Learner Guide

Faculty of Information Technology

Cloud Computing 600

Year 2 *Semester 1*

RICHFIELD

richfield.ac.za

Registered with the Department of Higher Education as a Private Higher Education Institution under the Higher Education Act, 1997.
Registration Certificate No. 2000/HE07/008

FACULTY OF INFORMATION TECHNOLOGY

LEARNER GUIDE

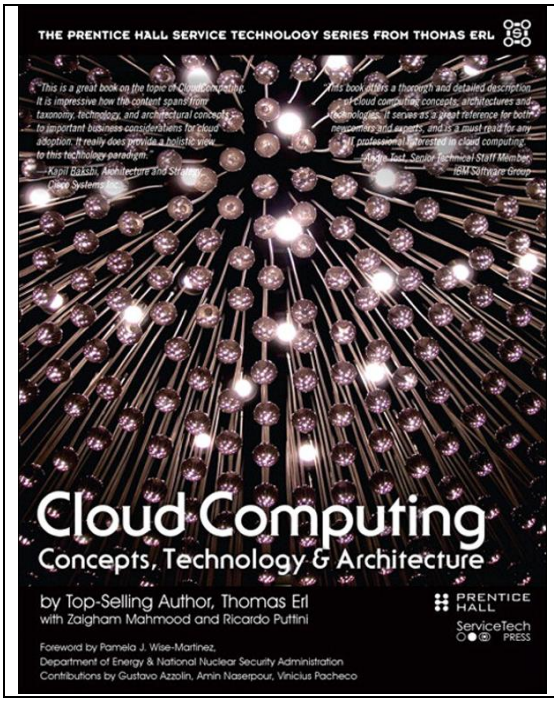
MODULE: CLOUD COMPUTING 600 (1ST SEMESTER)

PREPARED ON BEHALF OF
RICHFIELD GRADUATE INSTITUTE OF TECHNOLOGY (PTY) LTD

GRADUATE INDTITUTE OF TECHNOLOGY (PTY) LTD
Registration Number: 2000/000757/07

Table of Contents

PRESCRIBED OR RECOMMENDED BOOKS



PRESCRIBED

Cloud Computing: Concepts, Technology and Architecture. Zaigham Mahmood, Ricardo Puttini, Thomas Erl

Prentice Hall, Service Tech Press

ISBN: 978-0-133-38756-8

Chapter 1: Introduction to Cloud Computing

Chapter 2: Cloud Computing Architecture and Platforms

Chapter 3: Virtualization

Chapter 4: Cloud Security

Chapter 5: Parallel Processing, Distributed Computing and Storage System

Chapter 6: Web 2.0

Chapter 1: Introduction to Cloud Computing

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- Understand the origin and history of cloud computing
- Explain the evolution of cloud computing.
- Identify factors to consider before deploying application to cloud
- Describe the different concepts and terminologies in cloud computing
- Explain the goals and benefits of cloud computing.
- Identify the risks associated with cloud computing
- Discuss the challenges associated with cloud computing.

1.1 History and Evolution of Cloud Computing

The idea of computing in a "cloud" can be traced back to the origins of utility computing which was a concept proposed in 1961 by a computer scientist called John McCarthy. However, in 1969, Leonard Kleinrock, a chief scientist of the Advanced Research Projects Agency Network or ARPANET project that seeded the Internet, stated:

> *"As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of 'computer utilities' ...".*

The general public has been leveraging forms of Internet-based computer utilities since the mid-1990s through various incarnations of search engines (Yahoo!, Google); e-mail services (Hotmail, Gmail); open publishing platforms (MySpace, Facebook, YouTube), and other types of social media (Twitter, LinkedIn). Though consumer-centric, these services popularized and validated core concepts that form the basis of modern-day cloud computing.

In the late 1990s, Salesforce.com pioneered the notion of bringing remotely provisioned services into the enterprise. In 2002, Amazon.com launched the Amazon Web Services (AWS) platform, a suite of enterprise-oriented services that provide remotely provisioned storage, computing resources, and business functionality.

A slightly different understanding of the term "Network Cloud" or "Cloud" was introduced in the early 1990s throughout the networking industry. It referred to an abstraction layer derived in the delivery methods of data across heterogeneous public and semi-public networks that were primarily packet-switched, although cellular networks used the "Cloud" term as well. The networking method at this point supported the transmission of data from one end-point (local network) to the "Cloud" (wide area network) and then further decomposed to another intended end-point. This is relevant, as the networking industry still references the use of this term, and is considered an early adopter of the concepts that underlie utility computing.

In 2006, the term "cloud computing" emerged in the commercial arena. It was during this time that Amazon launched its Elastic Compute Cloud (EC2) services that enabled organizations to "lease" computing capacity and processing power to run their enterprise applications. Google Apps also began providing browser-based enterprise applications in the same year, and three years later, the Google App Engine became another historic milestone.

1.2 Introduction to Cloud Computing

Cloud computing refers to a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies. It is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

In other words, cloud computing is a term which is used for storing and accessing data over the internet. It does not store any data on the hard disk of your PC, but helps access your data from a remote server.

Common cloud service providers

Some common cloud service providers are listed below;

- Microsoft Azure
- Amazon
- Apple icloud
- Google
- Dropbox
- Salesforce
- Oracle cloud, etc.

Characteristics of Cloud Computing

Some characteristics of cloud computing are briefly described below;

On Demand Self-Service

Cloud computing allow users to use web services and resources on demand. User can log on to a website at any time and use them.

Broad Network Access

Since cloud computing is completely web based, it can be accessed from anywhere and at anytime

Resource Pooling

Cloud computing allows multiple tenants to share a pool of resources. One can share single physical instance of hardware, database and basic infrastructure.

Rapid Elasticity

It is very easy to scale the resources vertically or horizontally at any time. Scaling of resources means the ability of resources to deal with increasing or decreasing demand of clients

Measured Services

Cloud provider controls and monitors all the aspects of cloud service. Resource optimisation, billing, capacity planning, etc depend on it.

Factors to Consider Before Deploying Applications to the Cloud

The following are some of the factors to consider for business requirements before deploying applications to the cloud;

- Data security
- Privacy
- Data backup
- Client-access
- Budget
- Data export
- Type of cloud – private, public or hybrid

Technologies behind Cloud computing

There are certain technologies working behind the cloud computing platforms making cloud computing flexible, reliable and usable. These technologies are briefly described below.

- **Virtualization**

  Virtualization is a technique which allows the sharing of single physical instance of an application or resource among multiple organizations or tenants (customers). It does this by assigning a logical name to a physical resource and providing a pointer to that physical resource when demanded.

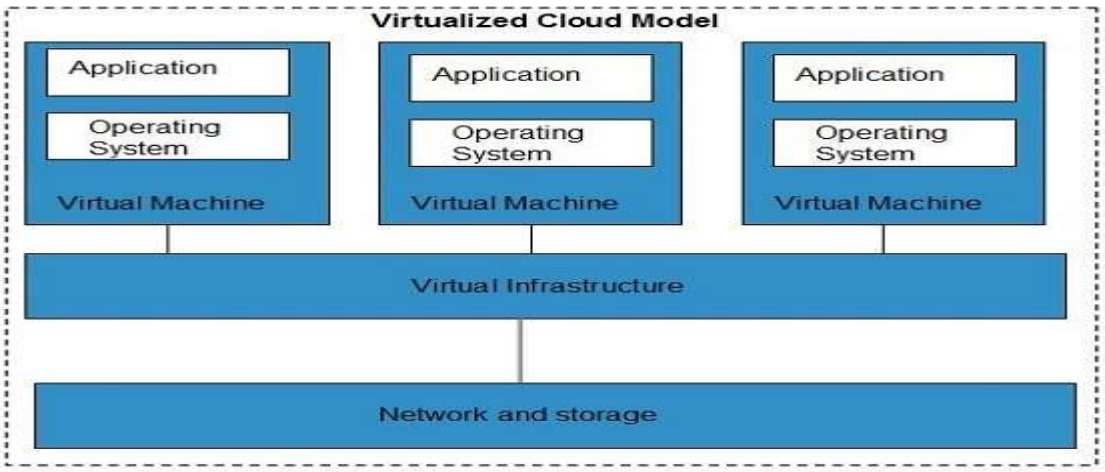


Fig 1.1: Virtualized cloud model

- **Grid computing**

  Grid Computing refers to distributed computing, in which a group of computers from multiple locations are connected with each other to achieve a common objective. These computer resources are heterogeneous and geographically dispersed. Grid Computing breaks complex task into smaller pieces, which are distributed to CPUs that reside within the grid.

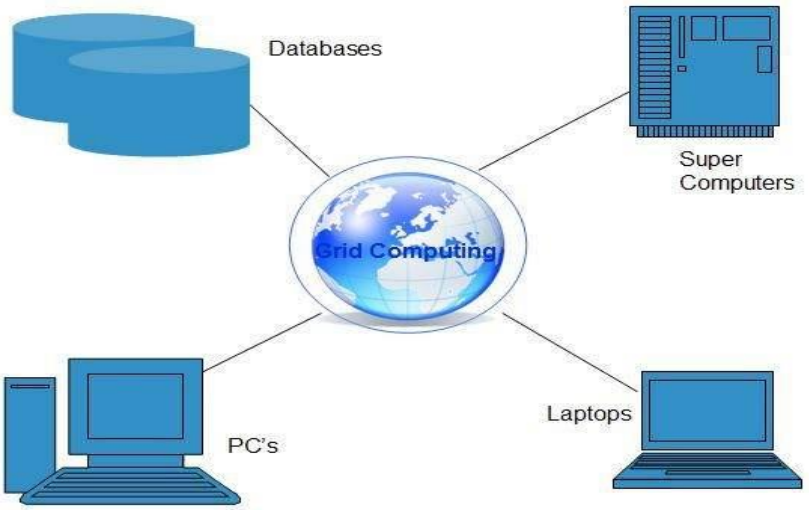


Figure 1.2: Grid Computing

- **Utility computing**

  Utility computing is based on Pay-per-Use model. It offers computational resources on demand as a metered service. Cloud computing, grid computing, and managed IT services are based on the concept of utility computing

Concepts and terminologies in the cloud refers to a set of basic terms that represent the fundamental concepts and aspects pertaining to the notion of a cloud. These terminologies are discussed below;

Cloud

A Cloud refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources. The term originated as a metaphor for the Internet which is, in essence, a network of networks providing remote access to a set of decentralized IT resources. Prior to cloud computing becoming its own formalized IT industry segment, the symbol of a cloud was commonly used to represent the Internet in a variety of specifications and mainstream documentation of Web-based architectures. This same symbol is now used to specifically represent the boundary of a cloud environment as shown in Figure 1.3 below

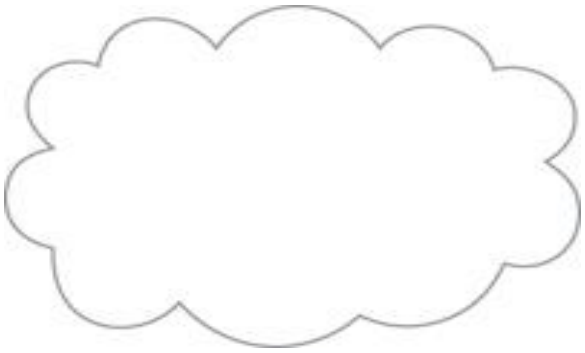


Figure 1.3: The cloud symbol used to denote the boundary of a cloud environment.

It is important to distinguish the term “cloud” and the cloud symbol from the Internet. As a specific environment used to remotely provision IT resources, a cloud has a finite boundary. There are many individual clouds that are accessible via the Internet. Whereas the Internet provides open access to many Web-based IT resources, a cloud is typically privately owned and offers access to IT resources that is metered.

Much of the Internet is dedicated to the access of content-based IT resources published via the World Wide Web. IT resources provided by cloud environments, on the other hand, are dedicated to supplying back-end processing capabilities and user-based access to these capabilities. Another key distinction is that it is not necessary for clouds to be Web-based even if they are commonly based on Internet protocols and technologies. Protocols refer to standards and methods that allow computers to communicate with each other in a pre-defined and structured manner. A cloud can be based on the use of any protocols that allow for the remote access to its IT resources.

