Cloud computing is a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to the external customers using the Internet technologies.

A **cloud** refers to a distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources.

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

Business Drivers

Capacity planning is the process of determining and fulfilling future demands of an organization's IT resources, products, and services.

- Lead Strategy – adding capacity to an IT resource in anticipation of demand
- Lag Strategy – adding capacity when the IT resource reaches its full capacity
- Match Strategy – adding IT resource capacity in small increments, as demand increases

Cost Reduction

Common forms of infrastructure-related operating overhead include the following:

- technical personnel required to keep the environment operational
- upgrades and patches that introduce additional testing and deployment cycles
- utility bills and capital expense investments for power and cooling
- security and access control measures that need to be maintained and enforced to protect infrastructure resources
- administrative and accounts staff that may be required to keep track of licenses and support arrangements

Organizational Agility is the measure of an organization's responsivness to change.

Technology Innovations: Existing technologies ar often used as inspiration.

- Clustering -> Independent IT resources that are interconnected and work as a single system (one failed component is replaced by another)
- Grid Computing -> A computing grid (or "computational grid") provides a platform in which computing resources are organized into one or more logical pools.
- Virtualization -> Virtualization represents a technology platform used for the creation of virtual instances of IT resources.

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.**

The party that provides cloud-based IT resources is the **cloud provider**.

The party that uses cloud-based IT resources is the **cloud consumer**.

Scaling is the ability of the IT resource to handle increased or decreased usage demands.

Horizontal Scaling – Scaling out and scaling in (The allocating or releasing of IT resources that are of the same type)

Vertical Scaling – Scaling up and scaling down (When an existing IT resource is replaced by another with higher or lower capacity)

Cloud service is any IT resource that is made remotely accessible via a cloud.

The **cloud service consumer** is a temporary runtime role assumed by a software program when it accesses a cloud service.

Goals and Benefits

Reduced Investment and Proportional Costs -> Opens the door for organizations to gain access to powerful infrastructure without having to purchase it themselves.

Increased Scalability -> Clouds can instantly and dynamically allocate IT resources to cloud consumers, on-demand or via the cloud consumer's direct configuration

Increased Availability and Reliability -> Accessible for longer periods of time and able to better avoid and recover from exception conditions.

Risks and Challenges

- Increased Security Vulnerabilities
- Reduced Operational Governance Control
- Limited Portability Between Cloud Providers
- Multi-Regional Compliance and Legal Issues

Chapter 4 – Fundamental Concepts and Models

Roles and Boundaries

Cloud Provider - Responsible for making cloud services available to cloud consumers.

Cloud Customer – Organization that has a formal contract or arrangement with a cloud provider to use IT resources made available by the cloud provider.

Cloud Service Owner – Person or organization that legally owns a cloud service.

Cloud Resource Administrator - Person or organization responsible for administering a cloud-based IT resource (including cloud services).

Additional Roles:

- *Cloud Auditor* - A third-party (often accredited) that conducts independent assessments of cloud environments
- *Cloud Broker* - Responsibility of managing and negotiating the usage of cloud services between cloud consumers and cloud providers.
- *Cloud Carrier* - Party responsible for providing the wire-level connectivity between cloud consumers and cloud providers.

An **organizational boundary** represents the physical perimeter that surrounds a set of IT resources that are owned and governed by an organization.

A **trust boundary** is a logical perimeter that typically spans beyond physical boundaries to represent the extent to which IT resources are trusted.

Cloud Characteristics

On-demand usage - > Once configured, usage of the self-provisioned IT resources can be automated, requiring no further human involvement by the cloud consumer or cloud provider.

Ubiquitous Access -> Ability for a cloud service to be widely accessible. (different devices, transfer protocols, interfacers, security technologies)

Multitenacity -> The characteristic of a software program that enables an instance of the program to serve different consumers (tenants) whereby each is isolated from the other. IT resources can be dynamically assigned and reassigned, according to cloud service consumer demands.

Elasticity -> Ability of a cloud to transparently scale IT resources, as required in response to runtime conditions or as pre-determined by the cloud consumer or cloud provider.

Measured Usage -> Ability of a cloud platform to keep track of the usage of its IT resources, primarily by cloud consumers. (billing purposes and monitoring)

Resiliency ->Distributes redundant implementations of IT resources across physical locations

Cloud Delivery Models

Represents a specific, pre-packaged combination of IT resources offered by a cloud provider.

Three common cloud delivery models:

Infrastructure-as-a-Service: 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. The environment can include hardware, network, connectivity, operating systems, and other raw IT resources. IT resources are typically virtualized and packaged into bundles that simplify up-front runtime scaling and customization of the infrastructure.

