MURANG'A UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF IT

SIT407: CLOUD COMPUTING
CREDIT HOURS: 3 HOURS

COURSE NOTES – 2021/2022

INTRODUCTION TO CLOUD COMPUTING

Objectives

i) To identify and explain cloud computing Properties and Characteristics
ii) Explain cloud Service models,
iii) Describe cloud deployment models

CLOUD COMPUTING CHARACTERISTICS

i) On-demand self-service

Cloud computing resources can be provisioned without human interaction from the service provider. In other words, a manufacturing organization can provision additional computing resources as needed without going through the cloud service provider. This can be a storage space, virtual machine instances, database instances, and so on.

Manufacturing organizations can use a web self-service portal as an interface to access their cloud accounts to see their cloud services, their usage, and also to provision and de-provision services as they need to.

CLOUD COMPUTING CHARACTERISTICS CONT'D

2. Broad network access

Cloud computing resources are available over the network and can be accessed by diverse customer platforms. In other words, cloud services are available over a network—ideally high broadband communication link—such as the internet, or in the case of a private clouds it could be a local area network (LAN).

Network bandwidth and latency are very important aspects of cloud computing and broad network access, because they relate to the quality of service (QoS) on the network. This is particularly important for serving time sensitive manufacturing applications

CLOUD COMPUTING CHARACTERISTICS CONT'D

3. Multi-tenancy and resource pooling

Cloud computing resources are designed to support a multi-tenant model. Multi-tenancy allows multiple customers to share the same applications or the same physical infrastructure while retaining privacy and security over their information.

It's similar to people living in an apartment building, sharing the same building infrastructure but they still have their own apartments and privacy within that infrastructure. That is how cloud multi-tenancy works.

CLOUD COMPUTING CHARACTERISTICS CONT'D

Resource pooling means that multiple customers are serviced from the same physical resources.

Providers' resource pool should be very large and flexible enough to service multiple client requirements and to provide for economy of scale.

When it comes to resource pooling, resource allocation must not impact performances of critical manufacturing applications.

CLOUD COMPUTING CHARACTERISTICS CONT'D

4. Rapid elasticity and scalability

One of the great things about cloud computing is the ability to quickly provision resources in the cloud as manufacturing organizations need them.

And then to remove them when they don't need them. Cloud computing resources can scale up or down rapidly and, in some cases, automatically, in response to business demands.

It is a key feature of cloud computing. The usage, capacity, and therefore cost, can be scaled up or down with no additional contract or penalties.

CLOUD COMPUTING CHARACTERISTICS CONT'D

With cloud computing scalability, there is less capital expenditure on the cloud customer side.

This is because as the cloud customer needs additional computing resources, they can simply provision them as needed, and they are available right away. Scalability is more planned and gradual.

For instance, scalability means that manufacturing organizations are gradually planning for more capacity and of course the cloud can handle that scaling up or scaling down.

CLOUD COMPUTING CHARACTERISTICS CONT'D

Just-in-time (JIT) service is the notion of requiring cloud elasticity either to provision more resources in the cloud or less. For example, if a manufacturing organization all of a sudden needs more computing power to perform some kind of complex calculation, this would be cloud elasticity that would be a just-in-time service. On the other hand, if the manufacturing organization needs to provision human-machine interface (HMI) tags in the database for a manufacturing project, that is not really just-in-time service, it is planned ahead of time. So it is more on the scalability side than elasticity.

CLOUD COMPUTING CHARACTERISTICS CONT'D

Another feature available for rapid elasticity and scalability in the cloud is related to testing of manufacturing applications. If a manufacturing organization needs, for example, a few virtual machines to test a supervisory control and data acquisition (SCADA) system before they roll it out in production, they can have it up and running in minutes instead of physically ordering and waiting for hardware to be shipped.

CLOUD COMPUTING CHARACTERISTICS CONT'D

In terms of the bottom line, when manufacturing organizations need to test something in the cloud, they are paying for what they use as they use it. As long as they remember to de-provision it, they will no longer be paying for it.

There is no capital expense here for computer resources.

Manufacturing organizations are using the cloud provider's investment in cloud computing resources instead. This is really useful for testing smart manufacturing solutions.

