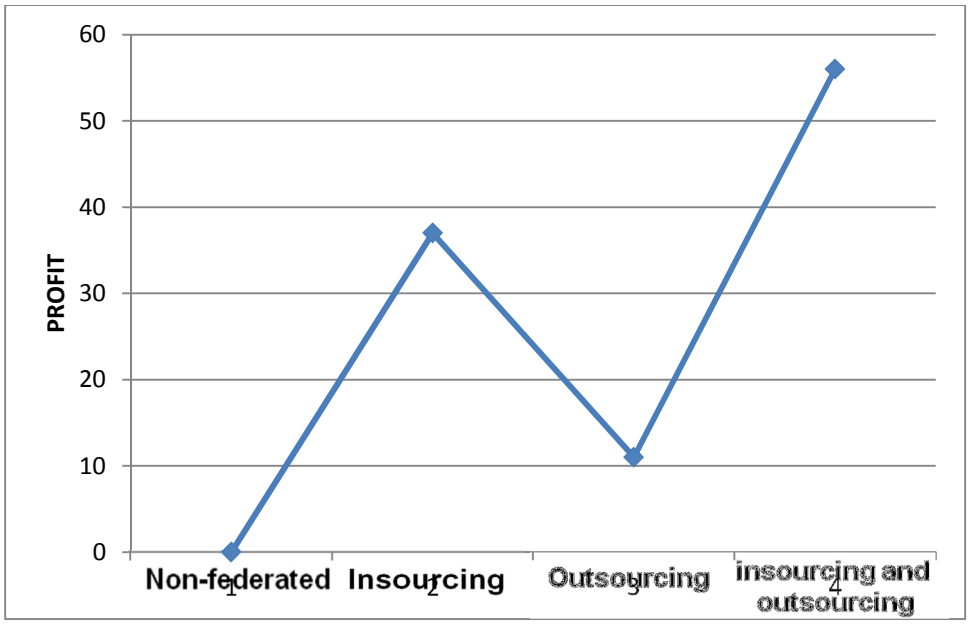

Figure 5.2:- Simulation Results for Profit and Revenue

The provider has many hosts, Hosts have same or different configuration. User's demands change with time which is discussed above. Using traditional resource management methods, the provider becomes incapable to handle the request and rejects the requests which have exceed its current maximum capacity. As a result, cloud provider loses many user requests during busy hours. When a user request is rejected, it results in loss to cloud provider in term of revenue and trust of the consumers. To avoid such situation outsourcing is performed. These results either in the table 1 and table 2 show provider profit is zero when perform experiment for non-federated scenario. Its shows, when provider does not have enough resources and not does able to use the resources of other provider, it rejects the request. Second experiment perform with the special value of contract matrix that is 0.5. It indicates both cloud provider one is requesting for resource of next provider and next provider are getting same amount of revenue.

Contract[Cij] = { {1,0.5,0.5,0.5}, {0.5,1,0.5,0.5}, {0.5,0.5,1,0.5}, {0.5,0.5,0.5,1} }.

	Revenue	**Profit**
Allocation within the provider	**0**	**0**
Insourcing	**30**	**26**
Outsourcing	**30**	**21**
Both insourcing and outsourcing	**60**	**51**