Platform-as-a-Service: Represents a pre-defined "ready-to-use" environment typically comprised of already deployed and configured IT resources. Reasons:

- Teh customer wants to extend on-premise environments into the cloud for scalability and economic purposes.
- The cloud consumer uses the ready-made environment to entirely substitute an on-premise environment.
- The cloud consumer wants to become a cloud provider and deploys its own cloud services to be made available to other external cloud consumers.

Software-as-a-Service: Positioned as a shared cloud service and made available as a "product" or generic utility. The cloud service consumer is given access the cloud service contract, but not to any underlying IT resources or implementation details.

Difference between services:

Cloud Delivery Model	Typical Level of Control Granted to Cloud Consumer	Typical Functionality Made Available to Cloud Consumer
SaaS	usage and usage-related configuration	access to front-end user-interface
PaaS	limited administrative	moderate level of administrative control over IT resources relevant to cloud consumer's usage of platform
IaaS	full administrative	full access to virtualized infra-structure-related IT resources and, possibly, to underlying physical IT resources

Cloud Delivery Model	Common Cloud Consumer Activities	Common Cloud Provider Activities
SaaS	uses and configures cloud service	implements, manages, and maintains cloud service monitors usage by cloud consumers
PaaS	develops, tests, deploys, and manages cloud services and cloud-based solutions	pre-configures platform and provisions underlying infrastructure, middleware, and other needed IT resources, as necessary monitors usage by cloud consumers
IaaS	sets up and configures bare infrastructure, and installs, manages, and monitors any needed software	provisions and manages the physical processing, storage, networking, and hosting required monitors usage by cloud consumers

Combining Cloud Delivery Models

IaaS + PaaS -> A PaaS environment will be built upon an underlying infrastructure comparable to the physical and virtual servers and other IT resources provided in an IaaS environment.

IaaS + PaaS + SaaS -> Different combinations of IaaS, PaaS, and SaaS are possible, depending on how cloud consumers and cloud providers choose to leverage the natural hierarchy established by these base cloud delivery models

Cloud Deployment Models

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

4 common cloud deployment models(+ Virtual private cloud and Inter-cloud)**:**

Public Clouds -> Publicly accessible cloud environment owned by a third-party cloud provider., provisioned via cloud delivery models. Offered to cloud customers at a cost or are commercialized via other avenues (advertisement).

Comunity Clouds -> Access is limited to specific community. Community share the responsibility for defining and envolving the community cloud.

Private Clouds -> Owned by a single organization. Private cloud can change how organizational and trust boundaries are defined and applied. Private clouds enable an organization to use cloud computing technology as a means of centralizing access to IT resources by different parts, locations, or departments of the organization.

Hybrid Clouds -> Cloud environment comprised of two or more different cloud deployment models.

Chapter 5. Cloud-Enabling Technology

Broadband Networks and Internet Architecture

Internet Service Providers provide national-wide or world-wide internet access service. They freely deploy, operate and manage their own network.

There are **three** tiers of world-wide connectivity:

- **Tier 1 (large-scale international ISPs)**
- **Tier 2 (large regional ISPs)**
- **Tier 3 (local ISPs)**

Two fundamental component of internetworking architecture:

- *Connectionless packet switching* -> End-to-end (sender-receiver pair) data flows are divided into packets of a limited size that are received and processed through network switches and routers, then queued and forwarded from one intermediary node to the next (IP and MAC address).
- *Router-based interconnectivity* -> A router is a device that is connected to multiple networks through which it forwards packets. Process and forward packet individually. Possibly multiple ISP networks between a cloud customer and its cloud provider.

Several internetworking layers are implemented:

Physical Network -> IP packets are transmitted through underlying physical networks that connect adjacent nodes, such as Ethernet, ATM network, and the 3G mobile HSDPA. Physical networks comprise a data link layer that controls data transfer between neighboring nodes, and a physical layer that transmits data bits through both wired and wireless media.

Transport Layer Protocol -> Use IP to provide standardized, end-to-end communication support that facilitates the navigation of data packets across the Internet (TCP-Transmission Control Protocol, UDP-User Datagram Protocol)

Application Layer Protocol -> Protocols such as HTTP, SMTP for e-mail, BitTorrent for P2P, and SIP for IP telephony use transport layer protocols to standardize and enable specific data packet transferring methods over the Internet.

Technical and Business Considerations

- **Connectivity issues**
- **Network Bandwidth and Latency Issues** (***Bandwidth***-how much data can be transferred within a unit time, ***End-to-End Bandwidth***- the transmission capacity of links, ***Latency***-time for a packet to travel from one node to another, ***Jitter***-how consistent the given latency is)
- **Cloud Carrier and Cloud Provider Selection** (The necessary level of connectivity and reliability, Required collaboration of the cloud carriers)

Data Center Technology

Modern data centers exist as specialized IT infrastructure used to house centralized IT resources, such as servers, databases, networking and telecommunication devices, and software systems.