IT Resource

An IT resource is a physical or virtual IT-related artifact that can be either software-based, such as a virtual server or a custom software program, or hardware-based, such as a physical server or a network device. This is shown in Figure 1.4 below

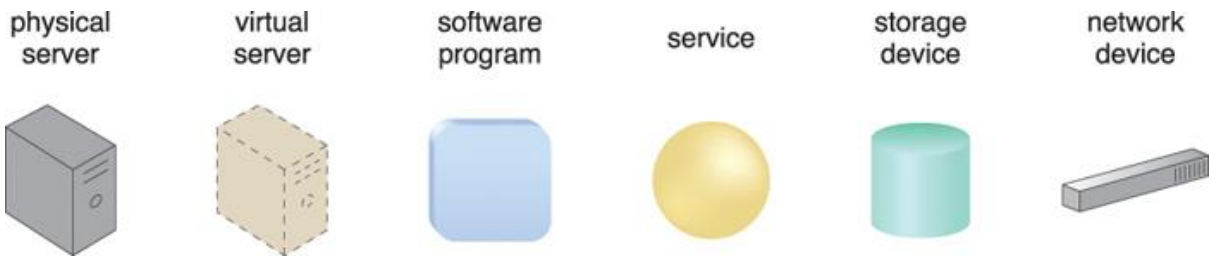
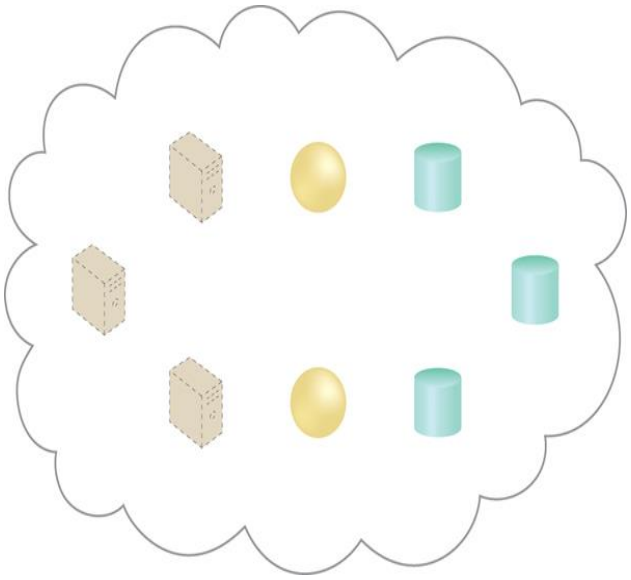


Figure 1.4 Common IT resources and their corresponding symbols.

Technology architectures and various interaction scenarios involving IT resources are illustrated in Figure 1.3 below. It shows how the cloud symbol can be used to define a boundary for a cloud-based environment that hosts and provisions a set of IT resources. The displayed IT resources are consequently considered to be cloud-based IT resources.



A cloud hosting eight IT resources: Three virtual servers, Two cloud services, and Three storage devices

Fig 1.5: Cloud symbol representing a cloud-based environment

On-premise

An IT resource that is hosted in a conventional IT enterprise within an organizational boundary (that does not specifically represent a cloud) is considered to be located on the premises of the IT

enterprise, or on-premise for short. In other words, the term “on-premise” is another way of stating “on the premises of a controlled IT environment that is not cloud-based.” This term is used to qualify an IT resource as an alternative to “cloud-based.” An IT resource that is on-premise cannot be cloud-based, and vice-versa.

When discussing on-premise, the following points need to be noted;

• An on-premise IT resource can access and interact with a cloud-based IT resource.

• An on-premise IT resource can be moved to a cloud, thereby changing it to a cloud-based IT resource.

• Redundant deployments of an IT resource can exist in both on-premise and cloud-based environments.

Cloud consumers and Cloud Providers

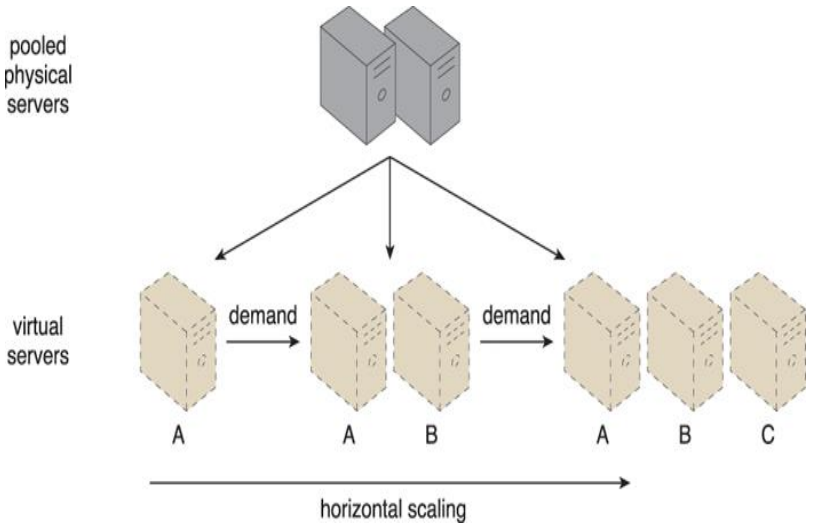
The cloud provider refers to the “party” that provides cloud-based IT resources, while the party that uses the cloud-based IT resources is the cloud consumer. These terms represent roles usually assumed by organizations in relation to clouds and corresponding cloud provisioning contracts.

Scaling

The term “scaling” from an IT resource perspective, represents the ability of the IT resource to handle increased or decreased usage demands. There are two types of scaling - Horizontal Scaling (scaling out and scaling in); Vertical Scaling (scaling up and down).

Horizontal Scaling

This refers to allocating or releasing of IT resources that are of the same type. The horizontal allocation of resources is referred to as *scaling out* and the horizontal releasing of resources is referred to as *scaling in*. Horizontal scaling is a common form of scaling within cloud environments.

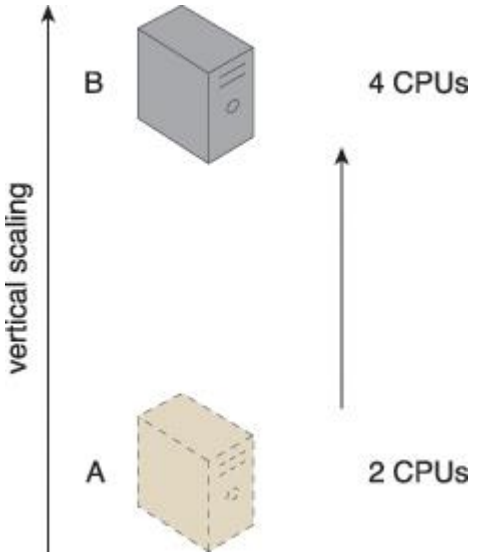


An IT resource (Virtual Server A) is scaled out by adding more of the same IT resources (Virtual Servers B and C).

Figure 1.6: Horizontal scaling

Vertical Scaling

Vertical scaling is when an existing IT resource is replaced by another with higher or lower capacity. Specifically, the replacing of an IT resource with another that has a higher capacity is referred to as scaling up, while the replacing an IT resource with another that has a lower capacity is considered scaling down. Vertical scaling is less common in cloud environments due to the downtime required while the replacement is taking place.



An IT resource (a virtual server with two CPUs) is scaled up by replacing it with a more powerful IT resource with increased capacity for data storage (a physical server with four CPUs).

Figure 1.7 : Vertical scaling

The table below provides an overview of pros and cons associated with horizontal and vertical scaling

Horizontal Scaling	Vertical Scaling
less expensive (through commodity hardware components)	more expensive (specialized servers)
IT resources instantly available	IT resources normally instantly available
resource replication and automated scaling	additional setup is normally needed
additional IT resources needed	no additional IT resources needed
not limited by hardware capacity	limited by maximum hardware capacity

Table 1.1 A comparison between horizontal and vertical scaling.

Cloud Service

A cloud service is any IT resource that is made remotely accessible via a cloud. Unlike other IT fields that fall under the service technology umbrella—such as service-oriented architecture—the term "service" within the context of cloud computing is especially broad. A cloud service can exist as a simple Web-based software program with a technical interface invoked via the use of a messaging protocol, or as a remote access point for administrative tools or larger environments and other IT resources.

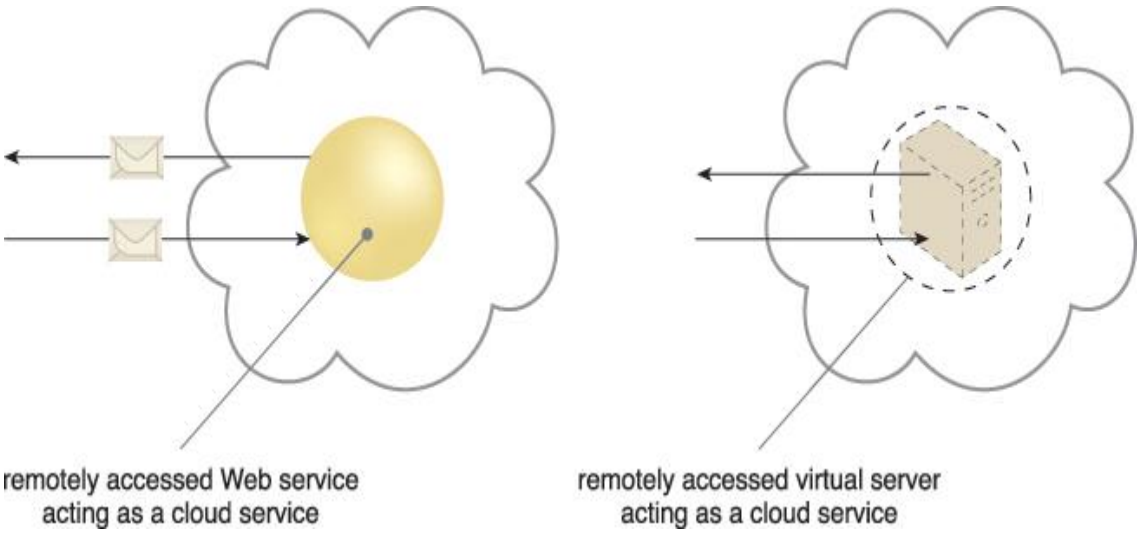


Fig 1.8: Cloud Service

A cloud service with a published technical interface is being accessed by a consumer outside of the cloud (left). A cloud service that exists as a virtual server is also being accessed from outside of the cloud's boundary (right). The cloud service on the left is likely being invoked by a consumer program that was designed to access the cloud service's published technical interface. The cloud service on the right may be accessed by a human user that has remotely logged on to the virtual server.

The driving motivation behind cloud computing is to provide IT resources as services that encapsulate other IT resources, while offering functions for clients to use and leverage remotely

Cloud service consumer

The cloud service consumer is a temporary runtime role assumed by a software program when it accesses a cloud service. Common types of cloud service consumers can include software programs and services capable of remotely accessing cloud services with published service contracts, as well as workstations, laptops and mobile devices running software capable of remotely accessing other IT resources positioned as cloud services.

Figure 1.9 Examples of cloud service consumers

An artifact labeled as a cloud service consumer may be a software program or a hardware device (in which case it is implied that it is running a software program capable of acting as a cloud service consumer).

1.4 Goals and Benefits of Cloud Computing

The driving motivation behind cloud computing is to provide IT resources as services that encapsulate other IT resources, while offering functions for clients to use and leverage remotely. Some common benefits of cloud computing are;

- Increased Scalability
- Increased Availability
- Increased Reliability
- One can manipulate and configure the applications online at any time.

- It does not require a user to install a software to access or manipulate cloud application.
- Cloud computing offers online development and deployment tools, programming runtime environment through PaaS model.
- Cloud resources are available over the network in a manner that provide platform independent access to any type of clients.
- Cloud computing offers on-demand self-service. The resources can be used without interaction with cloud service provider.
- Cloud computing is highly cost-effective because it operates at high efficiency with optimum utilization. It just requires an Internet connection.

1.5 Risks and Challenges associated with Cloud Computing

In as much as cloud computing has its benefits, there are also risks and challenges associated with cloud computing. Some of these risks and challenges are described below;

Security and Privacy

It is the biggest concern about cloud computing. Since data management and infrastructure management in cloud is provided by a third-party, it is always a risk to handover the sensitive data or information to cloud service providers. Although the cloud computing vendors ensure highly secured password protected accounts, any sign of security breach may result in loss of customers and businesses.

Limited portability between cloud providers

Due to a lack of established industry standards within the cloud computing industry. it can be challenging to migrate from one cloud provider to another. Portability is a measure used to determine the impact of moving cloud consumer IT resources and data between clouds

Multi-regional compliance and legal issues

Cloud consumers will often not be aware of the physical location of their IT resources and data when hosted by public clouds. This can pose serious legal concerns pertaining to industry or government regulations that specify data privacy and storage policies. Another potential legal issue pertains to the accessibility and disclosure of data. Countries have laws that require some types of data to be disclosed to certain government agencies or to the subject of the data.

Increased Security Vulnerabilities

The moving of business data to the cloud means that the responsibility over data security becomes shared with the cloud provider. The remote usage of IT resources requires an expansion of trust boundaries by the cloud consumer to include the external cloud. It can be difficult to establish a security architecture that spans such a trust boundary without introducing vulnerabilities, unless cloud consumers and cloud providers happen to support the same or compatible security frameworks—which is unlikely with public clouds.

Vendor Lock-in

It is very difficult for the customers to switch from one Cloud Service Provider (CSP) to another. It results in dependency on just a particular CSP for service.

Insecure or incomplete data deletion

It is possible that the data requested for deletion by a cloud consumer may not get deleted. It happens either because extra copies of data are stored but are not available at the time of deletion, or the disk that stores data of multiple tenants is destroyed.

Interoperability

Application on one platform should be able to incorporate services from other various platforms made possible via web services, but developing such web services is very complex.

1.6 Chapter Summary

Cloud computing refers to manipulating, configuring, and accessing hardware and software resources remotely with minimal management effort or service-provider interaction. Some of the factors to consider for business requirement before deploying application to the cloud are data security, data back-up, privacy, budget, client-access, etc. Some benefits of cloud computing include: increased scalability, increased availability, reliability, cost-effectiveness, high efficiency. Risks and challenges associated with cloud computing are: security, privacy, multiregional compliance and legal issues, limited portability between cloud service providers, Lock-in, insecure or incomplete data deletion, and increased security vulnerabilities.

