

- ~~Routing is a service of Network layer to find the best route.~~
- ~~Routing is performed by applying routing protocols and using the decision tables called *routing tables* in each router.~~
- ~~Forwarding is the action performed by a router on the basis of routing protocol and routing table according to the destination address of each packet received at any interface.~~
- ~~At network layer, each message from higher layer is broken down into *packets*.~~
- ~~A router performs packet switching.~~
- ~~Types of routing:~~
 - ~~Unicast routing: A router forwards the packet to only one of the attached networks.~~
 - ~~Multicast routing: A packet is forwarded to multiple attached networks.~~
- ~~Routing a packet from a source host to destination host can also be defined as routing a packet from a source router (the default router of the source host) to a destination router (the router connected to the destination network) through the intermediate routers using routing algorithms.~~
- ~~Types of routing:~~
 - ~~Connectionless routing: All packets of the same message are treated independently and may or may not follow the same route.~~
 - ~~Connection-oriented routing: All the packets of same message are labeled and routed through a *virtual circuit* or a fixed route.~~
- ~~An internet can be considered as a graph with each network as an edge and each router as a node.~~
- ~~In a weighted graph, each edge has a weight or cost.~~
- ~~Least cost routing can be performed. Example algorithms: Distance Vector routing, Link State routing~~

Lesson No. 07

ADVANCED TOPICS OF COMPUTER NETWORKS

Module No – 026: Broadband Networks & Internet:

- All clouds are inherently **dependent** upon **internetworking** or **Internet** for ubiquitously **remote** provisioning of **IT** resources.
- The cloud providers and consumers connect to **Internet** through **ISPs**.
- The **largest backbone** networks of the Internet are strategically interconnected by **core routers**.
- The **core-routers** connect the **international** networks.
- The Internet has become a dynamic and complex **aggregate** of **ISPs**.
- There is a **hierarchical topology** for worldwide interconnectivity composed of **tiers**.
- There are **three** tiers of worldwide connectivity:
 - **Tier 1** consists of **large-scale** international connectivity providers.
 - **Tier 2** consists of **large regional ISPs** connected to **tier 1**.
 - **Tier 3** consists of **local ISP providers** connected to **tier 2**.
- The **cloud providers** and **users** connect **directly** to **tier 3 providers**.

Module No – 027: Internet Architecture & Cloud Deployment:

- Internet supports the remote provisioning of IT resources.
- Cloud relies heavily upon Internet.
- The connectivity of end-users of cloud depends upon how the centralized resources of cloud are deployed.
- The cloud resources deployment can be either on-premises or Internet based.
- In cloud deployment using the on-premises, the provider sets up a fully controlled corporate network and a corporate Internet connection for the deployment of IT solutions and applications.
- In the on-premises deployment, the internal users access the cloud through corporate network. The remote users connect through internet by using virtual private network (VPN).
- A VPN creates a secure connection between a remote device and the corporate servers over the internet as if the device is inside the LAN.
- For the internet based deployment, the cloud provider has an Internet connection and all the internal and external users access the cloud resources through cloud provider's internet connection.
- In this deployment, there is an extra charge for internet connectivity.

Module No – 028: Scalable Computing over Internet:

- Scalable computing may refer to the dynamic resizing of the available computing resources (processing, memory, bandwidth, storage etc.) with demand.
- The growth of users and user demands for scalable computing over internet has been accompanied with matching growth in network, computing and resource management technologies.
- The computing platforms have evolved as follows
 - Mainframes (1950-70)
 - Minicomputers (1960-1980)
 - Personal computers (1970-1990)
 - Portable computers (1980-2000)
- Since 1990, the High Performance Computing (HPC) and High Throughput Computing (HTC) have been relying upon clusters, grids and the Internet clouds.
- The speed for HPC systems (supercomputers) has increased from Gflops in early 1990s to now Pflops in 2010.
- The network bandwidth has been doubling each year in the recent past (Gilder's law).
- Processor speed has been doubling every 18 months (Moore's law).
- Means that there has been a steady growth in these technologies.
- Fine grain (instruction level) parallelism and coarse grain (job level) parallelism are available.

- Ubiquitous computing is what refers to computing at any place and time using pervasive devices and wired or wireless communications.
- Utility computing works upon a business model in which the customers pay for computational resources from a provider.
- Cloud computing provides ubiquitous utility computing.

Module No – 029: Technologies for Network based Systems:

- The processor speed and network bandwidth have shown a remarkable growth in last few decades.
- The processor clock rate has risen from 10 MHz in 1970s to over 4GHz in 2010s.
- The network band has increased from 10 Mbps to over 100,000 Mbps
- The excessive heat generation from single processor core with high frequency has limited the maximum speed unless the chip technology matured.
- This has lead to the multi-core architecture of CPUs with dual, quad, six or more cores.
- The graphical processing unit (GPU) development has adopted a many-core architecture with hundreds to thousands of cores.
- Modern architecture of CPUs and GPUs have enhanced the instruction level parallelism (ILP) and the volume of millions of instructions per second (MIPS).
- Sun's Niagara CPU can provide 64 count for ILP.
- Intel's Core i7 990x can provide 159,000 MIPS execution rate
- The CPUs and GPUs are multithreaded, which means that each core can execute multiple processes or threads concurrently.
- A GPU unit has far more (but slower) cores than a multi-core CPU
- The DRAM memory chip capacity has increased from 16 KB in 1976 to 64 GB in 2011.
- The hard disk capacity has increased from 260 MB in 1981 to 3TB a few years ago.
- The flash memory and solid-state drives are rapidly evolving.
- Disk arrays are being utilized to enhance the storage.
- Servers can be connected to network storage such as disk arrays through storage area network (SAN)
- A disk array can be connected to client hosts through network attached storage (NAS)
- The high bandwidth networks in WAN scope can connect the host computers to network storage.
- A single host can be shared among multiple instances of operating systems through virtualization technology. ~~More on this latter.~~

Module No – 030: Web 2.0

- It is the second generation of world wide web.
- Lets people collaborate and share comments, media and information online.
- The web pages progressed from static to dynamic and interactive.