Table 5.2: Simulation Results for Profit and Revenue

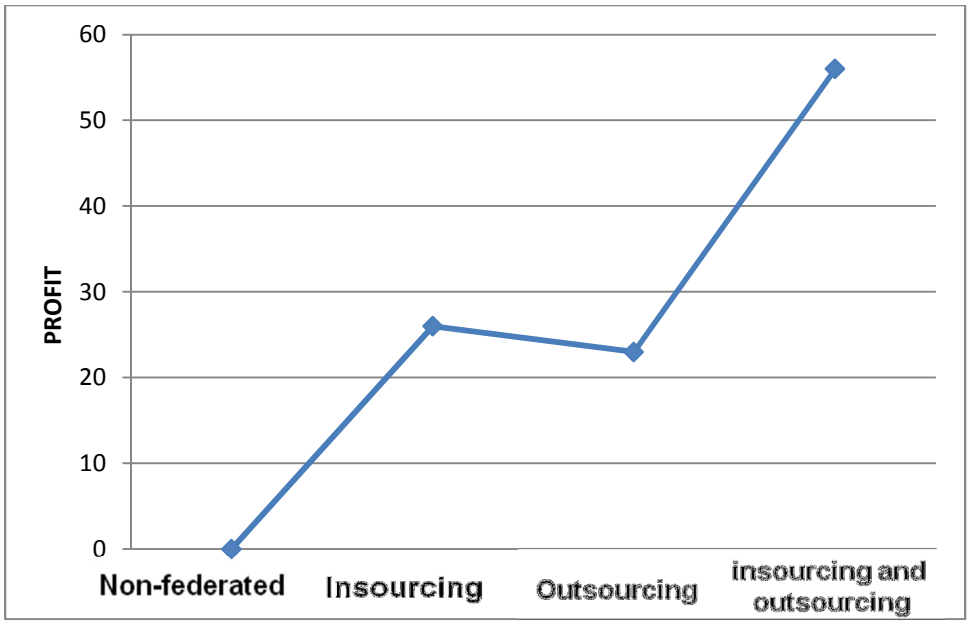

Figure 5.3:- Simulation Results for Profit and Revenue

CHAPTER 6

Conclusion and Scope for Future Work

6.1 Conclusion

In this thesis, a new algorithm is shown that deals with resources, to allocate these resources to virtual machine to enhance the profit in cloud federation environment and also shows how the providers can enhance their profit in cloud federation. Some equations are developed which helps to calculate the revenue and cost in case of outsourcing, insourcing and both (outsourcing and insourcing). Simulation experiments have evaluated these equations to find the result of some parameters in the provider's profit. These parameters include free resources to be sold, the cost of maintaining servers, the cost of outsourcing additional resources, the amount of outsourced resources, and workload.

Proposed algorithm is profitable if all the nodes are up. Minimum utilization and minimum price per hour are not considered in the calculations.

There are number of challenging issues still remain in this interesting topic of cloud computing and one can try to evaluate to obtain results in real cloud computing environments. Further VM provisioning policy based on QoS requirement mention by user can also be evaluated.

Reference:-

1. Google App Engine, http://appengine.google.com.
2. Amazon Elastic Compute Cloud.
3. http://www.amazon.com/ec2.
4. 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*, vol. 41, no. 1, pp. 23–50, Jan. 2011.
5. http://www.vmware.com/virtualization.
6. 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, no. 6, pp. 599–616, 2009.
7. http://en.wikipedia.org/wiki/Cloud_computing .
8. R. Buyya, R. Ranjan, and R. N. Calheiros, "InterCloud: Utility-oriented federation of Cloud computing environments for scaling of application services," in *Proceedings of the 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP'10), ser. Lecture Notes in Computer Science*, vol. 6081. Busan: Springer, May 2010, pp. 13–31.
9. B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, J. Caceres, M. Ben- Yehuda, W. Emmerich, and F. Galan, "The Reservoir model and architecture for open federated Cloud computing," *IBM Journal of Research and Development*, vol. 53, no. 4, pp. 1–11, Jul. 2009.

10. Borja Sotomayor, Kate Keahey, Ian Foster, and Tim Freeman, "Enabling cost-effective resource leases with virtual machines," *ACM/IEEE Internationa Symposium on High Performance Distributed Computing 2007 (HPDC 2007),* 2007.

11. 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.

12. A. Opitz, H. K¨onig, and S. Szamlewska, "What Does Grid Computing Cost?" *Journal of Grid Computing*, vol. 6, no. 4, pp. 385–397, 2008.

13. X. Chu, K. Nadiminti, C. Jin, S. Venugopal, and R. Buyya, "Aneka: Next-Generation Enterprise Grid Platform for e- Science and e-Business Applications," in *3rd IEEE International Conference on e-Science and Grid Computing*, 10–13 December, Bangalore, India, 2007, pp. 151–159.

14. A. Quiroz, H. Kim, M. Parashar, N. Gnanasambandam, and N. Sharma. Towards Autonomic Workload Provisioning for Enterprise Grids and Clouds. *Proceedings of the 10th IEEE/ACM International Conference on Grid Computing (Grid 2009)*, Banf, Alberta, Canada, October 13-15, 2009, IEEE Computer Society Press.

15. Dyna Mietzner, R.; Leymann, F.; "Towards Provisioning the Cloud: On the Usage of Multi-Granularity Flows and Services to Realize a Unified Provisioning Infrastructure for SaaS Applications," *Services - Part I, 2008. IEEE Congress on* , vol., no., pp.3-10, 6-11 July 2008 doi: 10.1109/SERVICES-1.2008.36.

16. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

17. http://en.wikipedia.org/wiki/Service-level_agreement .

18. Clark C, Fraser K, Hand S, "Live Migration of Virtual Machines[C],"*Proceedings of the 2ndInt'l Conference on Networked Systems Design & Implementation. Berkeley, CA, USA, 2005.*

19. Peter Mell and Tim Grance, Definition of *Cloud Computing*. Version 15,. National Institute of Standards and Technology, Information technology vol.24, Special Publication 800-145, 10-7-09.

20. Boghosian, B., Coveney, P., Dong, S., Finn, L., Jha, S., Karniadakis, G. E., and Karonis, N. T. (2006). Nektar, SPICE and Vortonics: Using federated Grids for large scale scientific applications. In IEEE Workshop on Challenges of Large Applications in Distributed Environments (CLADE), Paris, France. IEEE Computing Society.

21. Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., Zagorodnov, D.: The Eucalyptus Open-Source Cloud-Computing System. In: Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2009), Shanghai, China, May 18-May 21 (2010)

22. M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," *University of California at Berkeley*, Tech. Rep. EECS-2009-28, 2009.