1.7 Review Questions

1. **Explain the term Cloud Computing?**
2. **Identify some common cloud service providers.**
3. **What is the difference between a cloud consumer and cloud service provider. Give example of each**
4. **Define the term "scaling", and highlight the difference between vertical and horizontal scaling.**
5. **What are the pros and cons of vertical and horizontal scaling?**
6. **Differentiate between Virtualization and Grid computing? Use diagram to enhance your explanation.**
7. **Discuss risks and challenges associated with cloud computing?**
8. **Explain some benefits of cloud computing?**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 3,pg *77-102*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran

1.8 MCQs (Quick Quiz)

1. Which of the following is not a benefit of cloud computing?

a) Reliability
b) Scalability
c) Availability
d) Vulnerability

2. _____is the party that provides cloud-based IT resources?

a) Cloud consumer
b) Cloud manager
c) Cloud provider
d) Cloud customer

3. Which of the following is not a risk associated with cloud computing?

a) Security
b) Privacy
c) Incomplete data deletion
d) Reliability

4. The ability of IT resources to accommodate increasing fluctuations in demands can be termed as?

a) Reliability
b) Vulnerability
c) Availability
d) Scalability

5. _____ refers to the ability of the IT resource to handle increased or decreased usage demands.

a) Scaling
b) Reliability
c) Redundancy
d) Availability

6. The technique which allows to share single physical instance of application or resources among multiple organizations is called ?

a) Grid computing
b) Parallel computing
c) Virtualization'
d) Processing computing

Chapter 2: Cloud Computing Architecture and Platforms

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- **Identify the architecture and infrastructure of cloud computing.**
- **Evaluate various cloud service delivery models - IaaS, PaaS, and SaaS.**
- **Explain the characteristics of IaaS, PaaS and SaaS.**
- **Discuss the challenges associated with IaaS, PaaS and SaaS.**
- **Describe private, public, hybrid and community cloud.**
- **Discuss the advantages and disadvantages of private and public cloud.**
- **Differentiate between public and community cloud.**

2.1 Introduction

There are certain services and models working behind the scene making cloud computing feasible and accessible to end users. This chapter analyse various service model and deployment models in cloud computing.

2.2 Cloud Service Delivery Model

A cloud service delivery model represents a specific, pre-packaged combination of IT resources offered by a cloud provider. There are three (3) common cloud delivery models, namely;

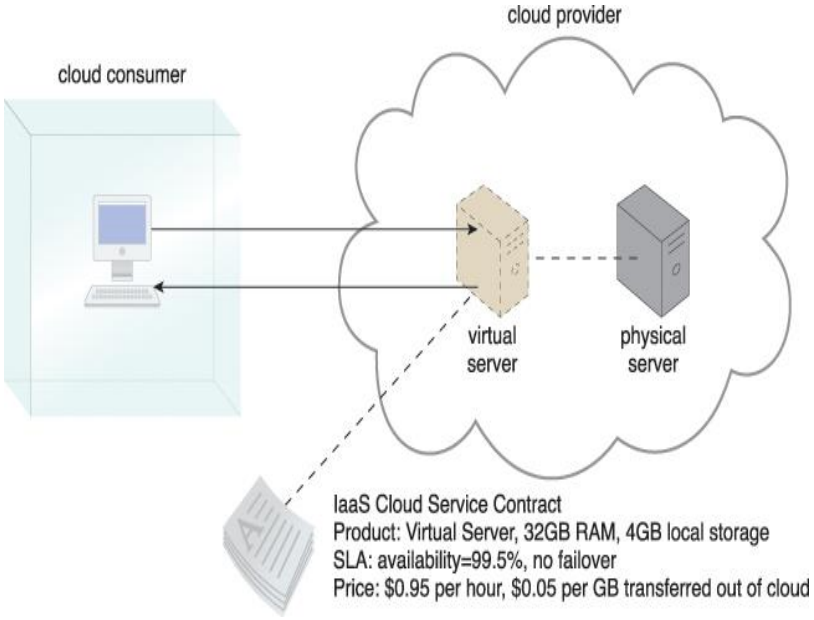
- Infrastructure-as-a- Service (IaaS)
- Platform-as-a- Service (PaaS)
- Software-as-a- Service (SaaS)

2.2.1 Infrastructure-as-a- Service (IaaS)

The IaaS delivery model represents a self-contained IT environment comprised of infrastructure-centric IT resources that can be accessed and managed via cloud service-based interfaces and tools. This environment can include hardware, network, connectivity, operating systems and other IT resources. In contrast to traditional hosting or outsourcing environment, with IaaS, IT resources are

typically virtualised and packaged into bundles that simplify up-front runtime scaling and customisation of the infrastructure. The general purpose of an IaaS environment is to provide cloud consumers with a high level of control and responsibility over its configuration and utilisation.

IaaS provide access to fundamental resources such as physical machines, virtual machines, virtual storage, etc. Iaas also offers Virtual machine disk storage, IP addresses, Virtual local area network (VLANS),etc.



A cloud consumer is using a virtual server within an IaaS environment. Cloud consumers are provided with a range of contractual guarantees by the cloud provider, pertaining to characteristics such as capacity, performance, and availability.

Fig 2.1: IaaS

Characteristics of IaaS

- Virtual machines with pre-installed software.
- Virtual machines with pre-installed operating systems such as Windows, Linux, and Solaris.
- On-demand availability of resources.
- Allows to store copies of particular data at different locations.
- The computing resources can be easily scaled up and down.

Benefits of IaaS

IaaS allows the cloud provider to freely locate the infrastructure over the Internet in a cost-effective manner. Some of the key benefits of IaaS are listed below:

- Full control of the computing resources through administrative access to VMs.
- Flexible and efficient renting of computer hardware.
- Portability, interoperability with legacy applications.

Challenges associated with IaaS

- Data erasure practices.
- Compatibility with legacy security and vulnerabilities.
- Robustness of VM level isolation

2.2.2 Platform-as-a- Service (PaaS)

The PaaS delivery model represents a pre-defined "ready-to-use" environment typically comprised of already deployed and configured IT resources. PaaS specifically relies on the usage of a ready-made environment that establishes a set of pre-packaged products and tools used to support the entire delivery lifecycle of custom applications. App Engine of Google and Force.com are example of PaaS offering vendors. Developer may log on to these websites and use the built-in Application Programming Interface (API) to create web-based applications. By working within a ready-made platform, the cloud consumer is spared the administrative burden of setting up and maintaining the bare infrastructure IT resources provided via the IaaS model.

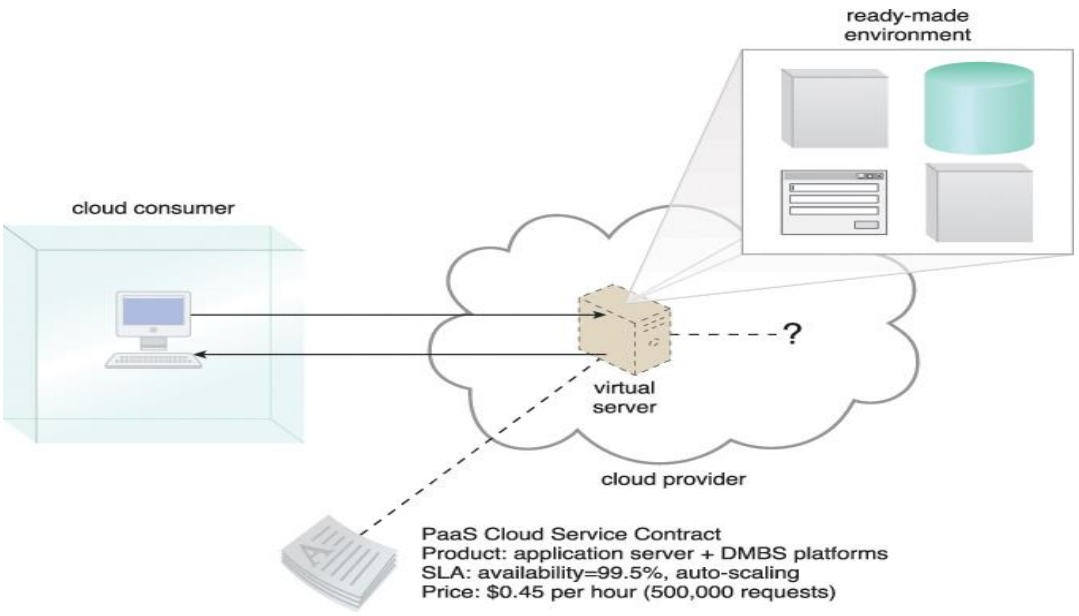


Fig 2.1: PaaS

Characteristics of PaaS

- PaaS offers browser-based development environment. It allows the developer to create database and edit the application code either via Application Programming Interface or point-and-click tools.
- PaaS provides built-in security, scalability and web service interfaces
- PaaS provides built-in tools for defining workflow, approval processes and business rules
- It is easy to integrate PaaS with other applications on the same platform.

Benefits of PaaS

- Scalable solutions
- Lower total cost of ownership
- Lower administrative overhead
- More current system software

Challenges associated with PaaS

Just like SaaS, PaaS also places significant burden on customer's browser to maintain reliable and secure connection to the provider's system. Other specific issues associated with PaaS are;

- Lack of portability between PaaS clouds
- Event based processor scheduling
- Security engineering of PaaS application

2.2.3 Software-as-a- Service (SaaS)

SaaS model provides software application as a service to end users. It refers to a software that is deployed on a host service and is accessible via the internet. Some common examples of SaaS application are - Billing and Invoicing system, Help desk application, Customer Relationship Management (CRM) application, etc.

Characteristics of SaaS

- SaaS makes the software available over the internet
- The software applications are maintained by the vendor
- SaaS applications are cost-effective since they do not require any maintenance on the end-user side.
- They are available on demand

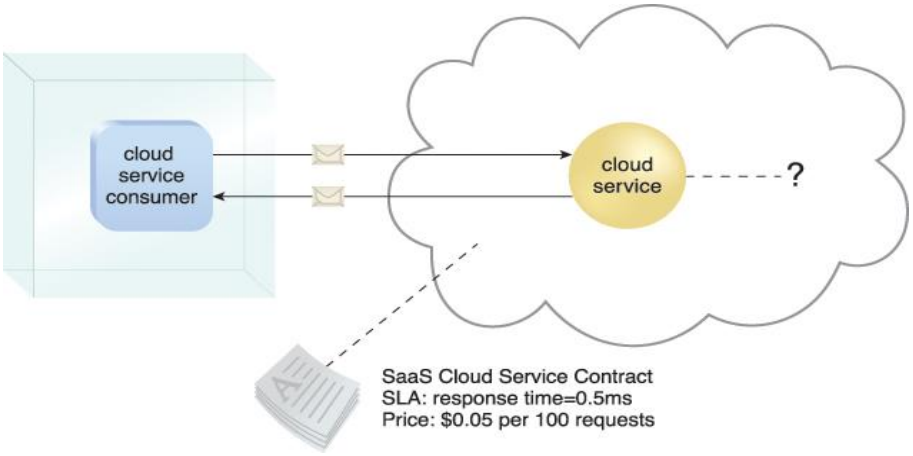


Fig 2.2: SaaS

Benefits of SaaS

- Scalable and efficient
- Centralised management and data
- Multitenant solutions
- The provider is responsible for managing platforms

Challenges associated with SaaS

- Network dependent
- Browser- based risks
- Multitenant solutions
- Lack of portability between SaaS clouds

2.3 Cloud Deployment Model

A cloud deployment model represents a specific type of cloud environment primarily distinguished by ownership, size and access.

The four (4) common cloud deployment models are;

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

2.3.1 Private cloud

A private cloud is owned by a single organisation. In a private cloud, systems and services are only operated and accessible within the organisation.

Benefits of private cloud

- High security
- High privacy
- More control
- Improved reliability

Disadvantages of private cloud

- Limited scalability
- Restricted area of operation

2.3.2 Public cloud

Public cloud allows systems and services to be easily accessible to the general public. Google, Amazon and Microsoft all offer cloud services via the internet.

Benefits of public cloud

- Highly flexible
- Highly scalable
- Highly reliable
- Cost-effective

Disadvantages of public cloud

- Low security
- Less customizable

2.3.3 Hybrid cloud

This is a combination of public and private cloud. A consumer may choose to deploy cloud services processing sensitive data to a private cloud; and less sensitive data to a public cloud

Benefits of hybrid cloud

- Scalable

- Cost-efficient
- Flexibility
- Security

Disadvantages of hybrid cloud

- Networking issues
- Security compliance
- Infrastructure dependency

2.3.3 Community cloud

Community cloud allows systems and services to be accessible by group of organisations. It is similar to a public cloud except that its access to limited to a specific community of cloud consumers.

Benefits of community cloud

- Cost-effective
- Centralised management and data
- More secure than public cloud, but less secure than private cloud.

Disadvantages of community cloud

- Challenging to allocate responsibilities of governance, security and cost among organisation
- Since all the data is located at one place, it might be accessible to others

2.4 Chapter Summary

The different types of cloud service delivery model are Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (Paas) and Software-as-a-Service (Saas). PaaS delivery model represents a pre-defined ready to use environment typically comprised of already deployed and configured IT resources. SaaS model provides software application as a service to end-users which can be accessed via the internet. Also, the different cloud deployment models are private cloud, public, hybrid and community cloud. The private cloud is owned and operated by a single organisation, the public cloud allows systems and services to be easily accessible to the general public, while the hybrid cloud is a combination of both private and public cloud.