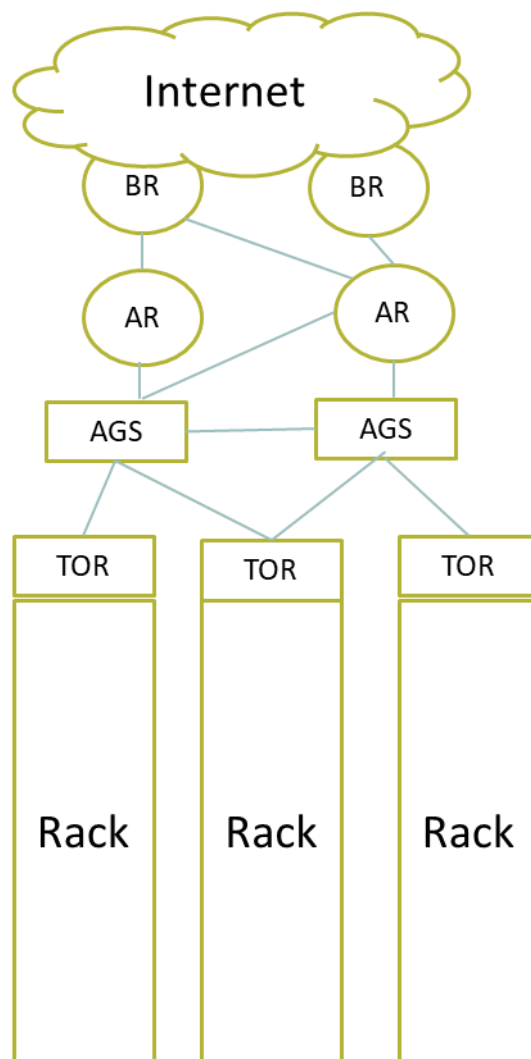
- Through **Asynchronous Javascript** and **XML** or **Ajax**, the **web** applications can send and receive data from a **web server** without interfering with the display and behavior of the existing page.
- Social networking and community oriented sites have emerged
- For example **myspace.com**, **facebook.com**, **twitter.com** etc.
- Users can contribute in web based blogs, wikis, online training, online education etc.
- **Really Simple Syndication (RSS)** feeds continuously keep the subscribers informed about **news**, **follow up updates** and **products**.
- Users can do **online chatting** and **share files through messenger** tools such as **Yahoo messenger**, **Skype**, **WhatsApp** etc.

Module No – 038: Virtual Private Network (VPN):

- A **VPN** extends a **private** network over **public** network and enables the users to communicate as if their devices are directly connected to the **private** network.
- A **VPN** creates a **secured** and **encrypted** network over a **less secured** network such as the **Internet**.
- Normally a **VPN** is provided and managed by a **service provider**.
- **VPN** allows the corporate employees to **securely** access the **applications** hosted over **enterprise LAN**.
- **VPN** is based upon **IP tunneling**.
- **IP tunneling**, or **port forwarding** is the transmission of **private** network **packets** over a **public network (Internet)** as the **payload** of **public network packets** such that the **routing devices** do not come to know about this.
- There are many protocols for VPN establishment and encryption: **IP Security (IPSec)**, **Secure Socket Layer(SSL)**, **Point-To-Point Tunneling Protocol (PPTP)**, **Multiprotocol Label Switching (MPLS)** etc.
- **VPN** although provide **secured** connectivity to extend a **private** network but the implementation may have **performance** issues.
- **VPN** is implementable over **Layer 1-3**.
- Types of VPN:
 - **Remote-access VPN**: A **VPN** client on user's device connected to **VPN** gateway of the enterprise.
 - **Site-to-site VPN**: Establishes a VPN between **two** networks over the **Internet** by using **VPN gateway**.
- **VPN** technology provides access to **cloud** resources. The **VPN** gateway exists in the **cloud** with a **secure link** provided by the **cloud provider**.

Module No – 040: Networking Structure of Cloud Hosting Data center:

- At the core of a cloud is a cluster of VMs/ physical servers.
- There can be tens of thousands of physical servers in a data center.
- With each physical server hosting multiple VMs, the cloud hosting networking structure becomes a little complicated.
- The cluster nodes are used for computations. Some nodes are used for workload allocation, some for monitoring and some for load balancing.
- Some node called gateway nodes provide the interface of cloud service/s to the outside world (through internet).
- Cloud hosting data center has a layered architecture for the Internet access.
- The servers are physically connected to layer 2 switches. There is a top of rack (TOR) in each rack. One server is connected to only one TOR switch.
- The TORs are connected to aggregate switches (AGS).
- The AGSs provide the cross rack inter VM connectivity.
- There are a few access routers (AR) and border routers (BR) at layer 3.
- The layer 2 AGSs are connected to BR through AR.
- The BRs are connected to Internet.
- Some problems solved by the hypervisor solutions:
 - The VMs hosted on one server may belong to different VLANs.
 - A single VLAN may span over multiple data centers.
 - A company may own multiple data centers and may want to migrate the VMs across data centers.

**Lesson No. 08****VIRTUALIZATION****Module No – 031:**

- **Virtualization** is a technology used to **enhance** the **utilization** of **computing resources**.
- A **single hardware machine** is **multiplexed** among **multiple virtual machines (VMs)**.
- A **software** based **virtual machine monitor/manager (VMM) or hypervisor** is a **program** that **manages** the **hardware** resources for the **VMs** and also keeps each **VM** from disrupting other **VMs**.
- Virtualization implementation levels:
 - **Instruction Set Architecture (ISA)** level: Executing **legacy code** over new machines using **ISA emulator tool** such as an **interpreter** which translate **one** instruction of **source code** into corresponding **instruction** of the target machine.

- **Hardware Abstraction level:** The **hardware** components (CPU, RAM, Disk, NIC) of a **physical** system are **virtualized** and shared among **virtual machines** using **Virtual Machine Monitor (VMM)** tool or **hypervisor** which performs as **abstraction layer**.
- **Operating System Level:** The **OS** running over a server accommodates multiple **containers** or **VMs**. The **host operating system** acts as the **abstraction layer** between **hardware** and the **containers**.
- **Library support level:** The **API** calls for **hardware** acceleration such as **vCUDA** stubs for **graphic processing units (GPUs)** are available at **VM** level.
- **Application level:** An **application** acts as a **VM** through **wrapping** of application in an **abstraction layer** which **isolates** it from **OS** and other applications. Another type is using **virtualization layer** as **programming** environment e.g; **Java Virtual Machine (JVM)**.

