UNIT - III

3 Virtualization in Cloud Computing

Syllabus

Introduction : *Definition of Virtualization, Adopting Virtualization, Types of Virtualization, Virtualization Architecture and Software, Virtual Clustering, Virtualization Application, Pitfalls of Virtualization.* ***Grid, Cloud and Virtualization :*** *Virtualization in Grid, Virtualization in Cloud, Virtualization and Cloud Security.* ***Virtualization and Cloud Computing :*** *Anatomy of Cloud Infrastructure, Virtual infrastructures, CPU Virtualization, Network and Storage Virtualization.*

Contents

3.1 Definition of Virtualization

- Virtualization is a broad term that refers to the abstraction of resources across many aspects of computing. For our purposes : One physical machine to support multiple virtual machines that run in parallel. Virtualization is a frame work or methodology of dividing the resources of computer into multiple execution environments.
- Virtualization is an abstraction layer that decouples the physical hardware from the operating system to deliver greater IT resource utilization and flexibility. It allows multiple virtual machines, with heterogeneous operating systems to run in isolation, side-by-side on the same physical machine.
- Virtualization means running multiple machines on a single hardware. The "Real" hardware invisible to operating system. OS only sees an abstracted-out picture. Only Virtual Machine Monitor (VMM) talks to hardware.
- It is "a technique for hiding the physical characteristics of computing resources from the way in which other systems, applications, or end users interact with those resources. This includes making a single physical resource appear to function as multiple logical resources; or it can include making multiple physical resources appear as a single logical resource."
- Fig. 3.1.1 shows concept of virtualization.

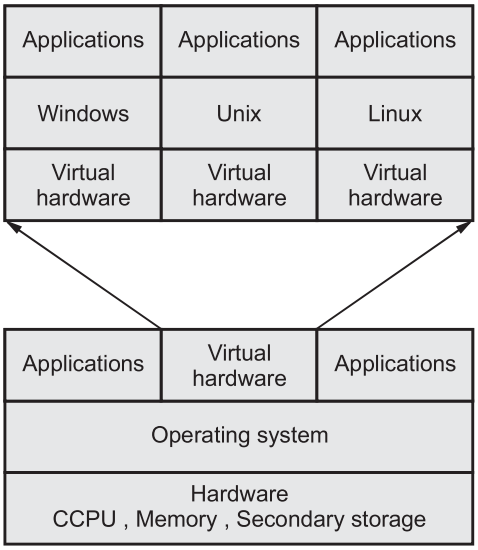


Fig. 3.1.1 Virtual machine

- It is divided into two main categories :
  1. Platform virtualization involves the simulation of virtual machines.

2. Resource virtualization involves the simulation of combined, fragmented, or simplified resources.

- Following are the reasons for using virtualizations :
  a) Virtual machines offer software developers isolated, constrained, test environments.
  b) The most important function of virtualization is the capability of running multiple operating systems and applications on a single computer or server.
  c) Virtualization can usually improve overall application performance due to technology that can balance resources, and provide only what the user needs.
  d) It provides fault and error containment.
  e) It helps in building secured computing platform.
  f) Server virtualization provides a way to implement redundancy without purchasing additional hardware.

3.2 Adopting Virtualization

- Virtualization is changing the IT environment where most companies have been able to adopt it as a measure that allows for better utilization of hardware and reduction of costs.
- Virtualization is particularly valuable to small to medium businesses because it lowers costs for hardware, and reduces systems administration and maintenance costs because fewer servers are in operation.
- Virtualization provides the server administrators a way to segment a large system into smaller sub-systems. The server can be put to use more efficiently to meet the different application needs for different users.
- Virtualization speeds up resource delivery through centralized and automated resource management. You can reduce the cost of a non-production environment through virtualization tools.
- Virtualization offers increased visibility and speed with which you can create a security-focused, non-destructible environment to reduce certain risks.
- Virtualizing the workstation environments can assist you with the simplification of administration, regained control, and access to data.
- Virtualization for enterprises deliver a seamless and standard quality user experience and improvised control and security.

3.3 Types of Virtualizations

SPPU : March-19

- Virtualization is mainly used to emulate execution environment, storage and network. Execution environment classified into two types : process level and system level.
- Fig. 3.3.1 shows taxonomy of virtualization.