Data centers are comprised of:

Virtualization:

Physical IT resource: the facility infrstructure that houses computing systems and equipment, together with hardware systems and their operating systems.

Virtual IT resource: comprised of operational and management tools that are often based on virtualization platforms that abstract the physical computing and networking IT resources as virtualized components that are easier to allocate, operate, release, monitor and control.

Standardization and Modularity:

It aggregates multiple identical building blocks of facility infrastructure and equipment to support scalability, growth, and speedy hardware replacements.

Automation:

Data centers have specialized platforms that automate tasks like provisioning, configuration, patching, and monitoring without supervision.

Remote Operation and Management:

Most of the operational and administrative tasks of IT resources in data centers are commanded through the network's remote consoles and management systems. Most of the tasks are carried out from the central room except hardware and cabling jobs.

High Availability:

Data centers are designed to operate with increasingly higher levels of redundancy to sustain availability.

Security-Aware Design, Operation, and Management:

Organizations are afraid of: possible malicious attack and keeping data outside of organization's physical boundary. (Network isolation, firewalls, monitoring tools, big data analysis)

Facilities:

Several layout areas, various power supplies, cabling, control stations that regulate heating, ventilation, air conditioning, fire protection.

Computing hardware:

Standardized commodity servers with a number of computing hardware technologies such as:

- *Rack technology:* standardized rack with interconnects for power, network, and internal cooling
- *CPU architecture:* support for various CPU types: x86-32bits, x86-64bits, RISC, CISC, etc.
- *Multi-core CPU architecture:* hundreds of physical & logical processing core in single unit of standardized racks
- *Redundancy & hot-swap technology:* hard disks, power supplies, network interfaces, storage controller cards, etc.

Storage hardware:

One of the most difficult task to deal with in data center and many different levels of technologies for fast access, data availability, massive data accommodation, etc.:

- **RAID** (Redundant Array of Independent/Inexpensive Disks) – divide and replicate data
- **IO caching** – disk access time, performance
- **Hot-swapping** –
- **Storage virtualization** –
- **Data replication** – memory snapshot, volume cloning, mirroring,

Storage topology:

DAS (Direct Attached Storage): storages directly attached to a host system via block-level channel protocol such as SCSI/FC

NAS (Network Attached Storage): storages attached to a number of host systems via file-level network protocols such as NFS/CIFS/SMB – while providing file-level data sharing among multiple hosts

SAN (Storage Area Network): storages attached to multiple hosts via block-level network protocols such as Fibre Channel, Infiniband, iSCSI, etc.

Network hardware:

Five network subsystems:

1. Carrier & external network interconnection firewalls and VPN gateways
2. Web-tier load balancing and acceleration XML pre-processors, encryption/decryption appliances (web acceleration), layer 7 switching devices (content-aware load balancing), etc.
3. LAN fabric layer 4 or lower switching devices, VLAN, load balancing, failover (redundant connectivity), etc.
4. SAN fabric
5. NAS fabric

DWDM(Dense Wavelength Driven Multiplexing) - purpose of high-speed real-time data replication between data centers

Virtualization Technology

Virtualization is the process of converting a physical IT resource into a virtual IT resource.

Most types of IT resources can be virtualized, including:

- Servers – A physical server can be abstracted into a virtual server.
- Storage – A physical storage device can be abstracted into a virtual storage device or a virtual disk.
- Network – Physical routers and switches can be abstracted into logical network fabrics, such as VLANs.
- Power – A physical UPS and power distribution units can be abstracted into what are commonly referred to as virtual UPSs.

Hardware Independance: Creating standardized soft (virtual) copies of physical IT resources → eliminating hardware dependency

Server Consolidation: Creating different multiple virtual servers on a single physical server

Resource Replication: Virtual servers are created as virtual disk images that contain binary file copies of hard disk content.

Operating System-Based Virtualization: Easy to deal with hardware compatibility issues even with absence of a specific hardware driver. Host OS services to be utilized: backup/recovery, integration to directory service, security management.

Hardware-Based Virtualization: This option represents 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.

Virtualization Management: *VIM (Virtualization Infrastructure Management) tools* – collectively manage virtual IT resources from a centralized & dedicated management computer (controller).

WEB Technology

Web technology is generally used as both the implementation medium and the management interface for cloud services.

Multitenant Technology:

The multitenant application design was created to enable multiple users (tenants) to access the same application logic simultaneously.

The **differences** lie in what is multiplied within a physical server acting as a host:

• **With virtualization:** Multiple virtual copies of the server environment can be hosted by a single physical server.

• **With multitenancy:** A physical or virtual server hosting an application is designed to allow usage by multiple different users. Each user feels as though they have exclusive usage of the application.