2.5 Review Questions

1. **Briefly describe IaaS and state the challenges associated with it.**
2. **Explain the difference between SaaS and PaaS?**
3. **Which category of cloud service delivery model is Microsoft, Google and Amazon. Justify your answer by describing the delivery model and also list the characteristics of the delivery model.**
4. **Critically analyse private, public and hybrid cloud deployment model?**
5. **What are the advantages and disadvantages of private and public cloud?**
6. **Evaluate the differences between public and community cloud?**
7. **Explain hybrid cloud deployment model and highlight its pros and cons?**
8. **Which of the various cloud deployment models do you think is the most secure? Justify your answer with suitable explanation?**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 3, pg *115-120*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran

2.6 MCQs (Quick Quiz)

1. **Which of the following is not a type of cloud service delivery model?**

a) Software-as-a-service
b) Infrastructure-as-a-Service
c) Hardware-as-as-a-Service
d) Platform-as-a-Service

2. **_____is a type of cloud comprised of 2 or more different models?**

a) Private cloud
b) Community cloud
c) Public cloud
d) Hybrid cloud

3. **Which of the following is not a benefit of private cloud deployment model?**

a) High security
b) Low security
c) Privacy
d) More control

4. **In which deployment model would security compliance be a disadvantage?**

a) Private cloud
b) Public cloud
c) Hybrid cloud
d) Community cloud

5. **Which service model provide application as a service to end users over the internet.**

a) SaaS
b) IaaS
c) PaaS
d) HaaS

6. **Which delivery model represent a pre-defined ready to use environment typically comprised of already deployed and configured IT resources?**

a) HaaS
b) Iaas
c) PaaS
d) SaaS

2.7 Case Study

A private college based in Umhlanga Durban is contemplating whether to adopt Private or hybrid cloud deployment model for their day-to-day operations. Assume you have been assigned as the I.T manager of the company, which deployment model would you advise the company to adopt?

Questions

1. **Describe the deployment model you would suggest to your company?**
2. **Explain the pros of the deployment model that you have suggested?**
3. **Elaborate on the cons of the other deployment model you did not suggest hence justifying what makes them unsuitable?**

Chapter 3: Virtualization

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- **Understand the concept of virtualization**
- **Explain the function of hypervisor**
- **Describe various virtualization approach**
- **Elaborate on the different types of virtualization**
- **Understand how the different resources can be virtualized.**
- **Discuss multitenancy, its advantages and disadvantages.**

3.1 Virtualization Technology

Virtualization is a technique to share single physical instance of an application or resource among multiple organisations or tenants (customers). It does so by assigning a logical name to a physical resource and providing a pointer to that physical resource on demand. Most types of IT resources can be virtualized.

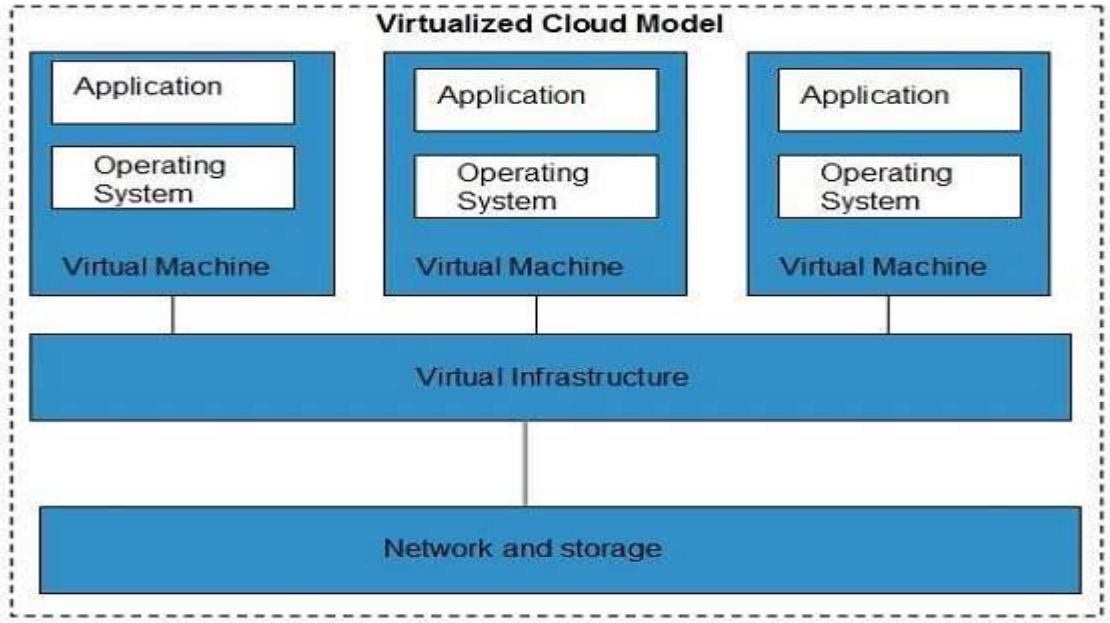


Fig 3.1: Virtualized cloud model

The concept of creating a virtual machine or server over an existing operating system and hardware is called Hardware virtualization. Virtual machines provide an environment that is logically separated from the underlying hardware. The machine on which the virtual machine is created is referred to as Host machine, while the virtual machine is called the guest machine. This virtual machine is managed by a software or firmware called "Hypervisor". Hypervisor is the software or firmware that manages the virtual machine. The hypervisor acts as a virtual machine manager. They help run multiple OS concurrently on a physical system sharing its hardware, hence, it allows multiple OS to share a single hardware host. The most important function of virtualization is the capability of running multiple OS and applications on a single computer or server.

The first step in creating a new virtual server or machine through virtualization software is the allocation of physical IT resources, followed by the installation of an operating system. Virtual servers use their own guest operating systems, which are independent of the operating system in which they were created.

3.2 Benefits of Virtualization

- Efficient resource utilization
- Lower costs and extended life of the technology
- Increased return on investments (ROI) for both the service providers and the consumers
- Promotes green IT by reducing energy wastage
- Dynamic data centre
- Improve disaster recovery
- Eases administration

3.3 Disadvantages of Virtualization

- High cost of implementation
- Scalability issues
- Requires high-end and powerful infrastructure
- Security risks
- Requires specialised skills set

3.4 Virtualization Approach

Some approach in the implementation of virtualization are described below;

Full virtualization

Full virtualization uses a special kind of software called hypervisor. The hypervisor interacts directly with the physical server's hardware resources (such as CPU and storage space), and acts as a platform for the virtual server's OS. It helps to keep each virtual server completely independent and unaware of the other virtual servers running on the physical machine. Each guest server or the virtual machine is able to run its own OS. Example is VirtualBox. A disadvantage of this virtualization is that overall system performance may be affected due to binary translation.

Para virtualization

In para virtualization, virtual machines do not simulate the underlying hardware, and this uses a special API that a modified guest OS must use. Also, partial simulation of the underlying hardware infrastructure is achieved. The guest OS is aware that it is running in a virtualized environment. Hypercalls are used for direct communication between guest OS and the hypervisor. An advantage of this approach is that it improves the overall system performance by eliminating the overhead of binary translation, while a disadvantage could be that a modification of the guest OS is required.

Hardware-Assisted virtualization

In hardware-assisted virtualization, hardware products supporting the virtualization are used. Hardware vendors like Intel and AMD have developed processors supporting the virtualization through the hardware extension. An advantage of this approach is that it eliminates the overhead of binary translation and paravirtualization, while a disadvantage is the lack of support from all vendors

3.5 Types of virtualization

The process of virtualization can be classified into the following depending on the resource that is being virtualized;

Server virtualization

In this type of virtualization, existing physical servers are moved into a virtual environment which is then hosted on a physical server. Modern servers can host more than one server simultaneously which allows the users to reduce the number of servers to be reserved for various purposes.

Operating System (OS) virtualization

Operating system-based virtualization is the installation of virtualization software in a pre-existing operating system, which is called the host operating system. Since the host operating system can provide hardware devices with the necessary support, operating system virtualization can rectify hardware compatibility issues even if the hardware driver is not available to the virtualization software. A concern with operating system-based virtualization is the processing overhead required to run the virtualization software and host operating systems. Implementing a virtualization layer will negatively affect overall system performance. Estimating, monitoring, and managing the resulting impact can be challenging because it requires expertise in system workloads, software and hardware environments, and sophisticated monitoring tools.

Memory virtualization

In main memory virtualization, the virtual main memory that is abstracted from the physical memory is allocated to various virtual machines to meet their memory requirements. The mapping of physical to virtual memory is performed by the hypervisor software.

Storage virtualization

Multiple physical hard drives are combined into a single virtualised storage environment. To different users, this is simply called “cloud storage”, and it could either be private storage (hosted by a company), or a public storage (hosted outside of a company), or mixed approach (i.e. private and public storage).

Application virtualization

In this type of virtualization, the single application installed on the central server is virtualized, and the various virtualized components of the application will be given to the users requesting the services.

Hardware virtualization

This is the installation of virtualization software directly on the physical host hardware so as to bypass the host operating system, which is presumably engaged with operating system-based virtualization Allowing the virtual servers to interact with hardware without requiring intermediary action from the host operating system generally makes hardware-based virtualization more efficient.

One of the main issues of hardware-based virtualization concerns compatibility with hardware devices. The virtualization layer is designed to communicate directly with the host hardware, meaning all of the associated device drivers and support software need to be compatible with the hypervisor. Hardware device drivers may not be as available to hypervisor platforms as they are to operating

systems. Host management and administration features may further not include the range of advanced functions that are common to operating systems.

3.6 Multitenant Technology

Multitenancy refers to a software architecture design in which a single instance of a software application serves multiple tenants (users). Multitenant application enables multiple users (tenants) to access the same application simultaneously. It allows multiple users (tenants) to work in a software environment all at the same time, each with their own separate user instances, resources and services. Multitenant application ensures that tenants do not have access to data and configuration information that is not their own. Tenants can individually customize features of the application such as: User interface, access control, data security, application upgrade, scalability, etc.

SaaS-based software applications delivered over the internet is a common example of multitenant architecture where a single application is accessed by many users globally. Multitenant application architecture is often significantly more complex than that of single-tenant applications because they need to support the sharing of various artefacts.

Advantages of multitenancy

- Helps to cut cost of investment
- It is relatively easy to add new tenants (customers)
- It is convenient to maintain same application, thereby maximising resource usage
- Holds multiple tenants all at the same time

Disadvantages of multitenancy

- Security
- Limited customization
- Rigid service levels
- Potential cost of re-architecture
- Increased operational and infrastructural costs and people's skills
- Service delivery

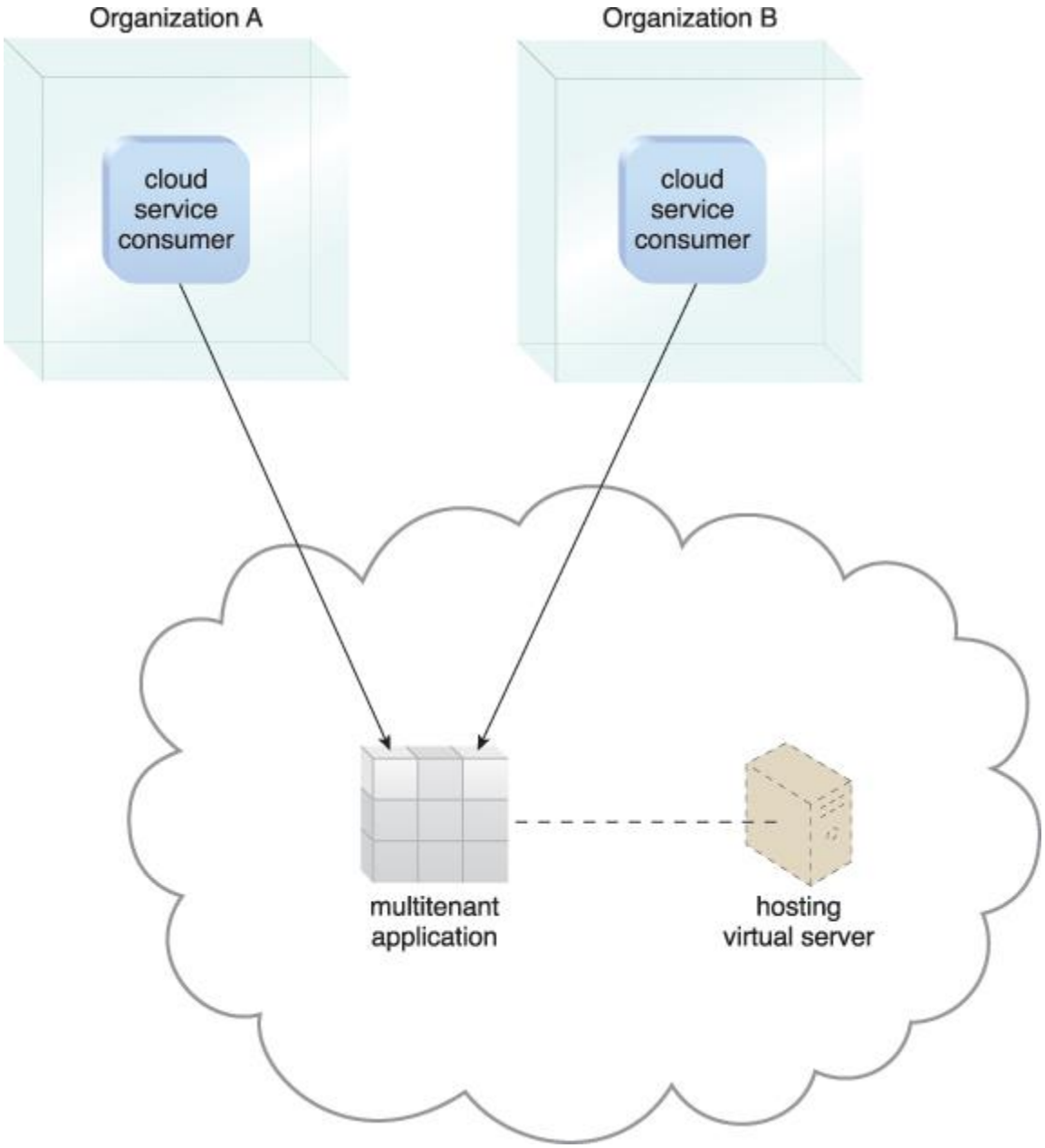


Fig 3.2: A multitenant application serving multiple cloud service consumers simultaneously.