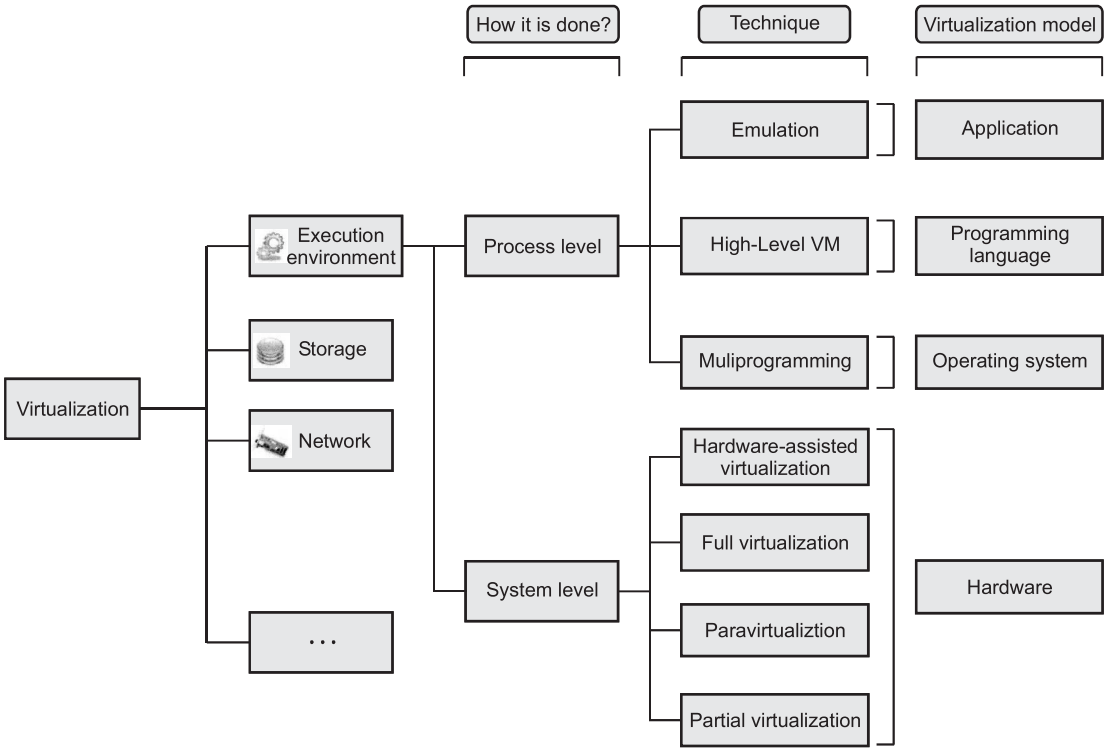


Fig. 3.3.1 Taxonomy of virtualization

- Process level is implemented on top of an existing operating system.
- System level is implemented directly on hardware and do not or minimum requirement of existing operating system.

Platform virtualization

- The creation of a virtual machine using a combination of hardware and software is referred to as platform virtualization. Platform virtualization is performed on a given hardware platform by "host" software, which creates a simulated computer environment for its "guest" software.
- The "guest" software, which is often itself a complete operating system, runs just as if it were installed on a stand-alone hardware platform. Typically, many such virtual machines are simulated on a given physical machine.

- For the "guest" system to function, the simulation must be robust enough to support all the guest system's external interfaces, which may include hardware drivers.

Resource virtualization

- The basic concept of platform virtualization was later extended to the virtualization of specific system resources, such as storage volumes, name spaces, and network resources. Resource aggregation, spanning, or concatenation combines individual components into larger resources or resource pools. For example : RAID and volume managers combine many disks into one large logical disk.
- Virtual Private Network (VPN), Network Address Translation (NAT), and similar networking technologies create a virtualized network namespace within or across network subnets. Multiprocessor and multi-core computer systems often present what appears as a single, fast processor.
- **Application - level virtualization :** It lets you emulate one application level interface on another. Examples include JVM. Another example is WINE that lets you run windows application on Linux or MAC by emulating the Win32 interface.
- **Desktop virtualization :** It supports various computing such as utility, testing, security and development.

3.3.1 Example : Wine

- Wine is a free and open-source compatibility layer that aims to allow application software and computer games developed for Microsoft Windows to run on Unix-like operating systems.
- Wine also provides a software library, named Winelib, against which developers can compile Windows applications to help port them to Unix-like systems.
- Wine is developed with x86 architecture and does not emulate as a processor. Fig. 3.3.2 shows wine, x86 based virtualization architecture.

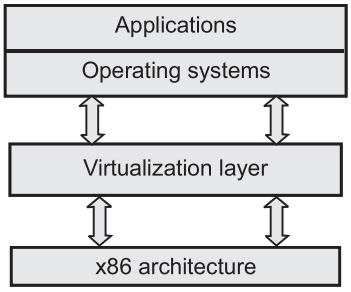


Fig. 3.3.2 wine, x86 based virtualization architecture

- x86 virtualization refers to hardware and software-based mechanisms to support virtualization for processors based on the x86 architecture. Using a hypervisor, it allows several operating systems to be run in parallel on an x86 processor and resources to be distributed in an isolated and efficient manner between the operating systems running in parallel.
  1. FreeBSD : FreeBSD is a UNIX-like operating system. FreeBSD is used by companies, Internet service providers, researchers, computer professionals, students and home users all over the world in their work, education and recreation.
  2. Hive : Hive allows users to read, write, and manage petabytes of data using SQL. Hive is built on top of Apache Hadoop, which is an open-source framework used to efficiently store and process large datasets.
  3. Nemesis : Nemesis is an operating system. Nemesis provides fine-grained guaranteed levels of all system resources including CPU, memory, network bandwidth and disk bandwidth. The OS has been built with the Multimedia in mind, its sole purpose of existence was the delivery and performance of multimedia content in the best way possible.

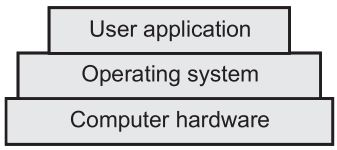
3.3.2 Server Virtualization

- Server virtualization is a software architecture that allows more than one server operating system to run as a guest on a given physical server host. The concept of Server Virtualization widely used in the IT infrastructure to minimizes the costs by increasing the utilization of existing resources.
- The ability to run multiple operating systems on a single physical system and share the underlying hardware resources. Virtual machines provide a layer of abstraction between the OS and the underlying hardware.
- Creating multiple logical server OS instances on one physical piece of hardware. All hardware drivers are virtualized and same virtual hardware regardless of physical hardware.
- Each virtual machine is completely independent of the others and doesn't 'realize' it's virtualized.
- Depending on the approach, server virtualization uses a number of different components. These include :
  1. A host machine, which is the physical server hardware where virtualization occurs.
  2. Virtual machines (VMs), which contain the assets that are abstracted from a traditional server environment.

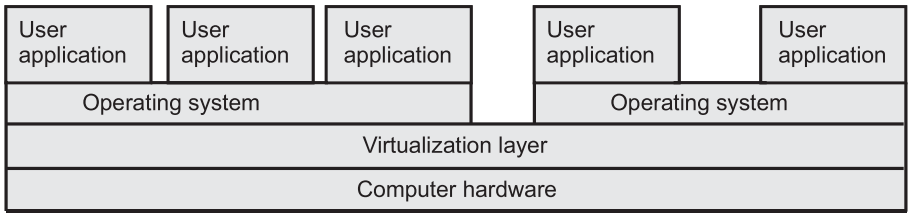
3. A hypervisor, which is a specialized software that creates and maintains virtual machines and can be run natively on bare metal servers or hosted on top of an existing operating system.
4. Hypercalls, which are messages sent between para-virtualized hypervisors and operating systems to share resources using an API.
5. Containers, which are unique user environments that are created in virtualized operating systems. With a container engine, multiple containers can make use of the same interfaces and shared libraries of the underlying host operating system. Containers are often deployed inside of hypervisors or virtual machines to offer an additional layer of isolation from the server's core host operating system.