CLOUD COMPUTING CHARACTERISTICS CONT'D

5. Measured service

Cloud computing resources usage is metered and manufacturing organizations pay accordingly for what they have used. Resource utilization can be optimized by leveraging charge-per-use capabilities. This means that cloud resource usage—whether virtual server instances that are running or storage in the cloud—gets monitored, measured and reported by the cloud service provider. The cost model is based on "pay for what you use"—the payment is variable based on the actual consumption by the manufacturing organization.

SUMMARY OF CHARACTERISTICS

On-demand self-services:

The Cloud computing services does not require any human administrators, user themselves are able to provision, monitor and manage computing resources as needed.

Broad network access:

The Computing services are generally provided over standard networks and heterogeneous devices.

Rapid elasticity:

The Computing services should have IT resources that are able to scale out and in quickly and on as needed basis. Whenever the user require services it is provided to him and it is scale out as soon as its requirement gets over.

SUMMARY OF CHARACTERISTICS CONT'D

Resource pooling:

The IT resource (e.g., networks, servers, storage, applications, and services) present are shared across multiple applications and occupant in an uncommitted manner. Multiple clients are provided service from a same physical resource.

Measured service:

The resource utilization is tracked for each application and occupant, it will provide both the user and the resource provider with an account of what has been used. This is done for various reasons like monitoring billing and effective use of resource.

CLOUD SERVICE MODELS

Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely:

i) Infrastructure as a Service

ii) Platform as a Service and

iii) Software as a Service.

These abstraction levels can also be viewed as a layered architecture where services of a higher layer can be composed from services of the underlying layer.

SERVICE MODELS CONT'D

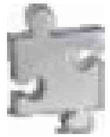
Service	Main Access &	Service content
 SaaS	Web Browser	Social networks, Office suites, CRM,
 PaaS	Cloud Development Environment	**Cloud Platform** Programming languages, Frameworks, Mashups editors, Structured data
 IaaS	Virtual Infrastructure	Compute Servers, Data Storage, 17

Figure 1: The cloud computing stack

SERVICE MODELS CONT'D

Figure 1 depicts the layered organization of the cloud stack from physical infrastructure to applications

The reference model explains the role of each layer in an integrated architecture.

A core middleware manages physical resources and the VMs deployed on top of them; in addition, it provides the required features (e.g., accounting and billing) to offer multi-tenant pay-as-you-go services

SERVICE MODELS CONT'D

Cloud development environments are built on top of infrastructure services to offer application development and deployment capabilities; in this level, various programming models, libraries, APIs, and mashup editors enable the creation of a range of business, Web, and scientific applications.

Once deployed in the cloud, these applications can be consumed by end users.

SERVICE MODELS CONT'D

i) Infrastructure as a Service

Offering virtualized resources (computation, storage, and communication) on demand is known as Infrastructure as a Service (IaaS).

A cloud infrastructure enables on-demand provisioning of servers running several choices of operating systems and a customized software stack. Infrastructure services are considered to be the bottom layer of cloud computing systems

SERVICE MODELS CONT'D

Amazon Web Services mainly offers IaaS, which in the case of its EC2 service means offering VMs with a software stack that can be customized similar to how an ordinary physical server would be customized.

Users are given privileges to perform numerous activities to the server, such as: starting and stopping it, customizing it by installing software packages, attaching virtual disks to it, and configuring access permissions and firewalls rules.

SERVICE MODELS CONT'D

ii) Platform as a Service

In addition to infrastructure-oriented clouds that provide raw computing and storage services, another approach is to offer a higher level of abstraction to make a cloud easily programmable, known as Platform as a Service (PaaS).

A cloud platform offers an environment on which developers create and deploy applications and do not necessarily need to know how many processors or how much memory that applications will be using. In addition, multiple programming models and specialized services (e.g., data access, authentication, and payments) are offered as building blocks to new applications

SERVICE MODELS CONT'D

Google App Engine (GAE), an example of Platform as a Service, offers a scalable environment for developing and hosting Web applications, which should be written in specific programming languages such as Python or Java, and use the services 'own proprietary structured object data store.