3.7 Chapter Summary

Virtualization is a technique used to share single physical instance of an application or resource among multiple organisation or tenants (customers). Some benefits of virtualization are - efficient resource utilization, increased return on investment for both service provider and consumer, promoting green IT and improve disaster recovery. The approach in the implementation of virtualization are – Full virtualization, para virtualization and Hardware-assisted virtualization. The various types of resource being virtualized are – server virtualization, OS virtualization, memory virtualization, storage virtualization, application virtualization and Hardware virtualization. Multitenant application enables multiple users (tenants) to access the same application simultaneously. Some benefits of multitenancy are – it is convenient to maintain same application, it holds multiple tenants all at the same time, and it is easy to add new tenants. Challenges are – security, limited customization, potential cost of re-architecture.

3.8 Review Questions

1. **Explain the concept of Virtualization?**
2. **What is an "Hypervisor", and explain its function?**
3. **Discuss the benefits of Virtualization?**
4. **What are the challenges associated with virtualization?**
5. **Briefly explain three (3) approach in the implementation of Virtualization**
6. **Discuss operating system virtualization, server virtualization, application virtualization and storage virtualization.**
7. **Explain the concept of multitenancy?**
8. **What are the advantages and disadvantages of multitenancy?**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 5, pg *148-154*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran, Chapter 7, pg 161 -187

3.9 MCQs (Quick Quiz)

1. Which of the following is not a benefit of Virtualization?

a) Efficient resource utilization
b) Disaster recovery
c) Promotes green computing
d) Require specialized skills set

2. In _____ virtualization, a disadvantage is that overall system performance may be affected due to binary translation?

a) Full
b) Para
c) Hardware-assisted
d) Partial

3 Which of the following is not a classification of resources that can be virtualized?

a) Server
b) Operating system
c) Storage
d) CMOS

4. In what type of virtualization does the single application installed on the central server is centralised?

a) Server virtualization
b) Application virtualization
c) Storage virtualization
d) OS virtualization

5. What is the software that manages the virtual machine called?

a) Operating system
b) Server
c) hypervisor
d) storage machine

6. Which of these is not a challenge associated with virtualization

a) Single point of failure
b) May lead to lower performance
c) Demands high-end and powerful infrastructure
d) Ease of administration

Chapter 4: Cloud Security

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- **Understand the fundamentals of cloud security**
- **Explain various cloud security terms and concepts**
- **Differentiate between Confidentiality, Integrity and Availability**
- **Identify risks and vulnerability in a system**
- **Understand various cloud security threats**
- **Develop practices to mitigate cloud security threats**

4.1 Introduction

Security is the biggest concern about cloud computing. Since data management and infrastructure management in cloud is provided by a third-party, it is always a risk to handover the sensitive data or information to cloud service providers. Although the cloud computing vendors ensure highly secured password protected accounts, any sign of security breach may result in loss of data, customers and even businesses. It is therefore necessary to have security measures to defend against threats and interference that arise malicious intent.

4.2 Basic Terms and Concept

Fundamental security terms and concepts relevant to cloud computing are briefly discussed below;

4.2.1 Confidentiality

Confidentiality refers to the characteristics of making data, information or resources accessible only to authorized parties. Within cloud environments, confidentiality is restricting access to data in transit and storage.

4.2.2 Integrity

Integrity refers to the characteristics of having data or information not been altered by an unauthorized party. Data integrity concerns itself with whether a cloud consumer can be guaranteed that the data it transmits to a cloud service matches the data received by that cloud service.

4.2.3 Authenticity

Authenticity is the characteristics of having data or information provided by authorized source. Authenticity encompasses Non-repudiation – which is the inability of a party to deny or challenge the authentication of an interaction.

4.2.4 Availability

Availability is the characteristics of having data or information being accessible and usable during a specified period of time. The availability of cloud services can be a responsibility that is shared by the cloud provider and the cloud carrier

4.2.5 Threat

A threat is a set of circumstances that has the potential to cause loss or harm to a system. A threat that is carried out result in an attack.

4.2.6 Vulnerability

A vulnerability is a weakness in a system which can be exploited either because it is protected by insufficient security controls, or existing security controls are overcome by an attack. IT resource vulnerabilities can have a range of causes, including configuration deficiencies, security policy weaknesses, user errors, hardware or firmware flaws, software bugs, and poor security architecture.

4.2.7 Risk

Risk is the possibility of loss or harm arising from performing an activity. It is measured according to its threat level and the number of possible or known vulnerabilities.

4.2.8 Security control

Security controls are countermeasures used to prevent or respond to security threats and to reduce or avoid risk. Details on how to use security countermeasures are typically outlined in the security policy, which contains a set of rules and practices specifying how to implement a system, service, or security plan for maximum protection of sensitive and critical IT resources.

4.2.9 Security policies

A security policy contains a set of security rules and regulations. Security policies will further define how these rules and regulations are implemented and enforced. For example, the positioning and usage of security controls and mechanisms can be determined by security policies.

4.2.10 Threat Agent

A threat agent is an entity that poses a threat because it is capable of carrying out an attack.

4.2.11 Trusted Attacker

A trusted attacker shares IT resources in the same cloud environment as the cloud consumer and attempts to exploit legitimate credentials to target cloud providers and the cloud tenants with whom they share IT resources

4.3 Cloud Security Threats

Cloud security threats can originate either internally or externally, from humans or software programs. Common examples of cloud security threats and vulnerability in a cloud-based environment are discussed below;

4.3.1 Traffic Eavesdropping

This occurs when data being transferred to or within a cloud is passively intercepted by a malicious service agent for illegitimate information gathering purposes. The aim of this attack is to directly compromise the confidentiality of the data; and sometimes the confidentiality of the relationship between the cloud consumer and cloud provider

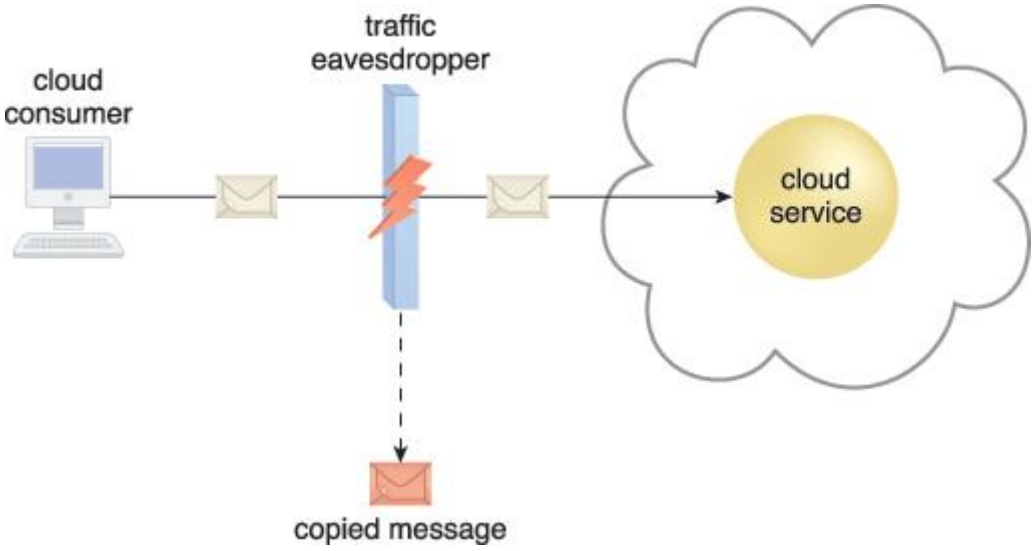


Fig 4.1: Traffic eavesdropping

4.3.2 Denial of Service

This occurs when IT resources are being overloaded to the extent that they cannot function properly.

This form of attack is commonly launched in one of the following ways:

- The workload on cloud services is artificially increased with imitation messages or repeated communication requests.
- The network is overloaded with traffic to reduce its responsiveness and cripple its performance.
- Multiple cloud service requests are sent, each of which is designed to consume excessive memory and processing resources.

Successful DoS attacks produce server degradation and/or failure.

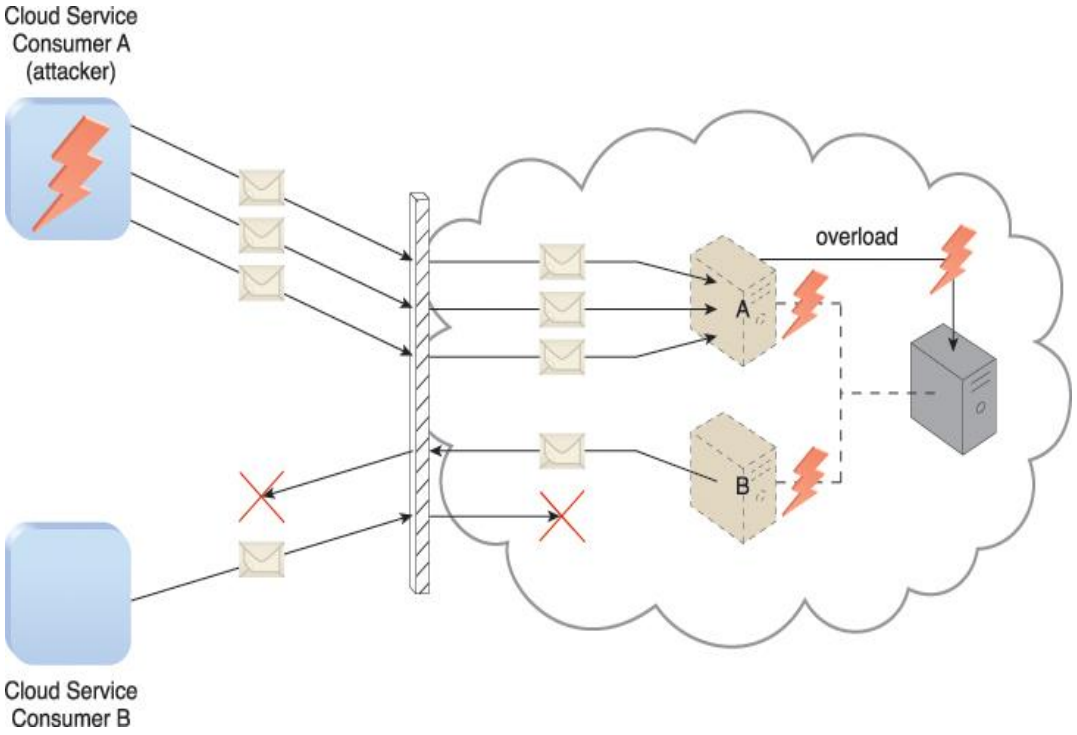


Fig 4.2: Denial of service attack

4.3.3 Virtualization Attack

This Virtualization attack exploits vulnerabilities in the virtualization platform in order to jeopardize its confidentiality, integrity and availability. An attacker successfully accesses a virtual server to compromise its underlying physical server. In the case of a public cloud where single physical IT resource may be providing virtualized IT resources to multiple consumers, such an attack can have huge consequences.

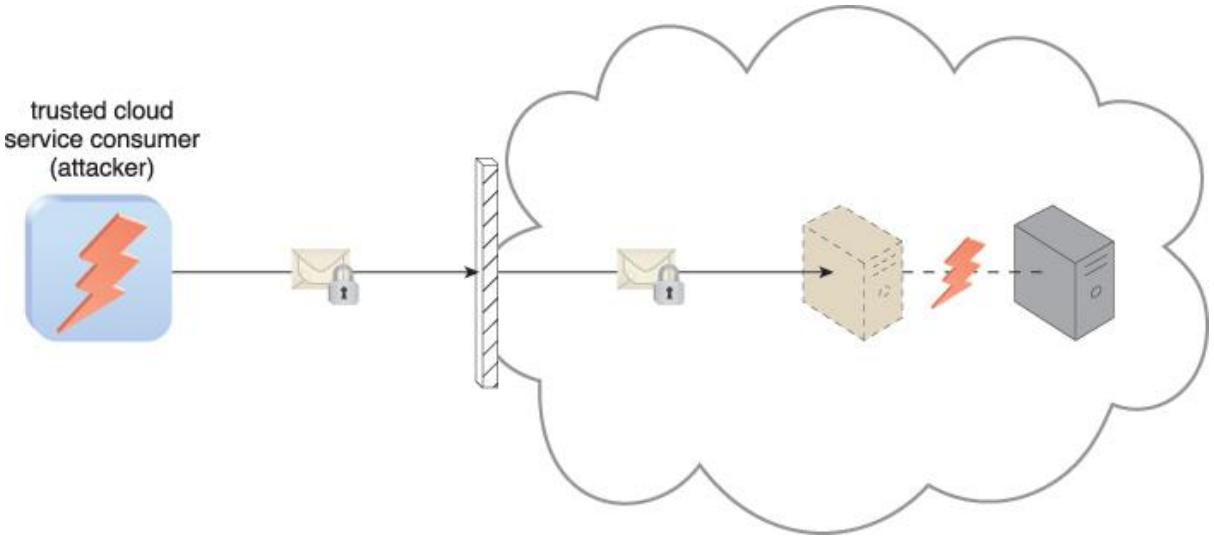


Fig 4.3: Virtualization attack

4.3.4 Malicious Intermediary

This threat arises when messages are intercepted and altered by a malicious service agent thereby potentially compromising the message's confidentiality and sometimes its integrity. It may also insert harmful data into the message before forwarding it to its destination.