- Requirements of server virtualization :
  1. **Consolidation :** It is common practice to dedicate each server to a single application. If several applications only use a small amount of processing power, the network administrator can combine several machines into one server running multiple virtual environments.
  2. **Redundancy :** Redundancy refers to running the same application on multiple servers. It's a safety measure, if a server fails for any reason, another server running the same application can take its place.
  3. **Legacy hardware :** Server hardware will eventually become obsolete and switching from one system to another can be difficult. In order to continue offering the services provided by these outdated systems, sometimes called legacy systems a network administrator could create a virtual version of the hardware on modern servers.
  4. **Migration :** Migration refers to moving a server environment from one place to another. With the right hardware and software, it's possible to move a virtual server from one physical machine in a network to another.
- Virtualization allows multiple operating system instances to run concurrently on a single computer; it is a means of separating hardware from a single operating system. Each "guest" OS is managed by a Virtual Machine Monitor (VMM), also known as a hypervisor.
- Because the virtualization system sits between the guest and the hardware, it can control the guests' use of CPU, memory and storage, even allowing a guest OS to migrate from one machine to another.
- By using specially designed software, an administrator can convert one physical server into multiple virtual machines. Each virtual server acts like a unique physical device, capable of running its own operating system.

- Fig. 3.3.3 shows server virtualization after and before.



(a) Before server virtualization



(b) After server virtualization

Fig. 3.3.3

- In server virtualization, the virtual servers are dedicated only to a particular task for their betterment in performance. Every virtual server performs like a distinctive physical device, that is capable of running its own operating system.
- Server virtualization is a cost-effective method that allows using resources efficiently and provides web hosting services effectively utilizing existing resources of IT infrastructure.
- By having each physical server divided into multiple virtual servers, server virtualization allows each virtual server to act as a unique physical device. Each virtual server can run its own applications and operating system. This process increases the utilization of resources by making each virtual server act as a physical server and increases the capacity of each physical machine.
- Types of server virtualization :
  1. Full virtualization
  2. Para-virtualization
  3. OS level virtualization
- Benefits of server virtualization
  1. Lower costs
  2. Consolidation
  3. Practice of redundancy

- Disadvantages of server virtualization
  1. Increase complexity of IT environment
  2. Physical failures become "serious"
  3. Bleed-over

3.3.3 Operating Level Virtualization

- Operating-system-level virtualization is a server-virtualization method where the kernel of an operating system allows for multiple isolated user-space instances, instead of just one. Such instances, which are sometimes called containers and software containers.
- This refers to an abstraction layer between traditional OS and user applications.
- This type of virtualization creates isolated containers on a single physical server and the OS instances to utilize the hard-ware and software in data centers.
- Containers behave like real servers. With containers you can create a portable, consistent operating environment for development, testing, and deployment.
- This virtualization creates virtual hosting environments to allocates hardware resources among a large number of mutually distrusting users.
- Operating-system-level virtualization usually imposes little to no overhead, because programs in virtual partitions use the operating system's normal system call interface and do not need to be subjected to emulation or be run in an intermediate virtual machine.
- Operating system-level virtualization is not as flexible as other virtualization approaches since it cannot host a guest operating system different from the host one, or a different guest kernel.
- Instead of trying to run an entire guest OS, container virtualization isolates the guests, but doesn't try to virtualize the hardware. Instead, you have containers for each virtual environment.
- With container-based technologies, you'll need a patched kernel and user tools to run the virtual environments. The kernel provides process isolation and performs resource management.

Why operating system level virtualization is required ?

- Operating system level virtualization provides feasible solution for hardware level virtualization issue. It inserts a virtualization layer inside an operating system to partition a machine's physical resources.

- It enables multiple isolated VMs within a single operating system kernel. This kind of VM is often called a virtual execution environment (VE), Virtual Private System (VPS), or simply container.
- From the user's point of view, virtual execution environment look like real servers.
- This means a virtual execution environment has its own set of processes, file system, user accounts, network interfaces with IP addresses, routing tables, firewall rules etc.
- Although VEs can be customized for different people, they share the same operating system kernel. Therefore, OS-level virtualization is also called single-OS image virtualization.

Challenges to cloud computing in OS level virtualization ?

- Cloud computing is transforming the computing landscape by shifting the hardware and staffing costs of managing a computational center to third parties.
- Cloud computing has at least two challenges :
  1. The ability to use a variable number of physical machines and virtual machine instances depending on the needs of a problem. For example, a task may need only a single CPU during some phases of execution but may need hundreds of CPUs at other times.
  2. It is related to slow operation of instantiating new virtual machine. Currently, new virtual machines originate either as fresh boots or as replicates of a template VM, unaware of the current application state. Therefore, to better support cloud computing, a large amount of research and development should be done.

Advantages of OS virtualization :

1. OS virtualization provide least overhead among all types of virtualization solution.
2. They offer highest performance and highest density of virtual environment.
3. Low resource requirements.
4. High Scalability.

Disadvantage of OS virtualization :

1. They support only one operating system as base and guest OS in a single server.
2. It supports library level virtualization.

3.3.4 Para-Virtualization

- Paravirtualization is a type of virtualization in which a guest operating system (OS) is recompiled, installed inside a virtual machine (VM), and operated on top of a hypervisor program running on the host OS.
- Para-virtualization refers to communication between the guest OS and the hypervisor to improve performance and efficiency.
- Para-virtualization involves modifying the OS kernel to replace non-virtualizable instructions with hyper-calls that communicate directly with the virtualization layer hypervisor.
- The hypervisor also provides hyper-call interfaces for other critical kernel operations such as memory management, interrupt handling and time keeping.
- Fig. 3.3.4 shows para-virtualization architecture.