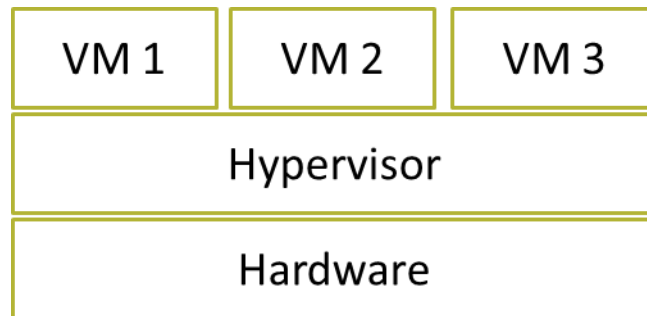
Module No – 032: Virtualization Structures:

- We know that the **virtualization layer** transforms the **physical hardware** into **virtual hardware**. There are **three** classes of **VM architectures**.
 - **Hypervisor Architecture:**
 - It is the **hardware** level **virtualization**. Also called the **bare-metal virtualization**
 - The **hypervisor** sits between the **hardware** and the **VMs** and manages the **VMs**.
 - Example: **Xen, VMware**
 - **Full-virtualization Architecture:**
 - The **guest operating system (OS)** or the **VM's OS** does not know that it is installed on a **VM**.
 - The **Virtualization layer** manages the **hardware acceleration**. For example **VMware**
 - The **virtualization layer** can be installed on **hardware** or on **host's OS**.
 - Some of the **instructions** of a **guest VM** are **directly** run on **hardware** to enhance the **performance**.
 - **Para-virtualization Architecture:**
 - The **guest OS** is **modified** to **comply** with **virtualization layer**. All **calls** for **hardware acceleration** are **handled** by **virtualization layer**.
 - For example: **KVM**

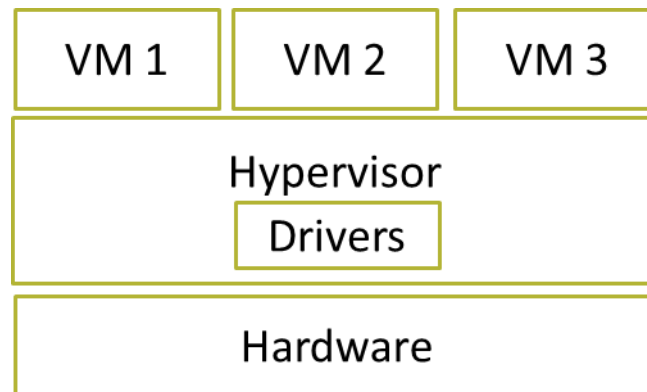
Module No – 033: Virtualization Architectures:

- **Hypervisor** transforms the **physical hardware** into **virtual hardware**.
- **Virtualization Architectures:**

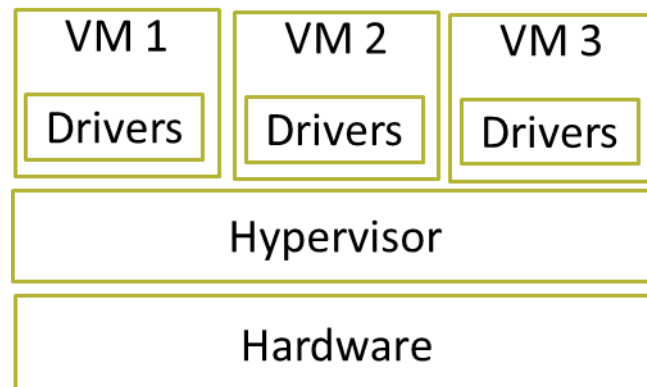
- **Generic:**



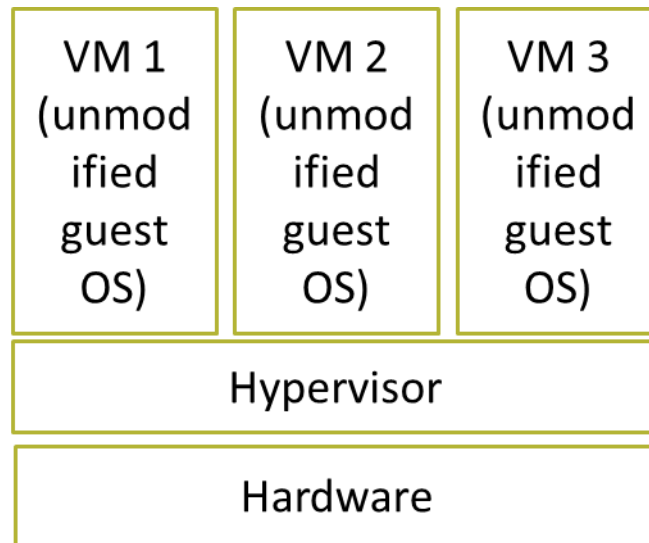
- **Monolithic:**



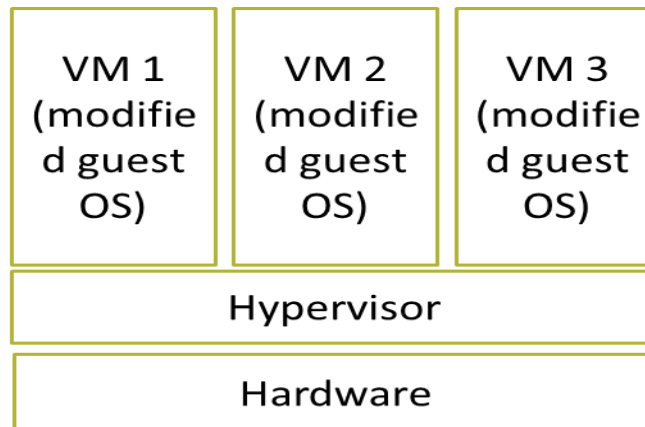
- **Microkernel:**



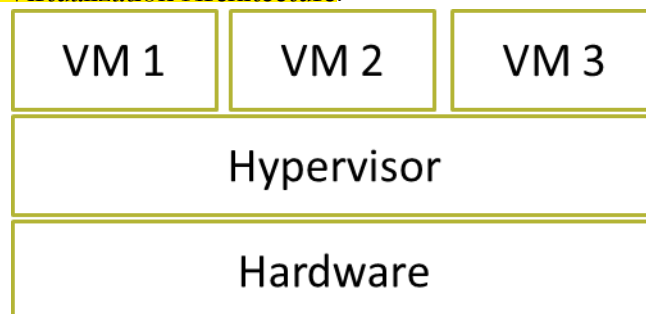
- Full Virtualization:



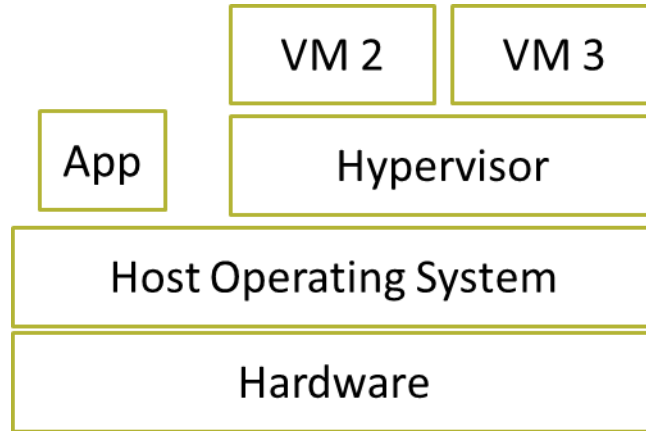
- Para Virtualization:



- Native Virtualization Architecture:



- Hosted Virtualization Architecture:



Module No – 034: Virtualization of CPU, Memory and I/O Devices:

- To support **virtualization**, processors such as **x86** architecture use a **special** mode and **instructions** known as **hardware-assisted virtualization**.
- In this way, the **hypervisor** is able to **trap** the sensitive instructions of the **guest OS** and its applications.
- The modern **processors** allow **multiple** processes to run **simultaneously**. Any process can execute a **critical** instruction to **crash** the whole system.
- Therefore, the critical instructions are executed in **privileged** or **supervisor** mode of the **processor**. The **OS** controls this mode on behalf of the **processes** being executed.
- The second type of instructions are **non-privileged** or **non-critical** instructions which are run in **user-mode** of the **processor**.
- **CPU Virtualization**: A **CPU** is **virtualizable** if it is able to run the **privileged** and **un-privileged instructions** of a **VM** in **user mode** and the **hypervisor** running in **supervisor** mode.
- **Memory Virtualization**: Traditionally, the **OS** performs the **mapping** of **virtual memory** to **machine memory** by using **page tables**.
- The modern **x86 CPUs** include the **memory management unit (MMU)** and **translation lookaside buffer (TLB)** to **optimize** virtual memory performance.
- However, in **virtualization** environment, the **memory virtualization** involves the **sharing** and **dynamic** allocation of **physical** memory of the system to the **physical** memory of the **VMs**.
- The **guest OS** performs the **virtual** to **physical memory mapping** of the **VM**, while the **hypervisor** performs **mapping** of **physical memory** to **machine memory**.
- **I/O Virtualization**:
 - It is done in either of the **three** ways:

- **Full device emulation:** The device is emulated in **software** located in **hypervisor**. The **hypervisor** interacts with the **real** device. The **VM** interacts with the **virtual device**.
- **Para-virtualization based I/O:** The **guest OS** interacts with the device through its **frontend driver**. The **frontend driver** interacts with a **backend driver** of the device. The **backend driver** interacts with the **device**.
- **Direct I/O virtualization:** This type of I/O virtualization allows the **VMs** to **directly access** the device.

Module No – 035: Virtual Clusters:

- A **virtual cluster** consists of several **VMs** hosted on a **physical cluster**.
- The **VMs** are interconnected through a **virtual network** across **multiple physical networks**.
- The **nodes** can be **physical** or **virtual** machines and can **grow or shrink dynamically**.
- The failure of a host can make the hosted VMs **unavailable** but the **virtual cluster** does not **collapse**.
- The **failure** of a **VM** does not **fail** the **host**.
- A **physical cluster** may host multiple **virtual clusters**.
- A **virtual cluster** may span over multiple **physical cluster**.
- In order to deploy a **virtual cluster**, several **VMs** with **installed OS** and **application software** are required.
- The **deployment time** is to be as **quick** as possible.
- **Templates** can be used to deploy the **VMs** from.
- A **template** is a **disk image** with **preinstalled OS** with or without certain **applications**.
- A **suitable template** can be **copied** as **disk image** of a **VM**. This saves **time** of **installing** and **configuring**.
- When the **VM** is ready and up, it is deployed to a **suitable** host.
- The **VM** then joins a **virtual cluster**.
- All of the above can be done **manually** as well as **full or partially automated**.
- Reasons of virtualization:
 - **Sharing of resources**
 - **Isolation of users of shared resource**
 - **Aggregation of smaller resources into a single big virtual resource (e.g., Storage)**
 - **Dynamic relocation/provisioning of virtual resources is easier than physical resources**
 - **Easier management of virtual resources/devices/machines.**

Module No – 036: Virtual Machine (VM) Migration:

- **VMs** can be migrated from **one** host to another for:
 - **Server Load balancing**
 - **Server consolidation**

- Remedy for failover hosts and VMs
 - Remedy for resource shortage for a VM
- A VM can be in any of the following states:
 - Powered-off
 - Suspended
 - Paused
 - Powered-on
- The following options are available for VM migration:
 - Cold migration: The VM has to be powered-off before migration.
 - Warm migration: Suspended VM migration.
 - Live migration: For powered-on VM with zero downtime and full availability.
- A VM is made of two basic components:
 - VM state: The processor and RAM contents
 - Virtual hard disk: Residing on network storage or on host's hard disk
- Live migration of VM means zero downtime of OS, connectivity and applications running on the VM.
- For live migration, the VM state is to be copied from source to destination host. The virtual disk can also be migrated through live storage migration feature of the hypervisor.
- Modern day hypervisors provide unbroken network connectivity of the VM during live migration.
- During the live migration, the state and storage of the VM keeps on working on source host to avoid down-time.
- For live migration of a VM with the virtual hard disk on network accessed shared storage, the virtual hard disk need not to be copied if the destination host can access that network based storage.
- Migrating the virtual hard disk is time consuming as well as network bandwidth consuming due to multi-Gigabyte migration.
- A better solution is to use the network storage.

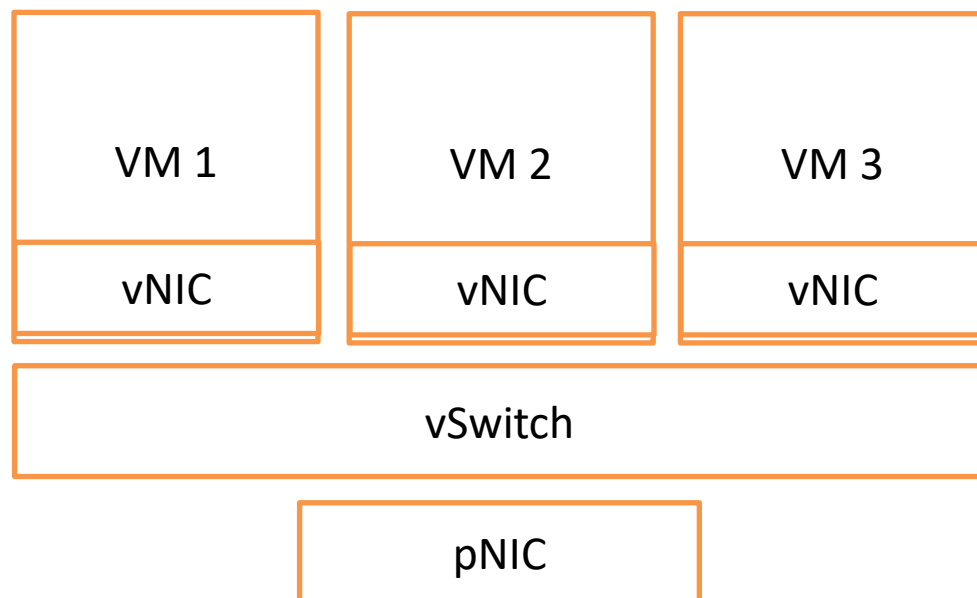
Module No – 037: Virtualization for Data Center Automation:

- A data center is a facility with networked computers and is used by businesses and other organizations to process, store and share large amounts of data.
- Companies like Google, Yahoo, Amazon, Microsoft, IBM, HP, Apple etc. have invested billions of dollars for constructing the data centers.
- Data center automation refers to the dynamic provisioning of hardware and software resources to millions of users simultaneously.
- Data centers can host Clouds.
- Data center automation is triggered by the growth of virtualization products.
- The data center owner has three major considerations:
 - Assuring Performance and QoS
 - Increase resource utilization
 - Saving costs

- Enhanced resource allocation (to jobs and/or VMs) may be performed in data centers to assure performance and QoS.
- The over allocation of computing resources may result in decrease in average utilization of these resources.
- This also leads to increased costs due to power consumption.
- Example: A VM hosted on a server with 1.5 GHz *4 cores and 16 GB of RAM is allocated 1.5GHz * 2 vCPUS, 4 GB vRAM (half of the processing and 1/4th RAM).
- Suppose if there are two such VMs. But the overall average workload of the hosted VMs keeps the physical utilization to less than 50%. This is a resource wastage as 50% of the resources remain idle.
- Server consolidation* is a technique by which more VMs are aggregated on a single server (by migrating jobs/VMs to it) while assuring performance and QoS.
 - This increases the resource utilization across data center.
 - More servers are available to take more workload. . Otherwise, the idle servers can be shut down to save power.
- Virtualization technology also helps in setting of virtual storage (over VMs) to offer virtual disks to other VMs.
- Virtualization can synchronize with cloud management systems to dynamically provision cloud services and billing systems.
- Hence, virtualization is essential for Cloud computing.

Module No – 039: Network Virtualization:

- Multiple virtual Network Interface Cards (vNIC) are linked to physical NIC or pNIC through a virtual Switch (vSwitch) inside a hypervisor.



Network Virtualization

- A **virtual network** consists of **virtual nodes** and **virtual links**.
- **Network virtualization** establishes the **coexistence** of multiple **virtual** networks.
- **Network virtualization** proposes the **decoupling** of traditional **ISP** functionalities such as **infrastructure setup and management** from the **creation and management of virtual networks**.
- It is possible to use **physical** infrastructures of multiple **providers** to **dynamically** compose **virtual network/s**.
- Network virtualization technologies:
 - **Virtual Local Area Network (VLAN)**: Logically grouping the hosts with common interest into a **single broadcasting domain**.
 - **Virtual Private Networks (VPN)**: A dedicated communications network of enterprise/s and user/s by using **tunneling** over **public** networks (**Internet**).

Lesson No. 09

ESSENTIAL CHARACTERISTICS OF CLOUD COMPUTING

Module No – 041:

- **On-demand self-service**: The **user** can **automatically** be **allocated** the computing resources without any **manual** operations (except the **initial signing up process**). The cloud management **software** handles the resource **management** and **provisioning**.
- **Broad Network Access**: The cloud resources can be accessed through network through broad range of **wired** and **wireless** devices. Various connectivity technologies are available.
- **Resource pooling**: Resources (**Computing, memory, storage, network**) are available in volumes and therefore can be **pooled**. The resources can be **physical** or **virtual**. **Multiple** users can **simultaneously** share these resources through **dynamic allocation** and **reallocation**.
- **Rapid elasticity**: The cloud resources are **virtually unlimited**. So much so, the provisioning of these resources can shrink and expand elastically according to demand.
- **Measured Service**: The resource **usage** is **charged** by the **provider** from **users**, according to usage.

Module No – 043: Revisiting NIST Definition of Cloud Computing

“Cloud computing 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.”

National Institute of Science and Technology (NIST) USA

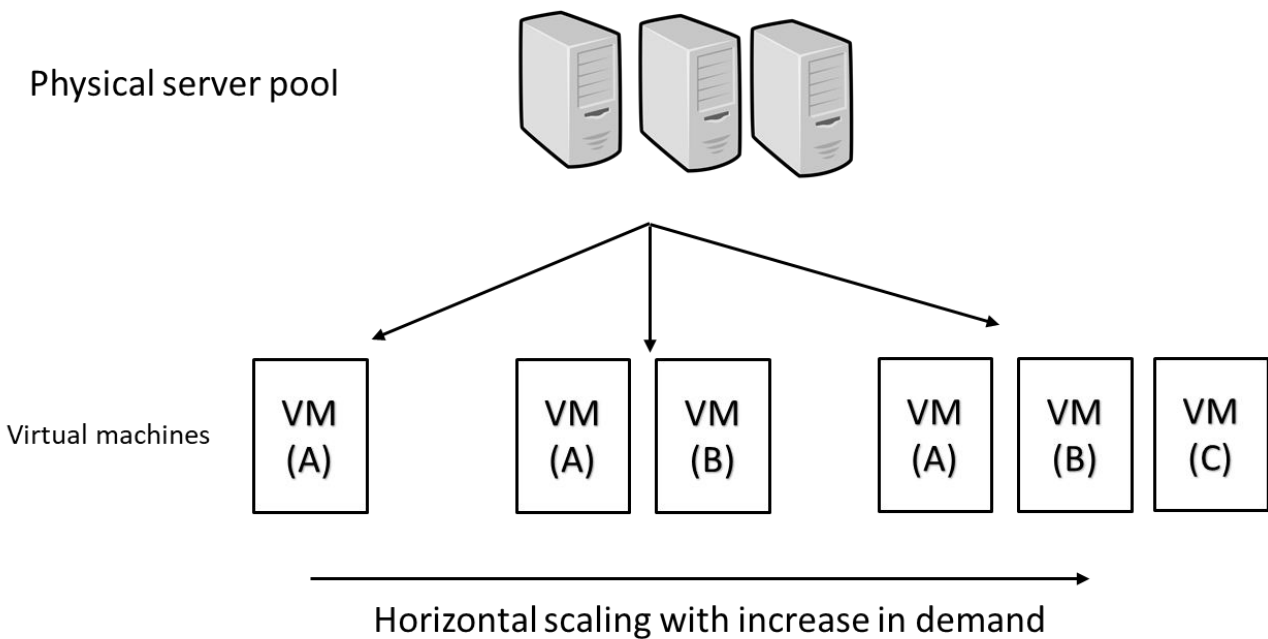
- Cloud computing can fulfill the business drivers such as
 - **IT Capacity Planning**
 - **Cost Reduction**
 - **Organizational Agility**