Building blocks include an in-memory object cache (memcache), mail service, instant messaging service (XMPP), an image manipulation service, and integration with Google Accounts authentication service.

SERVICE MODELS CONT'D

iii) Software as a Service

Applications reside on the top of the cloud stack. Services provided by this layer can be accessed by end users through Web portals.

Therefore, consumers are increasingly shifting from locally installed computer programs to on-line software services that offer the same functionally.

SERVICE MODELS CONT'D

Traditional desktop applications such as word processing and spreadsheet can now be accessed as a service in the Web.

This model of delivering applications, known as Software as a Service (SaaS), alleviates the burden of software maintenance for customers and simplifies development and testing for providers.

Salesforce.com, which relies on the SaaS model, offers business productivity applications (CRM) that reside completely on their servers, allowing customers to customize and access applications on demand.

CLOUD DEPLOYMENT MODELS

Although cloud computing has emerged mainly from the appearance of public computing utilities, other deployment models, with variations in physical location and distribution, have been adopted. In this sense, regardless of its service class, a cloud can be classified as public, private, community, or hybrid based on model of deployment as shown in Figure 2

CLOUD DEPLOYMENT MODELS CONT'D

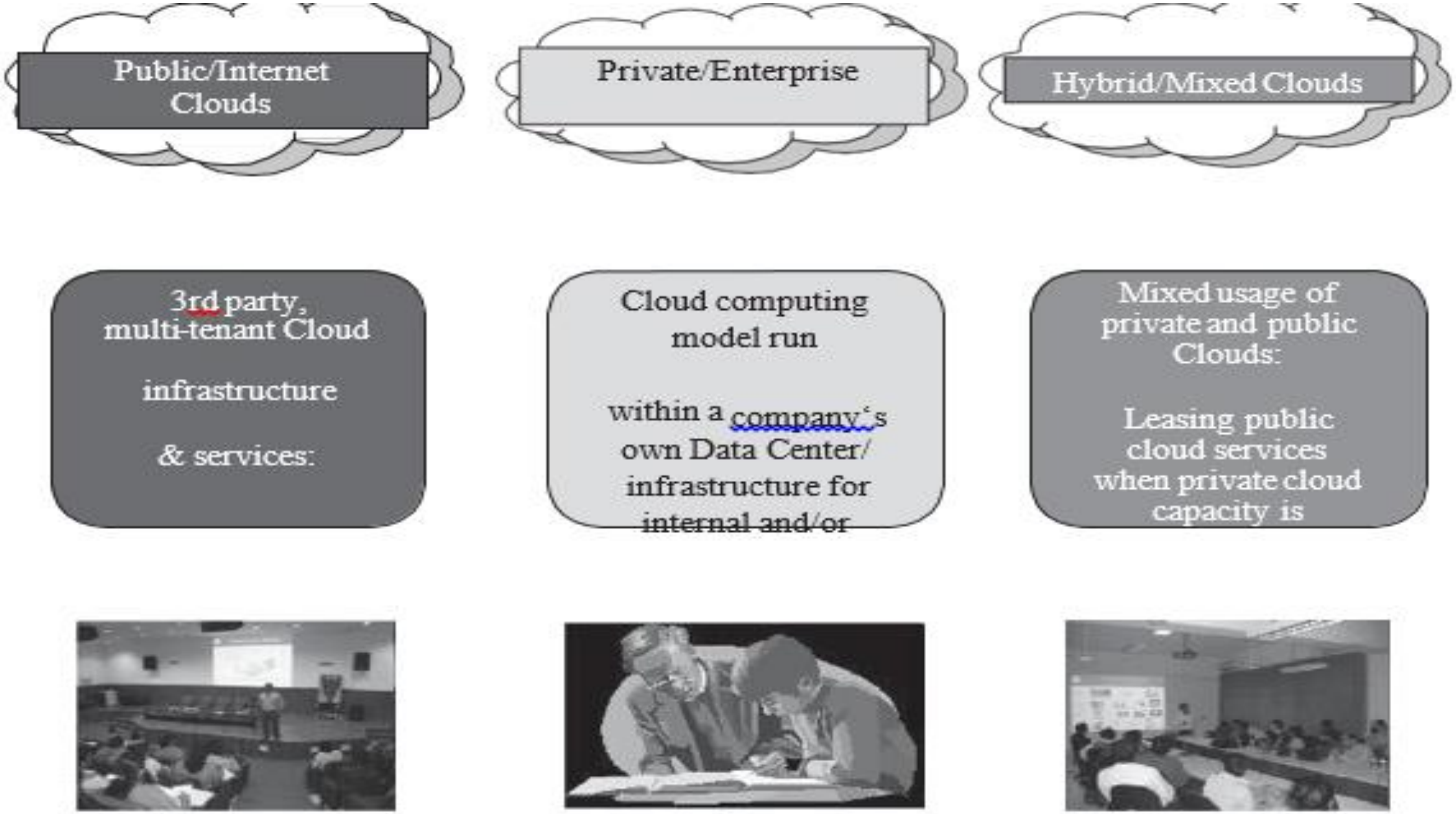


Figure 2:Types of clouds based on deployment models

CLOUD DEPLOYMENT MODELS CONT'D

i) Public cloud

Public cloud is a cloud made available in a pay-as-you-go manner to the general publicll and private cloud as internal data center of a business or other organization, not made available to the general public.

ii) Private Cloud

Establishing a private cloud means restructuring an existing infrastructure by adding virtualization and cloud-like interfaces. This allows users to interact with the local data center while experiencing the same advantages of public clouds, most notably self-service interface, privileged access to virtual servers, and per-usage metering and billing.

CLOUD DEPLOYMENT MODELS CONT'D

iii) Community cloud

A community cloud is —shared by several organizations and a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).

iv) Hybrid cloud

A hybrid cloud takes shape when a private cloud is supplemented with computing capacity from public clouds. The approach of temporarily renting capacity to handle spikes in load is known as cloud-bursting.

FEATURES OF A CLOUD

a) SELF-SERVICE