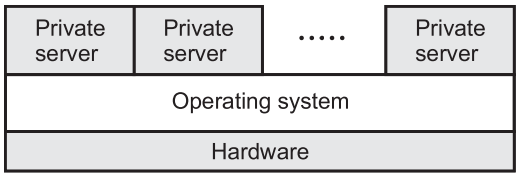


Fig. 3.3.4 Para-virtualization architecture

- In Para-virtualization, the virtual machine does not necessarily simulate hardware, but instead offers a special API that can only be used by modifying the "guest" OS. This system call to the hypervisor is called a "hypercall" in Xen.
- Xen is an open source para-virtualization solution that requires modifications to the guest operating systems but achieves near native performance by collaborating with the hypervisor.
- Microsoft Virtual PC is a para-virtualization virtual machine approach. User-mode Linux (UML) is another para-virtualization solution that is open source.
- Each guest operating system executes as a process of the host operating system. Cooperative Linux, is a virtualization solution that allows two operating systems to cooperatively share the underlying hardware.
- Linux-V server is an operating system-level virtualization solution for GNU/Linux systems with secure isolation of independent guest servers.

- The Linux KVM is virtualization technology that has been integrated into the mainline Linux kernel . Runs as a single kernel loadable module, a Linux kernel running on virtualization-capable hardware is able to act as a hypervisor and support unmodified Linux and Windows guest operating systems.
- Para-virtualization shares the process with the guest operating system.

Problems with para-virtualization

1. Para-virtualized systems won't run on native hardware
2. There are many different para-virtualization systems that use different commands, etc.

- The main difference between full virtualization and paravirtualization in Cloud is that full virtualization allows multiple guest operating systems to execute on a host operating system independently while paravirtualization allows multiple guest operating systems to run on host operating systems while communicating .

Review Question

1. *Explain different levels of virtualization implementation with neat diagram. Also give example of each.* **SPPU : March-19, In Sem, Marks 5**

3.4 Full Virtualization

SPPU : March-19, June-19

- Full Virtualization doesn't need to modify the host OS; it relies upon binary translation to trap and to virtualize certain sensitive instructions.
- Fig. 3.4.1 shows full virtualization.
- VMware Workstation applies full virtualization, which uses binary translation to automatically modify x86 software on-the-fly to replace critical instructions.
- Normal instructions can run directly on the host OS. This is done to increase the performance overhead - normal instructions are carried out in the normal manner, but the difficult and precise executions are first discovered using a trap and executed in a virtual manner.

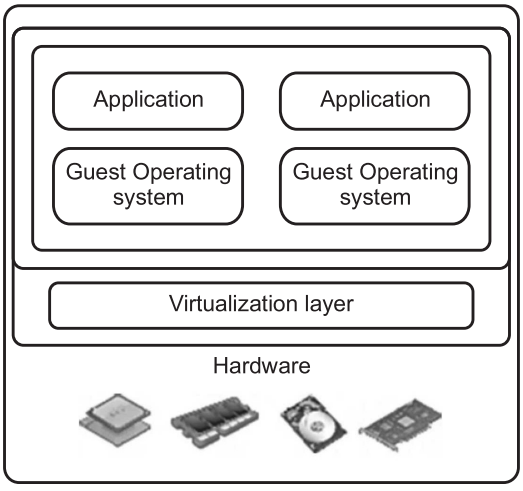


Fig. 3.4.1 Full virtualization

- This is done to improve the security of the system and also to increase the performance.

Host based virtualization :

- Virtualization implemented in a host computer rather than in a storage subsystem or storage appliance.
- Virtualization can be implemented either in host computers, in storage subsystems or storage appliances, or in specific virtualization appliances in the storage interconnect fabric.
- The guest OS are installed and run on top of the virtualization layer. Dedicated applications may run on the VMs. Certainly, some other applications can also run with the host OS directly.
- **Advantages** of host-based architecture :
  1. The user can install this VM architecture without modifying the host OS.
  2. The host-based approach appeals to many host machine configurations.

3.4.1 Memory Virtualization

- Memory virtualization features allow abstraction isolation and monitoring of memory on a per Virtual Machine (VM) basis. These features may also make live migration of VMs possible, add to fault tolerance, and enhance security.
- Example features include Direct Memory Access (DMA) remapping and Extended Page Tables (EPT), including their extensions: accessed and dirty bits, and fast switching of EPT contexts.
- The VMkernel manages all machine memory. The VMkernel dedicates part of this managed machine memory for its own use. The rest is available for use by virtual machines.
- Virtual machines use machine memory for two purposes : each virtual machine requires its own memory and the VMM requires some memory and a dynamic overhead memory for its code and data.
- The virtual memory space is divided into blocks, typically 4KB, called pages. The physical memory is also divided into blocks, also typically 4KB.
- When physical memory is full, the data for virtual pages that are not present in physical memory are stored on disk. ESX/ESXi also provides support for large pages.
- The VMM is responsible for mapping the guest physical memory to the actual machine memory.
- Each page table of a guest OS has a page table allocated for it in the VMM. The page table in the VMM which handles all these is called a shadow page table.

- As it can be seen all this process is nested and inter-connected at different levels through the concerned address.
- If any change occurs in the virtual memory page table or TLB, the shadow page table in the VMM is updated accordingly.

3.4.2 I/O Virtualization

- I/O Virtualization involves managing of the routing of I/O requests between virtual devices and shared physical hardware.
- The there are three ways to implement this are full device emulation, para-VZ and direct I/O.
- I/O virtualization features facilitate offloading of multi-core packet processing to network adapters as well as direct assignment of virtual machines to virtual functions, including disk I/O.
- Examples include Virtual Machine Device Queues (VMDQ), Single Root I/O Virtualization.
- Fig. 3.4.2 shows I/O virtualization.