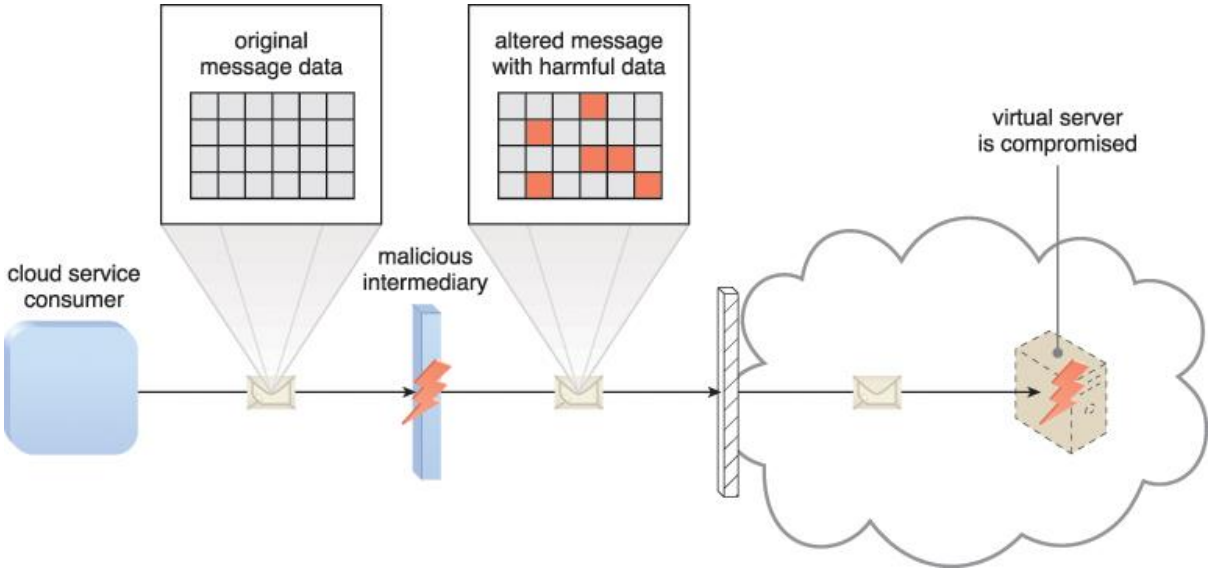


Fig 4.4: Malicious intermediary

4.3.5 Overlapping Trust Boundaries

If physical IT resources within a cloud are shared by different cloud service consumers, these cloud service consumers have overlapping trust boundaries. Malicious cloud service consumers can target shared IT resources with the intention of compromising cloud consumers or other IT resources that share the same cloud boundary. This could result in some or all of the other cloud service consumers being impacted by the attack, and the attacker could use virtual IT resources against others that also share the same trust boundary.

4.3.6 Flawed Implementation

Sub-standard design, configuration and implementation of cloud service deployment can have adverse consequences on the system. Security flaws or operational weakness on the cloud provider's software can be exploited to compromise integrity, confidentiality and available of the IT resources provided.

Some other cloud security threats are;

- Malware Injection
- Insecure Application Programming Interface (API)
- Insider Threats (Malicious insider)
- Advanced persistent threats (APT**)**

4.4 Mitigating Cloud Security Threats

Cloud security threats can be mitigated by adopting the following practices;

- Conduct a cloud security assessment regularly
- Implement cloud security monitoring
- Establish solid access management policies
- Create a disaster recovery plan
- Encryption
- Raise employee awareness

4.5 Chapter Summary

Security is the biggest concern about cloud computing. Fundamental security concepts are confidentiality, integrity, availability, authenticity. Confidential is the characteristics of make data or information accessible to only authorize parties. Integrity is having information unaltered by an unauthorized party. Availability refers to making resources accessible and usable during a specified period of time. Common examples of cloud-based security threats are: Traffic eavesdropping - which occurs when data transmitted to or within a cloud is passively intercepted by a malicious service agent. Denial of service – which occurs when IT resources are being overloaded to the extent that they cannot function properly. Flawed implementation – occurs when sub standard design, configuration and implementation of cloud service can have huge consequence on the system. Ways to mitigate these threats are: Encryption, implementation of cloud security monitoring, creating disaster recovery plan, etc.

4.6 Review Questions

1. **Define the term "Cloud Security"**
2. **Explain the difference between confidentiality, integrity, availability and authenticity?**
3. **Using appropriate example, differentiate between threat and vulnerability?**
4. **Explain a denial of service attack?**
5. **Discuss various security threats in a cloud?**
6. **Identify practices to adopt in order to secure a cloud environment?**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 6,pg *169-186*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran

4.7 MCQs (Quick Quiz)

1. **The ability of a system to make data/information accessible only to authorised parties is called?**

a) Confidentiality
b) Integrity
c) Availability
d) Authenticity

2. **_____is a characteristics of having data unaltered by an unauthorised party?**

a) Authenticity
b) Availability
c) Integrity
d) Confidentiality

3. **_____ refers to a weakness in a system which can be exploited?**

a) Threat
b) Low security
c) Risk
d) Vulnerability

4. **Which of the following is security threat that occurs when data is being transferred within a cloud environment is passively intercepted by a malicious service agent?**

a) Advanced persistent threat
b) Virtualization attack
c) Denial of service
d) Traffic eavesdropping

5. **A form of attack which occurs as a result of IT resource being overloaded to the extent that they cannot function properly is _____**

a) IT resource attack
b) Malware injection
c) Denial of service
d) Virtualization attack

6. **Which of the following practices can not be adopted to mitigate cloud security threats?**

a) Encryption
b) Cloud security monitoring
c) Disaster recovery plan
d) Denial of service

4.8 Case Study

https://www.researchgate.net/publication/276499055_Cloud_Security_Services_Risks_and_a_Case_Study_on_Amazon_Cloud_Services

Recent advances have witnessed the success and popularity of cloud computing which represents a new business model and computing paradigm. The feature of on-demand provisioning of computational storage and bandwidth resources has driven modern businesses into cloud services. The cloud is considered a cutting-edge technology and it is solely relied on by many large technology, business and media companies such as Netflix, or Salesforce.com. However, in addition to the benefits at hand, security issues have been a long-term concern for cloud computing, and are the main barriers of the widespread use of cloud computing.

Question

Using Amazon web services, describe some basic security issues/concerns that are of particular interest to cloud technology, and what do you think are future progression of cloud computing

[redacted]

Chapter 5: Parallel processing, Distributed Computing & Storage Systems in the Cloud

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- **Understand workload distribution architecture**
- **Explain resource pooling architecture**
- **Explain dynamic scalability architecture**
- **Describe elastic resource capacity architecture**
- **Discuss service load balancing architecture**
- **Identify cloud storage systems**
- **Explain distributed computing**
- **Analyse pros and cons of distributed computing**

5.1 Introduction

This chapter describes cloud architectural models. parallel processing, distributed computing

5.2 Workload Distribution Architecture

IT resources can be horizontally scaled via the addition of one or more identical IT resources, and a load balancer that provides runtime logic capable of evenly distributing the workload among the available IT resources. The resulting workload distribution architecture reduces both IT resource over-utilization and under-utilization to an extent dependent upon the sophistication of the load balancing algorithms and runtime logic.

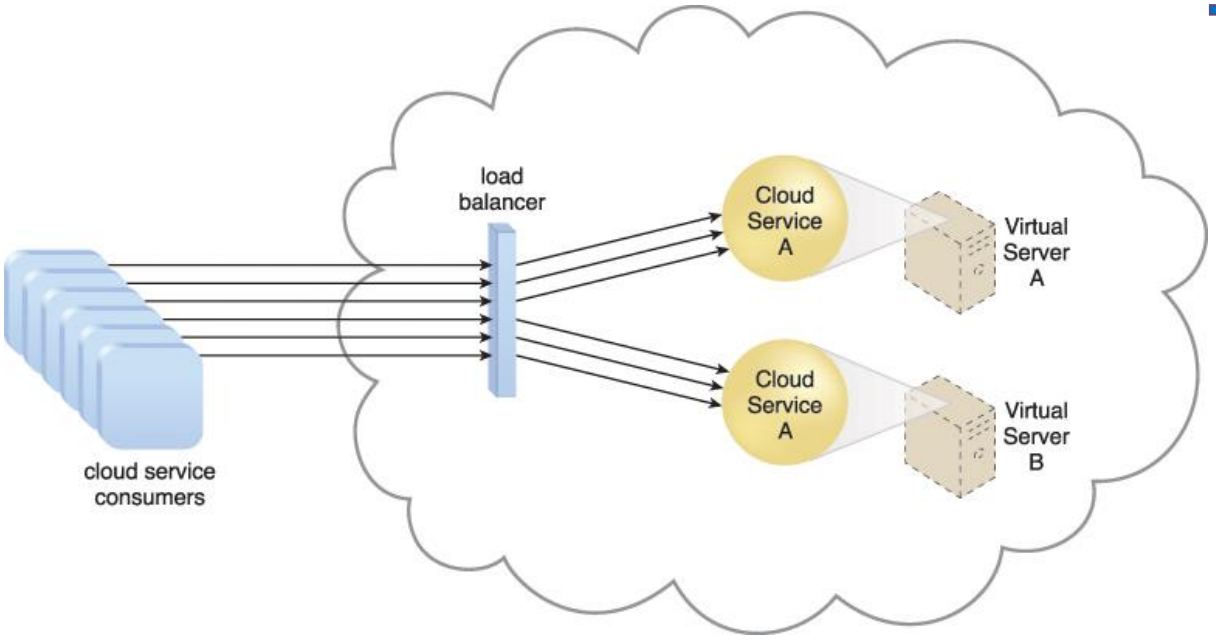


Fig 5.1: Distributed Architecture distribution

As seen above, a redundant copy of Cloud Service A is implemented on Virtual Server B. The load balancer intercepts cloud service consumer requests and directs them to both Virtual Servers A and B to ensure even workload distribution

This fundamental architectural model can be applied to any IT resource, with workload distribution commonly carried out in support of distributed virtual servers, cloud storage devices, and cloud services. Load balancing systems applied to specific IT resources usually produce specialized variations of this architecture that incorporate aspects of load balancing

The following mechanisms can also be part of this cloud architecture:

• Audit Monitor – When distributing runtime workloads, the type and geographical location of the IT resources that process the data can determine whether monitoring is necessary to fulfill legal and regulatory requirements.

• Cloud Usage Monitor – Various monitors can be involved to carry out runtime workload tracking and data processing.

• Hypervisor – Workloads between hypervisors and the virtual servers that they host may require distribution.

• Logical Network Perimeter – The logical network perimeter isolates cloud consumer network boundaries in relation to how and where workloads are distributed.

• Resource Cluster – Clustered IT resources in active/active mode are commonly used to support workload balancing between different cluster nodes.

• Resource Replication – This mechanism can generate new instances of virtualized IT resources in response to runtime workload distribution demands.

5.3 Resource Pooling Architecture

A resource pooling architecture is based on the use of one or more resource pools, in which identical IT resources are grouped and maintained by a system that automatically ensures that they remain synchronized

Common examples of resource pools are

Physical server pool

Physical server pools are composed of networked servers that have been installed with operating systems and other necessary programs and/or applications and are ready for immediate use.

Virtual server pool

Virtual server pools are usually configured using one of several available templates chosen by the cloud consumer during provisioning. For example, a cloud consumer can set up a pool of mid-tier Windows servers with 4 GB of RAM or a pool of low-tier Ubuntu servers with 2 GB of RAM

Storage pool

Storage pools, or cloud storage device pools, consist of file-based or block-based storage structures that contain empty and/or filled cloud storage devices.

Network pool

Network pools (or interconnect pools) are composed of different preconfigured network connectivity devices. For example, a pool of virtual firewall devices or physical network switches can be created for redundant connectivity, load balancing, or link aggregation

CPU pool

CPU pools are ready to be allocated to virtual servers, and are typically broken down into individual processing cores.

Memory pool

Pools of physical RAM can be used in newly provisioned physical servers or to vertically scale physical servers.