Module No – 044: Some key terms about Cloud Computing:

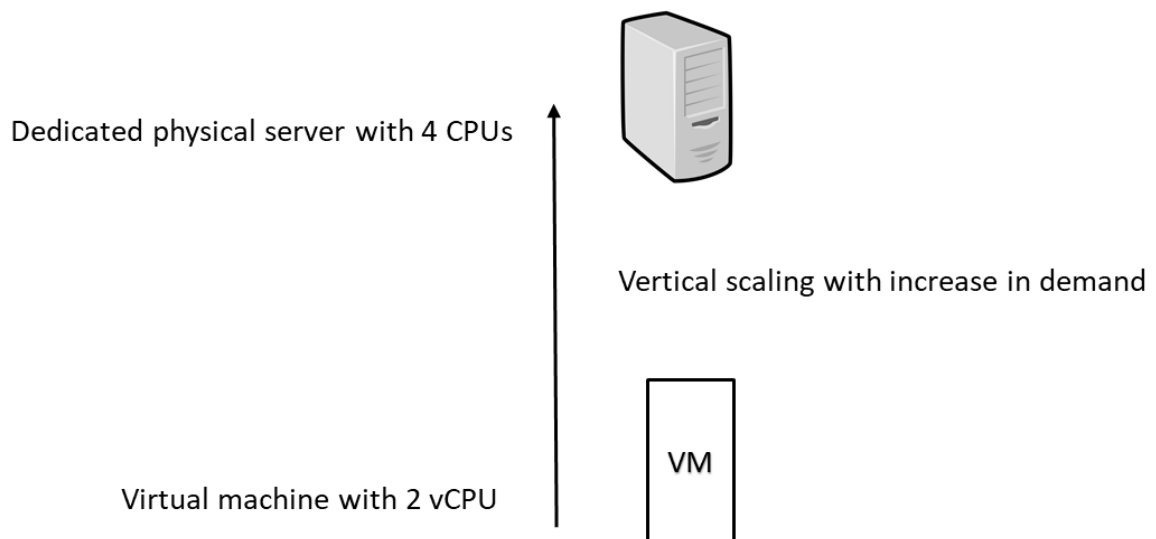
- Some key terms and concepts essential for understanding Cloud Computing course:
 - IT Resources
 - On-premises
 - Cloud Consumers
 - Cloud Providers
- Cloud IT Resources: Can be physical or virtual resources (virtual resources are implemented in software):
 - Physical/Virtual machines/servers
 - Physical/virtual storage
- On-premises: An IT resource which is hosted/located at the enterprise's premises.
 - It is different from a Cloud resource since a Cloud resource is hosted on Cloud.
 - An on-premises IT resource can be connected to a Cloud resource and/or can be moved to a Cloud.
 - However the distinction is difficult for private clouds.
- Cloud Providers: The party providing the cloud-based IT resources.
- Cloud Consumer: The user of cloud-based IT resources is called cloud consumer.

Module No – 045: Scaling, Cloud Service Providers & Consumers:

- Scaling: It refers to the ability of an IT resource to handle increased or decreased usage demands.
 - ~~○ Scaling: It refers to the ability of an IT resource to handle increased or decreased usage demands.~~
- Following are the types of scaling:
 - Horizontal scaling: It is the scaling out or scaling in of the IT resources of same type. The number of resources increases or decreases according to load.
 - Commodity hardware can do the work, instantly available IT resources, not limited by hardware capacity



- **Vertical scaling:** When an IT resource is **replaced** with a **resource** of **higher** capacity (**scaling up**) or when replaced with the resource of **lower** capacity (**scaling down**) according to **workload**.
- **Specialized server** are **required**, instantly available IT resources, additional setup is **required** (**downtime required during replacement**), **limited** by **maximum** hardware **capacity**, less common in **Cloud**.

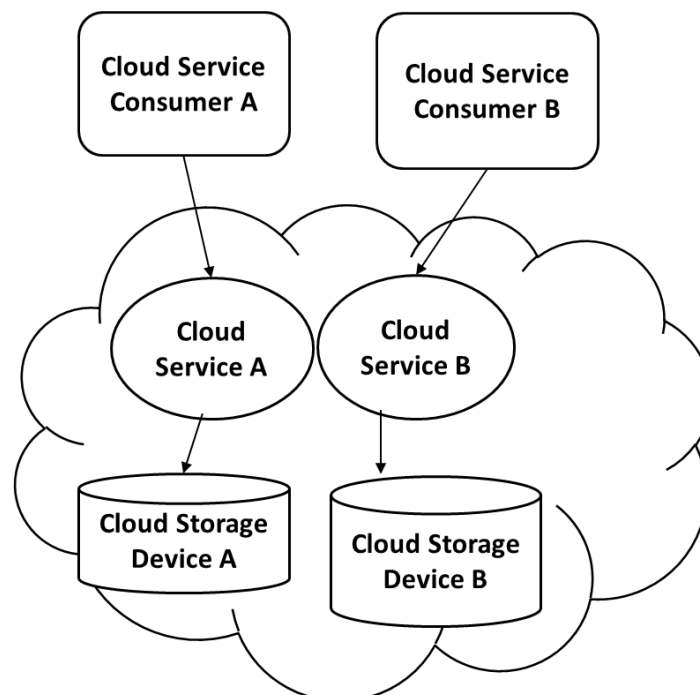


- **Cloud Service:**
 - Any IT resource (**software/VM**) that is made remotely available by the **cloud** provider.