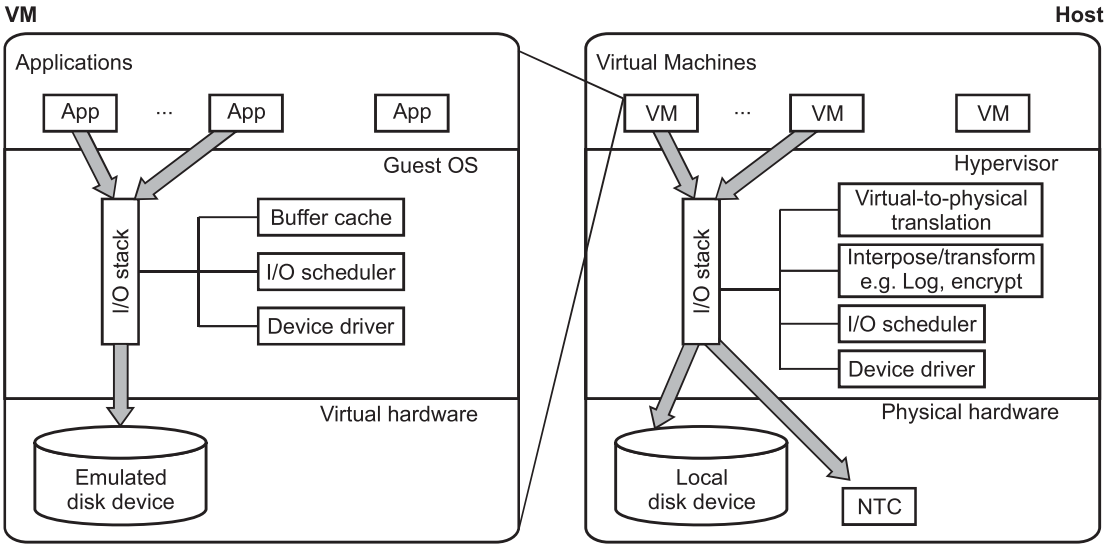


Fig. 3.4.2 I/O virtualization

1. **Full device emulation :** This process emulates well-known and real-world devices. All the functions of a device or bus infrastructure such as device enumeration, identification, interrupts etc. are replicated in the software, which itself is located in the VMM and acts as a virtual device. The I/O requests are trapped in the VMM accordingly.

2. **Para-virtualization :** This method of I/O VZ is taken up since software emulation runs slower than the hardware it emulates. In para-VZ, the frontend driver runs in Domain-U; it manages the requests of the guest OS. The backend driver runs in Domain-0 and is responsible for managing the real I/O devices. This methodology (para) gives more performance but has a higher CPU overhead.
3. **Direct I/O virtualization :** This lets the VM access devices directly; achieves high performance with lower costs. Currently, it is used only for the mainframes.

3.4.3 Difference between Full and Para Virtualization

Sr. No.	Full Virtualization	Para Virtualization
1.	Full Virtualization relies upon binary translation to trap and to virtualize certain sensitive instructions. Example : VMware	Para-Virtualization refers to communication between the guest OS and the hypervisor to improve performance and efficiency. Example : Xen architecture
2.	Full Virtualization doesn't need to modify the host OS.	Para-Virtualization involves modification of OS kernel.
3.	Normal instructions can run directly on the host OS.	Para-virtualized systems won't run on native hardware.
4.	Full Virtualization uses binary translation and direct execution.	Para-Virtualization uses hyper - calls.
5.	Performance is good.	Performance is better in certain cases.
6.	Guest software does not require any modification since the undelying hardware is fully simulated.	Hardware is not simulated and the guest software run their own isolated domains.

Review Questions

1. *Explain full and para virtualization with examples.* **SPPU : March-19, In Sem, Marks 5**
2. *Explain the following : i) CPU virtualization ii) Memory virtualization.* **SPPU : June-19, End Sem, Marks 5**

3.5 Storage Virtualization

- Storage virtualization today refers to a wide variety of products and technologies, from the simplest file systems all the way up-to the cutting edge storage abstraction layers that are capable of managing terabytes of heterogeneous storage spread across the world under a single coherent management framework.

- Storage virtualization refers to the abstraction of storage systems from applications or computers. It is a foundation for the implementation of other technologies, such as thin provisioning and data protection, which are transparent to the server.
- Storage virtualization provides the ability to pool storage systems into a consolidated, shared capacity that can be managed from a central point of control.
- Example of storage virtualizations are host-based volume management, LUN creation, tape storage virtualization and disk addressing.
- Storage virtualization has the following characteristics :
  1. The availability of logical volumes separate from physical hard disk constraints
  2. The capability of abstracting multivendor storage devices into one group and reallocating storage space independently of size or physical location
  3. The capability of having automated storage optimization and management.
- Fig. 3.5.1 shows virtualized storage environment.