Tenants can individually customize features of the application:

- *User Interface* (Special look for application interface)
- *Business Process* (customize rules, logic and workflows on the business process)
- *Data Model* (extend the data schema to include or exclude fields)
- *Access Control* (independently control the access rights for users and groups)

Characteristics of multitenant applications:

- *Usage Isolation* (usage behavior of one tenant does not affect the app availability)
- *Data Security* (tenants cannot access data that belongs to other tenants)
- *Recovery* (separately executed for the data of each tenant)
- *Application Upgrades* (tenants are not negatively affected by the synchronous upgrading of shared software artifacts)
- *Scalability* (the application can scale to accommodate increases in usage by existing tenants and/or increases in the number of tenants)
- *Metered Usage* (tenants are charged only for the application processing and features that are actually consumed)
- *Data Tier Isolation* (tenants can have individual databases, tables, and/or schemas isolated from other tenants)

Chapter 6. Fundamental Cloud Security

Information security is a complex ensemble of techniques, technologies, regulations, and behaviors that collaboratively protect the integrity of and access to computer systems and data.

Confidentiality is the characteristic of something being made accessible only to authorized parties.

Integrity is the characteristic of not having been altered by an unauthorized party (Data cloud customers stored **=** data cloud customers retrieved)

Authenticity is the characteristic of something having been provided by an authorized source.

Availability is the characteristic of being accessible and usable during a specified time period.

Threat is a potential security violation that can challenge defensesin an attempt to breach privacy and cause harm.

Vulnerability is a weakness that can be exploited either because it is protected by insufficient security controls, or because existing security controls are overcome by an attack.

Risk is the possibility of loss or harm arising from performing an activity.

Security Controls are contrameasures used to prevent or respond to security threats and to reduce or avoid risk.

Security Mechanisms are components comprising a defensive framework that protects IT resources, information and services.

Security Policies determine usage of security controls and mechanisms.

Threat Agent

A threat agent is an entity that poses a threat because it is capable of carrying out an attack. Internal and external threats by humans or software programs.

Anonymous Attacker is a non-trusted cloud service consumer without permission (External software programs) that launches network-level attacks through public networks.

Malicious Service Attacker is able to intercept and forward the network traffic that flows within a cloud. It typicaly exists as a service agent with compromised or malisious logic.

Trusted Attacker (Malicious tenants) 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.

Malicious Insider are a human threat agents acting on behalf of or in relation to the cloud provider. Typically current or former employees or third parties with acces to the cloud provider's promise.

Cloud Security Threats

Traffic Evaesdropping occurs when data being transferred to or within a cloud (usually from the cloud consumer to the cloud provider) is passively intercepted by a malicious service agent for illegitimate information gathering purposes.

Malicious Intermediary threat arises when messages are intercepted and altered by a malicious service agent, thereby potentially compromising the message's confidentiality and integrity.

Denial of Service is the attack which overloads IT resources to the point where they cannot function properly. Typically intentional overloading shared IT resource by generating excessive messages, consuming full network bandwidth, or sending multiple requests that consume excessive CPU time and memory.

Insufficient Authorization attack occurs when access is granted to an attacker by mistake, resulting in the attacker getting access to IT resources that are normally protected. Another case (**Weak Authentication**) when weak passwords or shared accounts are used to protect IT resources.

Virtualization Attack provides multiple cloud consumers with access to IT resources that share underlying hardware but isolate from each other.

Container Attack introduces a lack of isolation from the host operating system level.

Additional Consideration

Flawed Implementation: If the cloud provider’s software and/or hardware have inherent security flaws or operational weaknesses, attackers can exploit these vulnerabilities to impair the integrity, confidentiality, and/or availability of cloud provider IT resources and cloud consumer IT resources hosted by the cloud provider.

Security Policy Disparity: When a cloud consumer places IT resources with a public cloud provider, it may need to accept that its traditional information security approach may not be identical or even similar to that of the cloud provider.

Contracts: Cloud consumers need to carefully examine contracts and SLAs put forth by cloud providers to ensure that security policies, and other relevant guarantees, are satisfactory when it comes to asset security. The amount of reliability assumed by the cloud provider and/or the level of indemnity that the cloud provider may ask for must be specified in a clear language.

Risk Management: Before adopting cloud platform, potential cloud consumers are encouraged to perform a formal risk assessment as part of a risk management strategy.

- **Risk Assessment** is identifying potential vulnerabilities and shortcomings that threats can exploit. Statistics and other information about past attacks.
- **Risk Treatment** may agree to assume responsibility as part of its contractual obligations. Mitigation policies and plans are designed during the risk treatment stage with the intent of successfully treating the risks that were discovered during risk assessment.
- **Risk Control** is related to risk monitoring, a three-step process that is comprised of surveying related events, reviewing these events to determine the effectiveness of previous assessments and treatments, and identifying any policy adjustment needs.