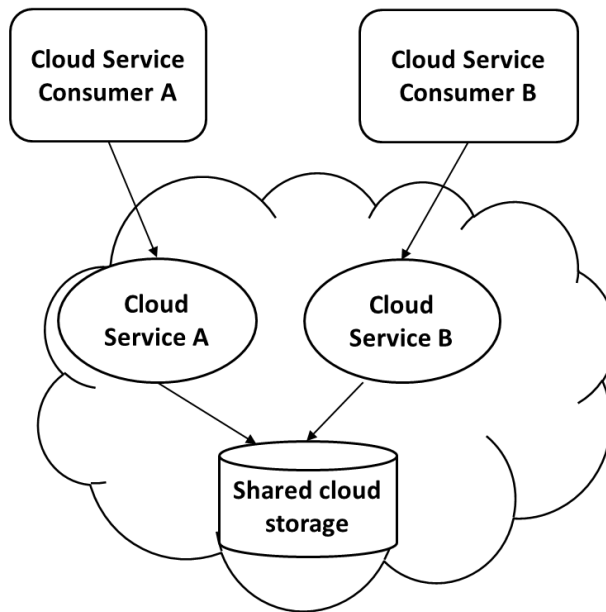
- Remember that not all the IT resources deployed in a **cloud** environment are **remotely accessible**. Some resources are used within the **Cloud** for **support** and **monitoring** etc.
- The **human** users interact with a **leased VM**.
- **Client programs** interact with **cloud software** service/s through **API calls**.
- The **software program** and **service** accessing a **cloud service** is called a **cloud service consumer**

Module No – 053: Multitenancy:

- A **software architecture** consisting of **software executing** over a **server** and serves different **users (tenants)** whereby each **tenant** is **isolated** from the others.
- **Cloud computing** serves different **cloud consumers** by using **virtualization software** frequently.



In single-tenant environment, there is a separate IT resource for each tenant.

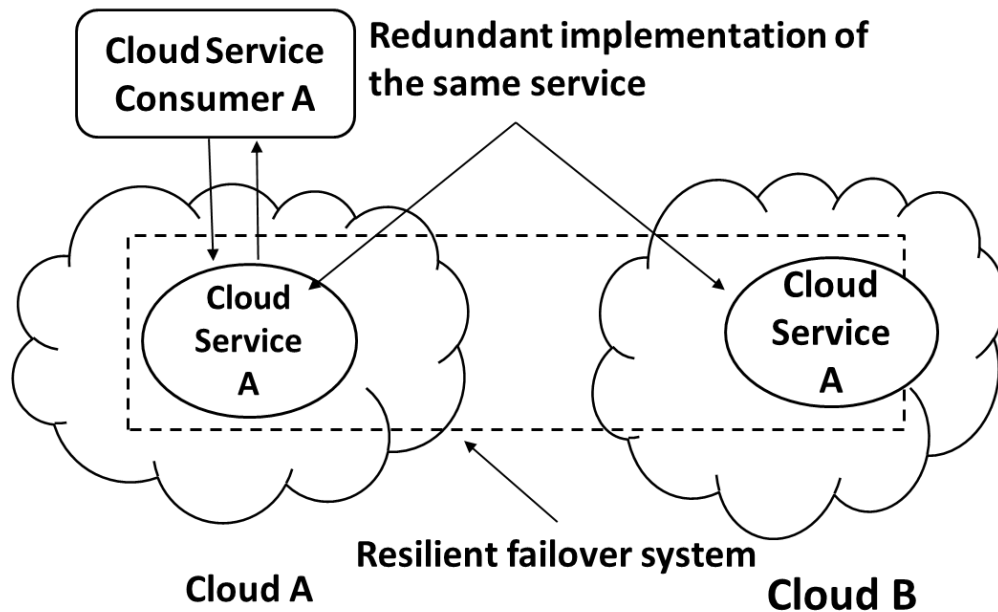


Multi-tenant environment, a single instance of an IT resource such as Cloud storage device serves multiple consumers.

- The cloud provider pools the IT-resources by using multitenancy technology to dynamically assign and reassign these resources according to cloud consumers' demands.
- The physical as well as virtual resources are multi-tenanted (or shared) by using statistical multiplexing.

Module No – 054:

- **Resiliency:** The ability of a computer system to recover from a failure is called resiliency.
 - The redundant implementation of IT-resources paves the way to a resilient system.
 - The whole system is pre-configured so that as soon as a resource fails, the processing is automatically handed over to the redundant resource.
 - Resiliency is one of the features of cloud computing whereby the redundancy of IT-resources is implemented at different physical locations and/or in different clouds.
 - For example the data can be kept at two different locations and replicated. If the primary hard disk fails, the secondary drive takes the data connectivity.
- A cloud service can be configured at two different VMs (A and B) and each VM is placed on a separate server or a different cloud. VM B is kept as failsafe resource. In case VM A fails, the VM B starts processing the user service user/s requests.



Lesson No. 10

BENEFITS OF CLOUD COMPUTING

Module No – 046:

- The immediate benefit of using **Cloud** is the **reduction** in **initial cost**.
- The initial costs include:
 - **Infrastructure costs:**
 - **IT equipment**
 - **Software**
 - **Networking**
 - **Construction costs**
 - **Installation costs**
 - The **infrastructure** costs can be regarded as **capital investments or ownership costs**. The **cloud** saves the initial upfront **ownership** costs. The cloud offers affordable and attractive packages for services obtained in **large** volume. The cloud reduces **investment** and **proportional** costs.
 - **Proportional cost or operational costs** (~~as discussed before~~): The **cloud rental** can replace this cost. The rental costs are **highly** competitive.
- The cloud provider can increase the **profit** by increasing the **resource utilization**, using proven practices and by **optimizing** the **cloud architecture**.
- Common measurable benefits for the cloud consumers are:
- **Pay-as-you-go rental for short term usage**

- The availability of **virtually unlimited resources** on **demand** with **negligible wait time** for **provisioning**.
- The IT resources can be **added or removed** in a **fine grained level** e.g., **1 GB** of storage increments
- **Applications** and **resources** can be migrated across regions if required.

Module No – 047: Increased Scalability, Availability & Reliability

- **Increased scalability**: The **cloud** can **dynamically** and **instantly** provide the computing resources.
- This provision can be on **demand** or as per **user configuration**.
- Similarly these IT resources can be released **automatically** or **manually** with the **decrease** in **processing** demand.
- This **dynamic scalability** avoids the **over-provisioning** and **under-provisioning** and the **associated** disadvantages.
- **Availability**: The **availability** of IT resources sometimes can be referred to **profit** and **customer retention**.
 - If an IT resource becomes unavailable (such as a **database dealing with clients' orders**) then this may result in **customer dissatisfaction and loss of business**.
- **Reliability**: The **reliability** of IT resources is very important for **continual** business **data processing and response time**.
 - The failure of any **IT** resource can be cause the **collapse** the **IT** system. For example failure of the **Ethernet switch** may **crash** a **distributed application**.
- The **modular structure** and **resource redundancy** in **cloud** **increases** the **availability** and **reliability**. **Cloud**, on the other hand provides a **guaranteed level** of **availability** and **reliability** through a **legal agreement** called **service level agreement (SLA)** between the **cloud provider** and **cloud user**.
- The **recovery time** after failure is the added **penalty**. **It** is the time when the system remains **unavailable**.
- The modular structure and resource redundancy in **cloud** increases the **availability** and **reliability**. It also improves the **recovery time**.