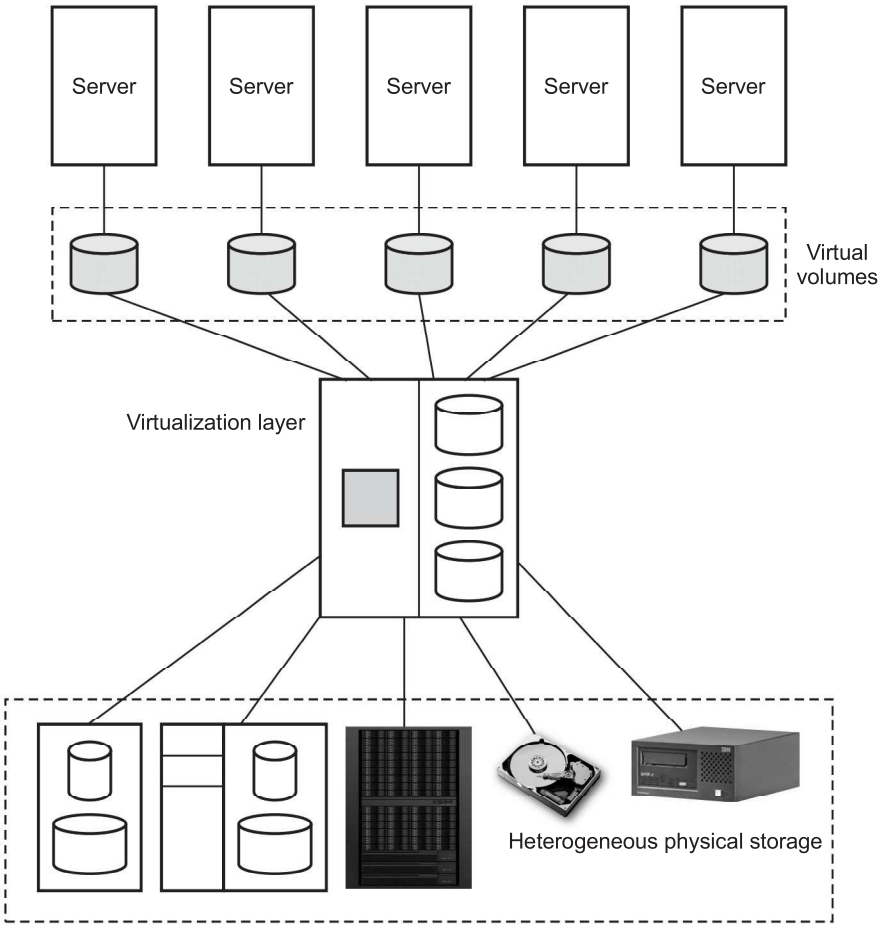


Fig. 3.5.1 Storage virtualization

- Top level servers assigned one virtual volume, which is currently in use by an application. These virtual volumes are mapped to the actual storage in the arrays. When an I/O is sent to a virtual volume, it is redirected through the virtualization at the storage network layer to the mapped physical array
- Primary types of storage virtualizations are block level virtualization and file virtualization.
- **Block level virtualization :** It separates physical and logical storage. File virtualization optimes use of server and storage consolidation.
- **Block-based :** Block-based storage virtualization is the most common type of storage virtualization being practiced across organizations. It identifies all available blocks on individual media/path irrespective of location or vendor, and then the engine leaves that data in the physical position and maps the address to a virtual storage device.
- **File-based :** File-level virtualization works over NAS devices. It has a challenge of its own because managing different NAS devices can be tedious work. Managing multiple appliances is time-consuming and costly. NAS devices require individual management, and users need to know the physical pathname to access a file. Migration of data from old to new NAS devices also remains a challenge as it results in downtime, leading to additional cost to the company.
- Currently there are three methods of storage virtualization :
  1. Server-based virtualization : This method places a management program on the host system and has the benefit of leveraging the SAN asset as it is.
  2. Fabric-based virtualization : This can be done via network switches or appliance servers. In both instances, independent appliances, such as switches, routers, and dedicated servers, are placed between servers and storage and have a storage virtualization function. The purpose behind this is to reduce the impact on the existing SAN and servers.
  3. Storage array-based virtualization : This is a virtualization implemented at the storage-system level.
- Benefits to storage virtualization :
  1. Data is stored in more convenient locations away from the specific host.
  2. The storage devices are able to perform advanced functions like de-duplication, replication, thin provisioning and disaster recovery functionality.
  3. By abstracting the storage level, IT operations can become more flexible in how storage is partitioned, provided and protected.
  4. Improved physical resource utilization.
  5. Lower total cost of ownership : Virtualized storage allows more to be done with the same or less storage.

3.5.1 Network Virtualization

- Network virtualization refers to the technology that enables partitioning or aggregating a collection of network resources and presenting them to various users in a way that each user experiences an isolated and unique view of the physical network.
- Network virtualization creates virtual networks whereby each application sees its own logical network independent of the physical network.
- A virtual LAN (VLAN) is an example of network virtualization that provides an easy, flexible, and less expensive way to manage networks.
- VLANs make large networks more manageable by enabling a centralized configuration of devices located in physically diverse locations.
- Fig. 3.5.2 shows network virtualization.
- Consider a company in which the users of a department are separated over a metropolitan area with their resources centrally located at one office.
- In a typical network, each location has its own network connected to the others through routers. When network packets cross routers, latency influences network performance.
- With VLANs, users with similar access requirements can be grouped together into the same virtual network. This setup eliminates the need for network routing.
- As a result, although users are physically located at disparate locations, they appear to be at the same location accessing resources locally.

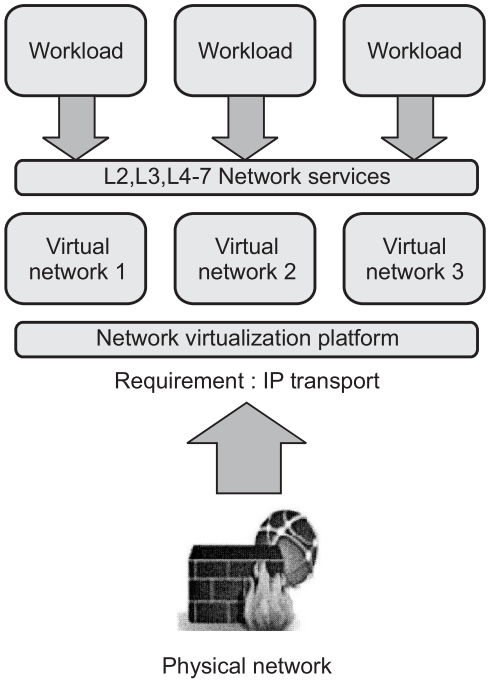


Fig. 3.5.2 Network virtualization

- In addition to improving network performance, VLANs also provide enhanced security by isolating sensitive data from the other networks and by restricting access to the resources located within the networks.
- Network virtualization decouples the roles of the traditional Internet service providers (ISPs) into infrastructure providers (InPs) and service providers (SPs).

- Benefits :
  1. Reduces the number of physical devices needed.
  2. Easily segment networks.
  3. Permits rapid change / scalability and agile deployment.
  4. Security from destruction of physical devices.

3.6 Virtual Clustering

SPPU : March-20

- A computer cluster is a set of connected computers (nodes) that work together as if they are a single machine. All processor machines share resources such as a common home directory and have a software such as a Message Passing Interface (MPI) implementation installed to allow programs to be run across all nodes simultaneously.
- Computer clusters are often used for cost-effective High Performance Computing (HPC) and High Availability (HA) by businesses of all sizes. A computer cluster help to solve complex operations more efficiently with much faster processing speed, better data integrity than a single computer and they only used for mission-critical applications.