Dedicated pools can be created for each type of IT resource and individual pools can be grouped into a larger pool, in which case each individual pool becomes a sub-pool

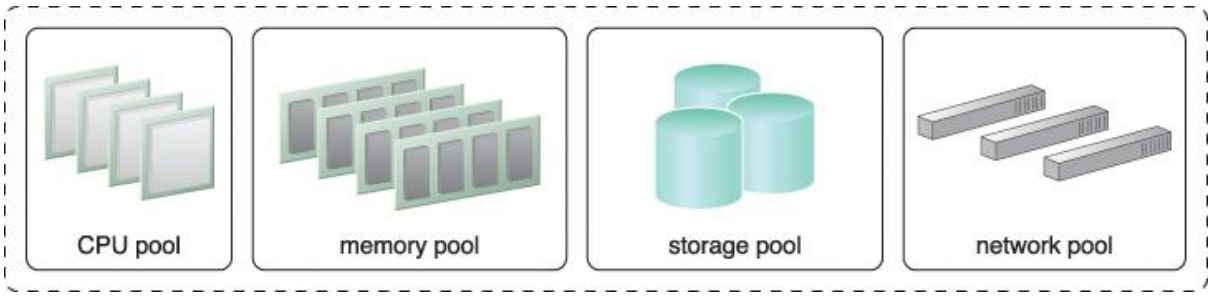


Fig 5.2: A sample resource pool comprising four sub-pools

In addition to cloud storage devices and virtual servers, which are commonly pooled mechanisms, the following mechanisms can also be part of this cloud architecture:

• Audit Monitor – This mechanism monitors resource pool usage to ensure compliance with privacy and regulation requirements, especially when pools contain cloud storage devices or data loaded into memory.

• Cloud Usage Monitor – Various cloud usage monitors are involved in the runtime tracking and synchronization that are required by the pooled IT resources and any underlying management systems.

• Hypervisor – The hypervisor mechanism is responsible for providing virtual servers with access to resource pools, in addition to hosting the virtual servers and sometimes the resource pools themselves.

• Logical Network Perimeter – The logical network perimeter is used to logically organize and isolate resource pools.

• Pay-Per-Use Monitor – The pay-per-use monitor collects usage and billing information on how individual cloud consumers are allocated and use IT resources from various pools.

• Remote Administration System – This mechanism is commonly used to interface with backend systems and programs in order to provide resource pool administration features via a front-end portal.

• Resource Management System – The resource management system mechanism supplies cloud consumers with the tools and permission management options for administering resource pools.

• Resource Replication – This mechanism is used to generate new instances of IT resources for resource pools.

5.4 Dynamic Scalability Architecture

The dynamic scalability architecture is an architectural model based on a system of predefined scaling conditions that trigger the dynamic allocation of IT resources from resource pools. Dynamic allocation enables variable utilization as dictated by usage demand fluctuations, since unnecessary IT resources are efficiently reclaimed without requiring manual interaction

The following types of dynamic scaling are commonly used:

Dynamic Horizontal Scaling – IT resource instances are scaled out and in to handle fluctuating workloads. The automatic scaling listener monitors requests and signals resource replication to initiate IT resource duplication, as per requirements and permissions.

Dynamic Vertical Scaling – IT resource instances are scaled up and down when there is a need to adjust the processing capacity of a single IT resource. For example, a virtual server that is being overloaded can have its memory dynamically increased or it may have a processing core added.

Dynamic Relocation – The IT resource is relocated to a host with more capacity. For example, a database may need to be moved from a tape-based SAN storage device with 4 GB per second I/O capacity to another disk-based SAN storage device with 8 GB per second I/O capacity.

5.5 Elastic Resource Capacity Architecture

The elastic resource capacity architecture is primarily related to the dynamic provisioning of virtual servers, using a system that allocates and reclaims CPUs and RAM in immediate response to the fluctuating processing requirements of hosted IT resources.

5.6 Service Load Balancing Architecture

The service load balancing architecture can be considered a specialized variation of the workload distribution architecture that is geared specifically for scaling cloud service implementations. Redundant deployments of cloud services are created, with a load balancing system added to dynamically distribute workloads.

The duplicate cloud service implementations are organized into a resource pool, while the load balancer is positioned as either an external or built-in component to allow the host servers to balance the workloads themselves. Depending on the anticipated workload and processing capacity of host server environments, multiple instances of each cloud service implementation can be generated as part of a resource pool that responds to fluctuating request volumes more efficiently.

The load balancer can be positioned either independent of the cloud services and their host servers or built-in as part of the application or server's environment. In the latter case, a primary server with the load balancing logic can communicate with neighboring servers to balance the workload.

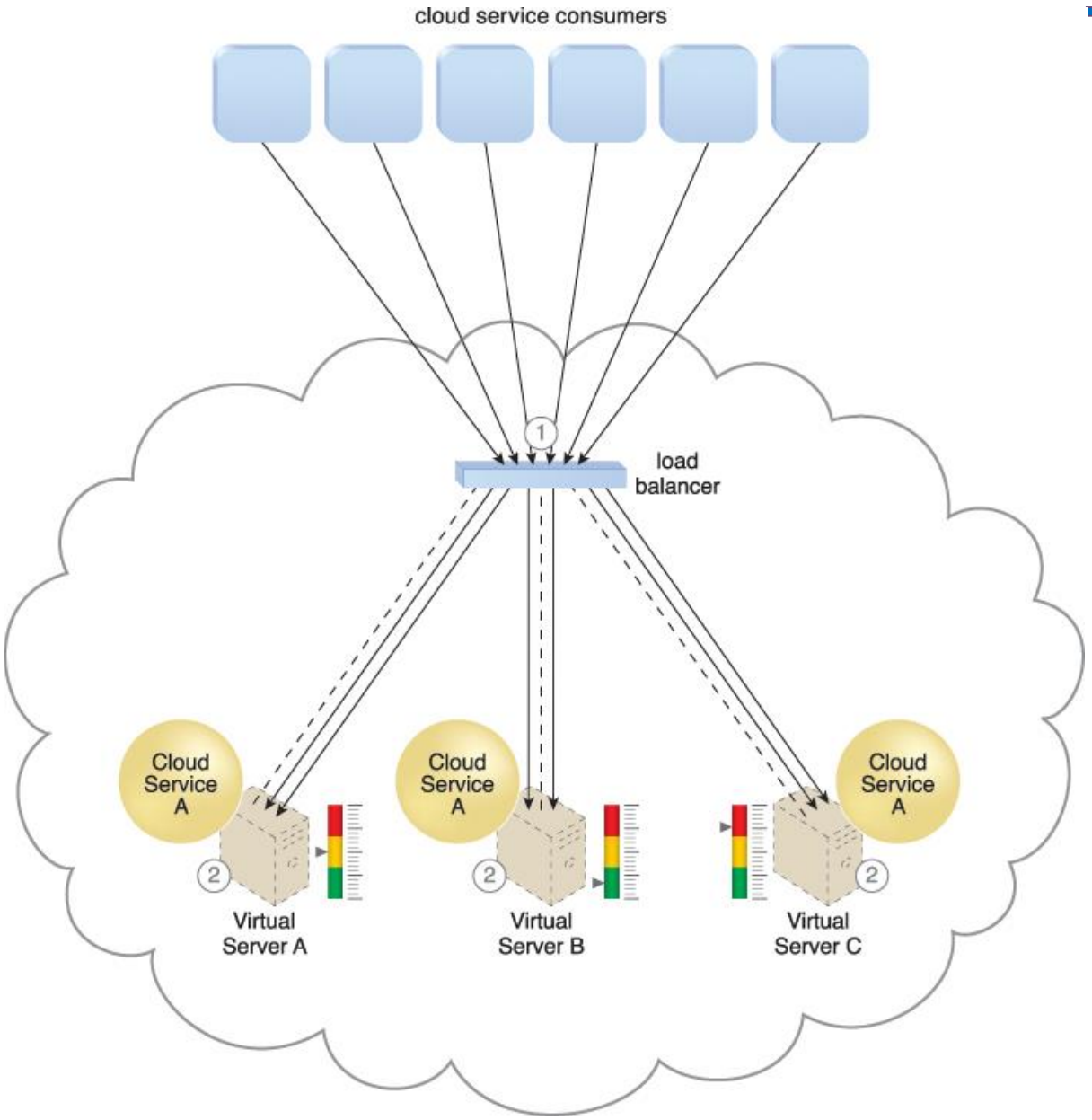


Fig 5.3: Load balancing Architecture

5.7 Cloud Bursting Architecture

The cloud bursting architecture establishes a form of dynamic scaling that scales or "bursts out" on-premise IT resources into a cloud whenever pre-defined capacity thresholds have been reached. The corresponding cloud-based IT resources are redundantly pre-deployed but remain inactive until cloud bursting occurs. After they are no longer required, the cloud-based IT resources are released and the architecture "bursts in" back to the on-premise environment.

Cloud bursting is a flexible scaling architecture that provides cloud consumers with the option of using cloud-based IT resources only to meet higher usage demands. The foundation of this architectural model is based on the automated scaling listener and resource replication mechanisms.

5.8 Cloud Storage Device

The cloud storage device mechanism represents storage devices that are designed specifically for cloud-based provisioning. Instances of these devices can be virtualized. Cloud storage devices are commonly able to provide fixed-increment capacity allocation in support of the pay-per-use mechanism, and they can be exposed for remote access via cloud storage services.

Security, integrity and confidentiality of data are primary concern related to cloud storage, and this becomes more prone to being compromised when entrusted to external cloud providers and other third parties. There can also be legal and regulatory implications that result from relocating data across geographical or national boundaries

5.8.1 Cloud Storage Levels

Cloud storage device mechanisms provide common logical units of data storage, such as:

Files– Collections of data are grouped into files that are located in folders.

Blocks - The lowest level of storage and the closest to the hardware, a block is the smallest unit of data that is still individually accessible.

Datasets – Sets of data are organized into a table-based, delimited, or record format.

Objects – Data and its associated metadata are organized as Web-based resources.

Each of these data storage levels is commonly associated with a certain type of technical interface which corresponds to a particular type of cloud storage device and cloud storage service used to expose its API

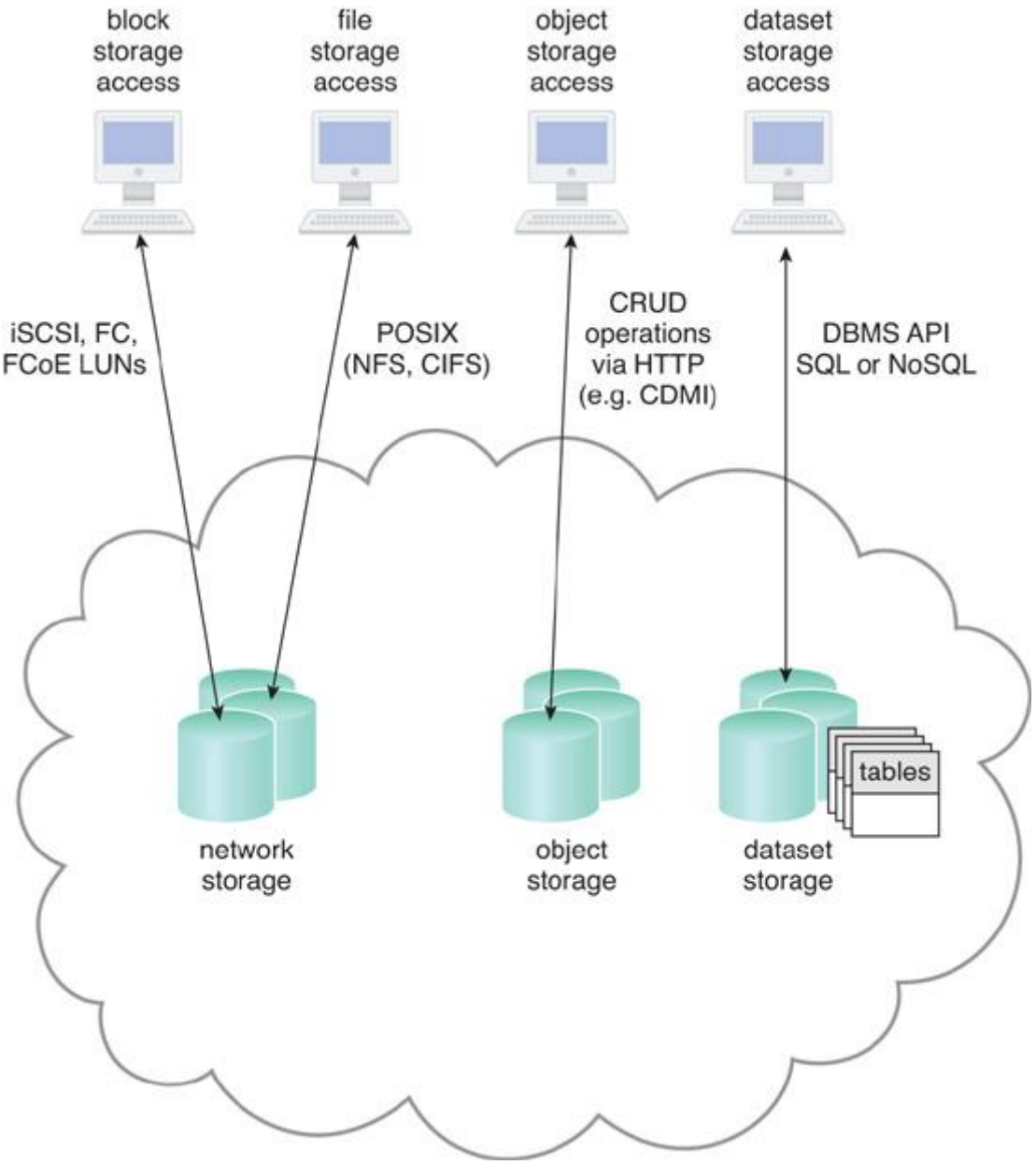


Fig 5.4: Cloud storage

As seen in fig 5.4 above, different cloud service consumers utilize different technologies to interface with virtualized cloud storage devices

Network storage interfaces

It includes storage devices in compliance with industry standard protocols, such as SCSI for storage blocks and the server message block (SMB), common Internet file system (CIFS), and network file system (NFS) for file and network storage. File storage entails storing individual data in separate files that can be different sizes and formats and organized into folders and subfolders. Original files are often replaced by the new files that are created when data has been modified.