Consumers of cloud computing services expect on-demand, nearly instant access to resources. To support this expectation, clouds must allow self-service access so that customers can request, customize, pay, and use services without intervention of human operators.

b) PER-USAGE METERING AND BILLING

Cloud computing eliminates up-front commitment by users, allowing them to request and use only the necessary amount.

FEATURES OF A CLOUD CONT'D

Services must be priced on a short-term basis (e.g., by the hour), allowing users to release (and not pay for) resources as soon as they are not needed. For these reasons, clouds must implement features to allow efficient trading of service such as pricing, accounting, and billing.

Metering should be done accordingly for different types of service (e.g., storage, processing, and bandwidth) and usage promptly reported, thus providing greater transparency.

FEATURES OF A CLOUD CONT'D

c) ELASTICITY

Cloud computing gives the illusion of infinite computing resources available on demand. Therefore, users expect clouds to rapidly provide resources in any Quantity at any time. In particular, it is expected that the additional resources can be

i) Provisioned, possibly automatically, when an application load increases and

ii) Released when load decreases (scale up and down).

d) CUSTOMIZATION

In a multi-tenant cloud a great disparity between user needs is often the case. Thus, resources rented from the cloud must be highly customizable. In the case of infrastructure services, customization means allowing users to deploy specialized virtual appliances and to be given privileged (root) access to the virtual servers.

CLOUD INFRASTRUCTURE MANAGEMENT

A key challenge IaaS providers face when building a cloud infrastructure is managing physical and virtual resources, namely servers, storage, and net- works.

The software toolkit responsible for this orchestration is called a virtual infrastructure manager (VIM). This type of software resembles a traditional operating system but instead of dealing with a single computer, it aggregates resources from multiple computers, presenting a uniform view to user and applications. The term cloud operating system is also used to refer to it.

CLOUD INFRASTRUCTURE MANAGEMENT CONT'D

The availability of a remote cloud-like interface and the ability of managing many users and their permissions are the primary features that would distinguish cloud toolkits from VIMs.

Virtually all VIMs we investigated present a set of basic features related to managing the life cycle of VMs, including networking groups of VMs together and setting up virtual disks for VMs.

The end.

Q & A