Characteristics Virtual Cluster :

1. Virtual machine or physical machine is used as virtual cluster nodes. Multiple VM running with different types of OS can be deployed on the same physical node.
2. Virtual machine runs with guest operating system. Host OS and VM OS are different but it manages the resources in the physical machine.
3. Virtual machine can be replicated in multiple servers and it support distributed parallelism, fault tolerance and disaster recovery.
4. Number of nodes of a virtual cluster may change accordingly.
5. It virtual machine failes, it can not affect the host machine.

- **Virtual cluster is managed by four ways :**
  1. We can use a guest-based manager, by which the cluster manager resides inside a guest OS. Ex. : A Linux cluster can run different guest operating systems on top of the Xen hypervisor.
  2. We can bring out a host-based manager which itself is a cluster manager on the host systems. Ex. : VMware HA (High Availability) system that can restart a guest system after failure.
  3. An independent cluster manager, which can be used on both the host and the guest - making the infrastructure complex.

4. Finally, we might also use an integrated cluster (manager), on the guest and host operating systems; here the manager must clearly distinguish between physical and virtual resources.

Review Question

1. *Explain in brief virtual clusters and resource management.*

SPPU : March-20, In Sem, Marks 5

3.7 Virtualization Application

- Virtualization at the application level virtualizes an application as a VM. On a traditional OS, an application often runs as a process. Therefore, application-level virtualization is also known as process-level virtualization.
- A fully virtualized application is not installed in the traditional sense, although it is still executed as if it were. The application behaves at runtime like it is directly interfacing with the original operating system and all the resources managed by it, but can be isolated to varying degrees.
- Full application virtualization requires a virtualization layer. Application virtualization layers replace part of the runtime environment normally provided by the operating system.
- The layer intercepts all disk operations of virtualized applications and transparently redirects them to a virtualized location, often a single file.
- The application remains unaware that it accesses a virtual resource instead of a physical one. Since the application is now working with one file instead of many files spread throughout the system, it becomes easy to run the application on a different computer and previously incompatible applications can be run side-by-side.
- The most popular approach is to deploy High Level Language (HLL) VMs. Here the virtualization layer sits as an application program on top of the operating system, and the layer exports an abstraction of a VM that can run programs written and compiled to a particular abstract machine definition. Any program written in the HLL and compiled for this VM will be able to run on it.
- **Benefits :**
  1. Application virtualization uses fewer resources than a separate virtual machine.
  2. Application virtualization also enables simplified operating system migrations.
  3. Applications can be transferred to removable media or between computers without the need of installing them, becoming portable software.

- **Limitations :**
  1. Not all computer programs can be virtualized
  2. Lower performance

3.8 Pitfalls of Virtualization

a) Pros

1. Data center and energy-efficiency savings : As companies reduce the size of their hardware and server footprint, they lower their energy consumption.
2. Operational expenditure savings : Once servers are virtualized, your IT staff can greatly reduce the ongoing administration and management of manual work.
3. Reduced costs : It reduced cost of IT infrastructure.
4. Data does not leak across virtual machine.
5. Virtual machine is completed isolated from host machine and other virtual machine.
6. Simplifies resource management by pooling and sharing resources.
7. Significantly reduce downtime.
8. Improved performance of IT resources.

b) Cons

1. Not all hardware or software can be virtualized.
2. Not all servers are applications are specifically designed to be virtualization-friendly.

3.9 Grid, Cloud and Virtualization

3.9.1 Virtualization in Grid

- A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational facilities.
- Grid computing is a distributed computing system where a group of computers are connected to create and work as one large virtual computing power, storage, database, application and service.
- Grid computing can be used in a variety of ways to address various kinds of application requirements. The three primary types of grids are given below.
  1. Computational grid : A computational grid is focused on setting aside resources specifically for computing power. In this type of grid, most of the machines are high-performance servers.

2. Scavenging grid : A scavenging grid is most commonly used with large numbers of desktop machines. Machines are scavenged for available CPU cycles and other resources. Owners of the desktop machines are usually given control over when their resources are available to participate in the grid.
3. Data grid : A data grid is responsible for housing and providing access to data across multiple organizations. Users are not concerned with where this data is located as long as they have access to the data. For example, you may have two universities doing life science research, each with unique data. A data grid would allow them to share their data, manage the data and manage security issues such as who has access to what data.

3.9.2 Virtualization in Cloud

1. The cloud computing adoption model

- Cloud Adoption is a strategic move by organizations of reducing cost, mitigating risk and achieving scalability of data base capabilities. Fig. 3.9.1 shows cloud computing adoption model.