Lesson No. 11

RISKS AND CHALLENGES OF CLOUD COMPUTING

Module No – 048:

- The term **vulnerability** refers to a state of **being attacked**.
 - Moving the business data to **cloud** can introduce **vulnerabilities** and **security risks**.
- The term **security framework** refers to the **procedures** and **practices** for **securing** the **resource** such as **data, network and IT** infrastructure.

- Unless the cloud **provider** and cloud **user** are covered under **same security framework**, the **vulnerabilities** are **unavoidable**.
- The **cloud provider and user** have to be in a **trust relationship**. The factors affecting the trust may include the following facts:
 - The data is being accessed remotely.
 - There are multiple users sharing the cloud based IT resources such as virtual storage.
 - The cloud provider has a privileged access to the users' data.
 - The security of the data depends upon the security policies of the provider and the consumer.
 - There can be malicious consumers (human and automated) who can benefit from the security vulnerabilities of the cloud environment by stealing and/or damaging the business data.

Module No – 049:

- **Reduced operational governance control:** The cloud **consumer** gets a **lesser privileged** control over the **resources leased** from the **cloud**.
- There can be risks arising as to how the cloud provider manages the **cloud**.
- An **unreliable cloud provider** may not abide by the guarantees offered in **SLA** of the cloud **services**. This will directly affect the **quality of cloud consumer** solutions (**enterprise software**) which rely upon these services.
- The **cloud consumer** should keep track of **actual level of service** being provided by the **cloud provider**.
 - The **SLA violation**s can lead to **penalties** receivable from the **cloud provider**.
- Limited **portability** between cloud providers: Due to lack of industry standards for **cloud computing**, the **public clouds environments** remain **proprietary** to their **providers**.
- It is quite challenging to move a **custom-built software** from one **cloud** to another if it has **dependencies** upon the **proprietary** environment (such as **security framework**) of the former **cloud**.
- **Multi-regional compliance and legal issues:** Cloud providers tend to set their **data centers** in regions favoring **affordability** and/or **convenient**. This may lead to **legal** issues for cloud **provider** as well as cloud **consumers**.
- Some countries such as some **UK** laws require the **personal data of UK citizens** to be hosted inside **UK**.
- Thus a **cloud provider** with multi-regional data centers including **UK**, can not migrate the **UK citizen's personal data** outside UK.
- The **UK citizens** are legally bound to keep the **personal data** on **clouds** hosted in **UK** only.
- Some countries such as **USA** allows government agencies' **access** to data hosted inside **USA**.
- Despite that the owners of this data are neither residing inside nor the citizens of USA, but still their data is accessible by the **USA government agencies** if hosted inside **USA**.

ROLES AND BOUNDARIES IN CLOUD COMPUTING

Module No – 050:

- **Cloud provider:** The organization that provides the IT resources.
 - Responsible for providing IT resources as per SLA.
 - Also performs the management and administrative tasks to assure flawless provisioning of cloud services.
 - A cloud provider usually owns the IT resources of the cloud.
 - It is also possible that the cloud provider resells the cloud services leased from another cloud providers.
- **Cloud consumer:** The organization or individual who has contracted with cloud provider to lease/rent the cloud IT-resources through user interface and/or through software API calls.
 - In the later case, a cloud consumer uses a cloud service consumer (a software program) to interact/use a cloud service.
- **Cloud Service Owner:** Is the one who owns the cloud service. Can be:
 - Cloud consumer: If the deployed service is on leased IT-resources.
 - Cloud provider: If the cloud provider has deployed the service on cloud IT-resources.
 - A cloud service owner may not be the owner of the cloud IT-resource.

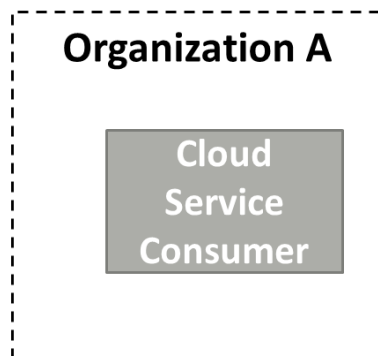
Module No – 051:

- **Cloud Resource Administrator:** This role is responsible for administering the cloud resources (including cloud services).
 - Cloud resource administrator can be:
 - Cloud consumer (as cloud service owner)
 - Cloud provider (when the service resides inside the cloud)
 - Third party contracted to administer a cloud service
- Additional roles:
 - **Cloud Auditor:** Provides an unbiased assessment of trust building features of the cloud. These include the security, privacy impact and performance of the cloud. The cloud consumer may rely upon the cloud audit report for choosing a cloud.
 - **Cloud Broker:** A party that provides mediation services to cloud providers (seller) and cloud consumers (buyer) for the purchasing of cloud services.
 - **Cloud Carrier:** The party responsible for providing connectivity between cloud provider and cloud consumer. The ISPs can be assumed as cloud carriers.
 - The cloud provider and cloud carrier are in legal agreement (SLA) to assure a certain level of connectivity and network security.

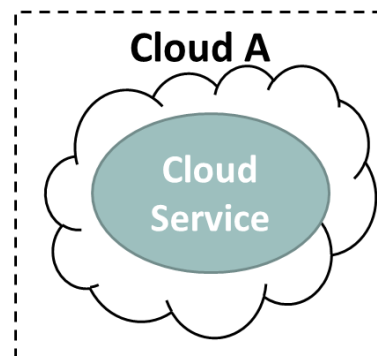
Module No – 052:

- **Organizational boundary:** This is a boundary of ownership and governance of IT assets of an organization.

- Similarly, the **cloud** has its **organizational** boundary.
- ~~This is a boundary of ownership and governance of IT assets of an organization.~~
- ~~Similarly, the cloud has its organizational boundary.~~



Organizational boundary



Organizational boundary

- **Trust boundary:** When an **organization** takes the role of **cloud** consumer, then it has to extend its **trust boundary** to include the **cloud resources**. A **trust boundary** represents a **border** around **trusted IT-resources**.