When a cloud storage device mechanism is based on this type of interface, its data searching and extraction performance will tend to be sub-optimal.

Object storage interfaces

Various types of data can be referenced and stored as Web resources. This is referred to as object storage, which is based on technologies that can support a range of data and media types. Cloud Storage Device mechanisms that implement this interface can typically be accessed via REST or Web service-based cloud services using HTTP as the prime protocol. The Storage Networking Industry Association's Cloud Data Management Interface (SNIA's CDMI) supports the use of object storage interfaces.

Database storage interfaces

Cloud storage device mechanisms based on database storage interfaces typically support a query language in addition to basic storage operations. Storage management is carried out using a standard API or an administrative user-interface

This classification of storage interface is divided into two main categories according to storage structure. They are Relational and Non-relational Data Storage

- **Relational Data Storage**

Traditionally, many on-premise IT environments store data using relational databases or relational database management systems (RDBMSs). Relational databases (or relational storage devices) rely on tables to organize similar data into rows and columns. Tables can have relationships with each other to give the data increased structure, to protect data integrity, and to avoid data redundancy (which is referred to as data normalization). Working with relational storage commonly involves the use of the industry standard Structured Query Language (SQL). Challenges with cloud-based relational databases commonly pertain to scaling and performance. Scaling a relational cloud storage device vertically can be more complex and cost-ineffective than horizontal scaling. Databases with complex relationships and/or containing large volumes of data can be afflicted with higher processing overhead and latency, especially when accessed remotely via cloud services.

- **Non-relational Data Storage**

Non-relational storage (also commonly referred to as *NoSQL* storage) moves away from the traditional relational database model in that it establishes a "looser" structure for stored data with less emphasis on defining relationships and realizing data normalization. The primary motivation for using non-

relational storage is to avoid the potential complexity and processing overhead that can be imposed by relational databases. Also, non-relational storage can be more horizontally scalable than relational storage. Cloud providers often offer non-relational storage that provides scalability and availability of stored data over multiple server environments

5.9 Distributed Computing

It is a model in which components of a software system are shared among multiple computers to improve efficiency and performance. In distributed computing, multiple computer systems work on a single problem. The ultimate of goal of distributed computing is to maximise performance by connecting users and IT resources in a cost-effective, transparent and reliable manner. It also ensures fault tolerance and enable resource accessibility in the event that one component fails.

A distributed system allows resource-sharing including software by systems connected to their network. Examples of distributed systems/ applications of distributed computing are internet, airline reservation system, electronic banking, telecommunication networks, etc.

5.9.1 Benefits of Distributed Computing

The benefits of distributing computing are discussed below;

Scalability

Distributed systems are inherently scalable. A user can add another machine to handle the increasing workload instead of having to update a single system over and over again.

Fault tolerance and redundancy

Distributed systems are more fault-tolerance than single machines. A business running a cluster of a machine across two data centres means its applications would work even if one data centre goes offline.

Improved efficiency

Distributed systems allow breaking complex problems into smaller pieces and have multiple work on them in parallel which can help cut down on the time needed to solve those problems.

Cost effectiveness

They are much more cost-effective compared to very large centralised systems. A distributed system made up of many mini-computers can be more cost-effective than a mainframe machine.

Reliability – high fault tolerance

A system crash on one server does not affect others.

Other benefits of distributed computing are;

- Flexibility
- Resource sharing
- High performance

5.9.2 Disadvantages of Distributed Computing

Some disadvantages of a distributed computing are listed below;

- Some messages can be lost in the network system
- Security concern
- Overloading
- Higher initial cost
- Network reliance
- Complexity

5.10 Chapter Summary

A resource pooling architecture is based on the use of one or more resource pools, in which identical IT resources are grouped and maintained by a system that automatically ensures that they remain synchronized. Examples of resource pools are physical server pool which are composed of networked servers that have been installed with operating systems. Network pool which are composed of different pre-configured network connectivity devices. Other examples of resource pools are CPU pool, memory pool, etc. Cloud storage is a service that allows to save data on offsite storage systems managed by third-party and is made accessible by a web service API. The primary concern associated with cloud storage are security, integrity and confidentiality. In Distributed computing, multiple computers can host different software components. Advantages of distributed computing are: efficiency, fault-tolerance, scalability, low latency, high performance, etc. Some disadvantages of distributed computing are: security concern, higher initial cost, etc

5.11 Review Questions

1. **Discuss resource pooling architecture, and briefly describe examples of resource pool**
2. **Explain the mechanism of dynamic scalability architecture**
3. **Differentiate between dynamic horizontal scaling and dynamic vertical scaling**
4. **Discuss cloud storage interfaces**
5. **Explain the difference between relational and non-relational storage system**
6. **Describe the concept of distributed computing**
7. **List the benefits and disadvantages of distributed computing**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 11,pg *291-314*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran

5.12 MCQs (Quick Quiz)

1 Which of the following is not an example of resource pool?

a) Virtual server pool
b) Memory pool
c) Storage pool
d) IP pool

2. In _____ scaling, IT resource instances are scaled up and down when there is a need to adjust the processing capacity of a single IT resource?

a) Dynamic vertical
b) Dynamic horizontal
c) Dynamic relocation
d) Memory

3. A database which rely on tables to organise similar data into rows and columns is called?

a) Relational
b) Non-relational
c) Dynamic
d) Non-dynamic

4. A type of scaling where IT resources are scaled out and in to handle fluctuation workloads is called?

a) Dynamic horizontal scaling
b) Dynamic vertical scaling
c) Upside scaling
d) Inward scaling

5. A type of resource pool which is usually configured using one of several available template chosen by the cloud consumer during provisioning is?

a) Network pool
b) Storage pool
c) Physical server pool
d) Virtual server pool

Chapter 6: Web 2.0

LEARNING OUTCOMES

After reading this Section of the guide, the learner should be able to:

- **Understand Web 2.0 Technology**
- **Explain basic concepts in web 2.0 technology**
- **Discuss Hypertext transfer protocol**
- **Explain URL, HTML and XML**

6.1 Introduction

This section describes basic web technologies and its relationship to cloud services. As a result of cloud computing's fundamental reliance on internetworking, web browser, the ease of web-based service development, web technology is generally used as both the implementation medium and the management interface for cloud services.

6.2 Basic Web Technology

The world wide web (www) is simply a system of interlinked resources that are accessed over the internet. The two basic components of the web are: Web browser client and the Web server. Other components such as proxies, caching services, gateways and load balancers are used to improve scalability and security of web application.

The technology architecture comprises three fundamental elements. These are;

Uniform Resource Locator

Uniform resource locator (URL) is a standard syntax used for creating identifiers that point to web-based resources. It is often structured using a logical network location.

Hypertext Transfer Protocol

HTTP is the primary communication protocol used to exchange content and data throughout the world wide web. The Hypertext Transfer Protocol (HTTP) is an application-level protocol for distributed, collaborative, hypermedia information systems. This is the foundation for data communication for the World Wide Web (i.e. internet) since 1990. HTTP is a generic and stateless protocol which can be used for other purposes as well using extensions of its request methods, error codes, and header. URLs are transmitted via HTTP.

There are three basic features that make HTTP a simple but powerful protocol:

- **HTTP is connectionless:** The HTTP client, i.e., a browser initiates an HTTP request and after a request is made, the client waits for the response. The server processes the request and sends a response back after which client disconnect the connection. So client and server knows about each other during current request and response only. Further requests are made on new connection like client and server are new to each other.
- **HTTP is media independent:** It means, any type of data can be sent by HTTP as long as both the client and the server know how to handle the data content. It is required for the client as well as the server to specify the content type using appropriate MIME-type.
- **HTTP is stateless:** As mentioned above, HTTP is connectionless and it is a direct result of HTTP being a stateless protocol. The server and client are aware of each other only during a current request. Afterwards, both of them forget about each other. Due to this nature of the protocol, neither the client nor the browser can retain information between different requests across the web pages.

The HTTP protocol is a request/response protocol based on the client/server based architecture where web browsers, robots and search engines, etc. act like HTTP clients, and the Web server acts as a server

Mark-up Languages

Mark-up languages (HTML, XML) provide a lightweight means of expressing web-centric data and meta. The two primary mark-up languages are HTML (used for expressing the presentation of web pages); and XML (which allows for the definition of vocabularies used to associate meaning to web-based data via metadata)

6.3 Web Applications

A distributed application that uses Web-based technologies is typically considered a Web application. These applications can be found in all kinds of cloud-based environments due to their high accessibility.

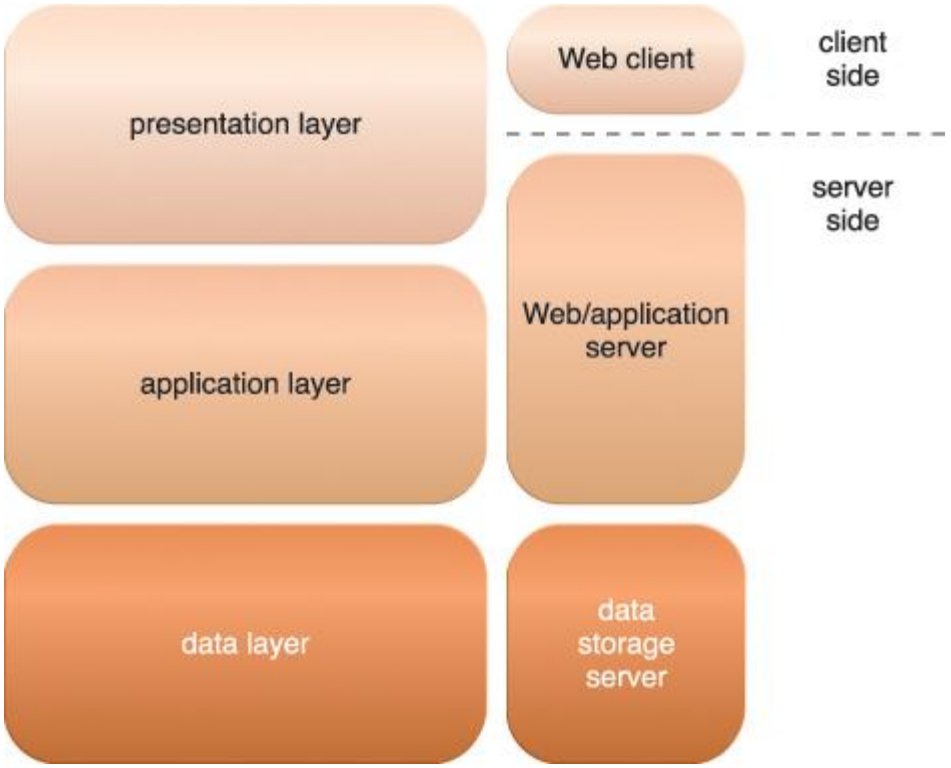


Fig 6.1: The three basic architectural tiers of Web applications

Fig 6.1 presents a common architectural abstraction for Web applications that is based on the basic three-tier model. The first tier is called the presentation layer, which represents the user-interface.

The middle tier is the application layer that implements application logic, while the third tier is the *data* layer that is comprised of persistent data stores.

The presentation layer has components on both the client and server-side. Web servers receive client requests and retrieve requested resources directly as static web content and indirectly as dynamic web content, which is generated according to the application logic. Web servers interact with application servers in order to execute the requested application logic, which then typically involves interaction with one or more underlying databases

6.4 Chapter Summary

The world wide web is a system of interconnected resources which are accessed over the internet. The two basic components of the web are web browser client and the web server. Uniform resource locator (URL) is a standard syntax used for creating identifiers that point to web-based resources. It is often structured using a logical network location. HTTP is the primary communication protocol used to exchange content and data throughout the world wide web. URLs are transmitted via HTTP. Mark-up languages (HTML, XML) provide a lightweight means of expressing web-centric data and meta. The two primary mark-up languages are HTML which is used for expressing the presentation of web pages; and XML which allows for the definition of vocabularies used to associate meaning to web-based data via metadata.

6.5 Review Questions

1. **What is the difference between Web browser client and a web server?**
2. **Briefly describe the function of a URL?**
3. **Explain the mechanism that goes on between the client's side and server side when a client request for a webpage?**
4. **Differentiate between HTML and XML**

Read

Cloud Computing: Concepts, Technology and Architecture, 9th edition, 2015 Zaigham Mahmood, Ricardo Puttini, Thomas Erl, Chapter 5 ,pg *154-157*

Essentials of Cloud Computing, 6th Edition, 2015, K. Chandrasekaran

6.6 MCQs (Quick Quiz)

1. What is the full meaning of WWW

a) World wide wizard
b) World wide world
c) Wide world web
d) World wide web

2. _____ is a standard syntax used for creating identifiers that point to web-based resources?

a) URL
b) FTP
c) HTTP
d) XML

3. What is the full meaning of URL?

a) Universal resource locator
b) Universe resource locator
c) Uniform resource locator
d) Universal resources locator

4. Which one of the following is not a feature that make HTTP a simple but powerful protocol?

a) HTTP is connectionless
b) HTTP is media independent
c) HTTP is stateless
d) HTTP is connection oriented

5. Which of the following is used in the presentation of a web-based?

a) FTTP
b) XML
c) HTML
d) HTTP