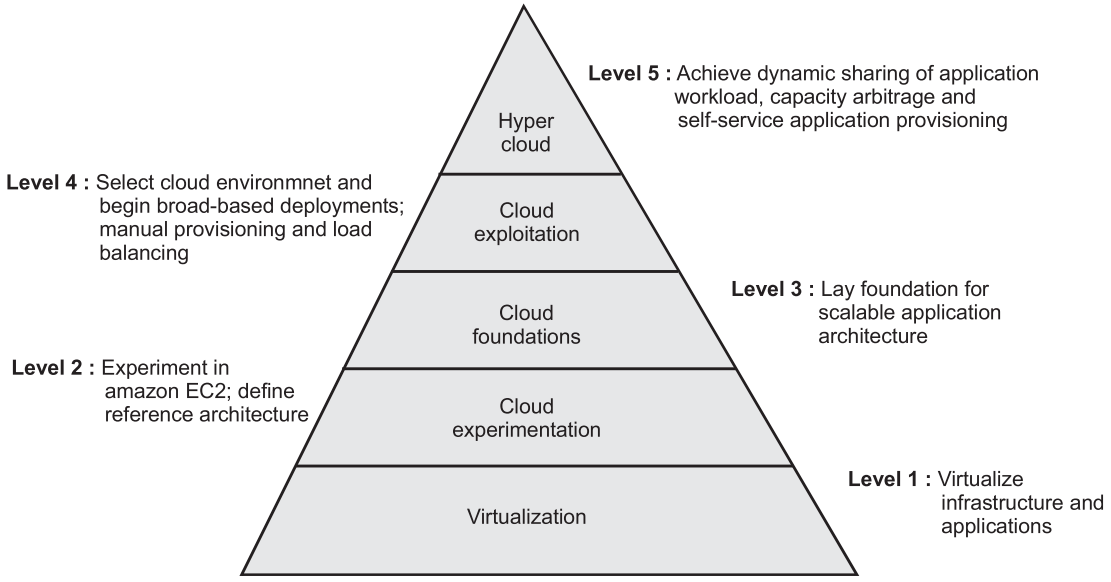


Fig. 3.9.1 Cloud computing adoption model

- Cloud adoption model consists of following layers :
  a) Hyper cloud : It provides dynamic sharing and self-service application.
  b) Cloud foundations : It performs load balancing, deployments etc.

c) Cloud exploitation : Where foundations for scalable application architecture are carried out.

d) Cloud experimentation : Various architectures are experimented.

e) Virtualization : Infrastructure and applications are virtualized.

3.9.3 Difference between Cloud and Virtualization

Sr. No.	Virtualization	Cloud Computing
1.	Virtualization is the process of creating a virtual environment on an existing server to run your desired program, without interfering with any of the other services provided by the server or host platform to other users.	Cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive.
2.	Location of virtual machine is on a specific host.	Location of virtual machine is on any host.
3.	Instance storage is persistent.	Instance storage is shortly lived.
4.	Virtualization uses customizable VM resource like CPU and RAM.	Cloud computing uses standard VM resource like CPU and RAM
5.	Recovery from failures: attempt to recover failed VM.	Recovery from failures : Discard instance spin up new one.

3.9.4 Virtualization and Cloud Security

- Cloud computing security challenges fall into three broad categories :
  1. Data protection : Securing your data both at rest and in transit.
  2. User authentication : Limiting access to data and monitoring who accesses the data.
  3. Disaster and data breach : Contingency planning.
- Data protection : Data needs to be encrypted at all times, with clearly defined roles when it comes to who will be managing the encryption keys.
- User authentication : Data resting in the cloud needs to be accessible only by those authorized to do so, making it critical to both restrict and monitor who will be accessing the company's data through the cloud. In order to ensure the integrity of user authentication, companies need to be able to view data access logs and audit trails to verify that only authorized users are accessing the data.
- Contingency planning : With the cloud serving as a single centralized repository for a company's mission-critical data, the risks of having that data compromised due to a data breach or temporarily made unavailable due to a natural disaster are real concerns.

- If information is encrypted while passing through the cloud, who controls the encryption/decryption keys ? Is it the customer or the cloud vendor ? Most customers probably want their data encrypted both ways across the Internet using Secure Sockets Layer protocol.
- They also most likely want their data encrypted while it is at rest in the cloud vendor's storage pool. Be sure that you, the customer, control the encryption/decryption keys, just as if the data were still resident on your own servers.
- Data integrity means ensuring that data is identically maintained during any operation.
- Cloud-based services will result in many mobile IT users accessing business data and services without traversing the corporate network. This will increase the need for enterprises to place security controls between mobile users and cloud-based services.
- Placing large amounts of sensitive data in a globally accessible cloud leaves organizations open to large distributed threats, attackers no longer have to come onto the premises to steal data, and they can find it all in the one "virtual" location.
- Virtualization efficiencies in the cloud require virtual machines from multiple organizations to be co-located on the same physical resources. Although traditional data center security still applies in the cloud environment, physical segregation and hardware-based security cannot protect against attacks between virtual machines on the same server.
- Operating system and application files are on a shared physical infrastructure in a virtualized cloud environment and require system, file, and activity monitoring to provide confidence and auditable proof to enterprise customers that their resources have not been compromised or tampered with.
- In the cloud computing environment, the enterprise subscribes to cloud computing resources, and the responsibility for patching is the subscriber's rather than the cloud computing vendor's.
- The need for patch maintenance vigilance is imperative. Lack of due diligence in this regard could rapidly make the task unmanageable or impossible, leaving you with "virtual patching" as the only alternative.
- Confidentiality : Confidentiality refers to limiting information access. Sensitive information should be kept secret from individuals who are not authorized to see the information.

- In cloud environments, confidentiality primarily pertains to restricting access to data in transit and storage.
- Integrity can extend to how data is stored, processed, and retrieved by cloud services and cloud-based IT resources.
- Some common cloud security threats include :
  a) Risks of cloud-based infrastructure including incompatible legacy IT frameworks, and third-party data storage service disruptions.
  b) Internal threats due to human error such as misconfiguration of user access controls.
  c) External threats caused almost exclusively by malicious actors, such as malware, phishing, and DDoS attacks.

3.10 Virtualization and Cloud Computing

3.10.1 Anatomy of Cloud Infrastructure

- Cloud anatomy can be simply defined as the structure of the cloud. Anatomy can be considered as a part of architecture. Fig. 3.10.1 shows cloud anatomy.

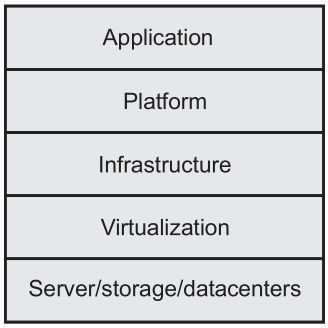


Fig. 3.10.1 Cloud anatomy

1. **Application :** The upper layer is the application layer. In this layer, any applications are executed.
2. **Platform :** This component consists of platforms that are responsible for the execution of the application. This platform is between the infrastructure and the application.
3. **Infrastructure :** The infrastructure consists of resources over which the other components work. This provides computational capability to the user.
4. **Virtualization :** Virtualization is the process of making logical components of resources over the existing physical resources. The logical components are isolated and independent, which form the infrastructure.

5. **Physical hardware :** The physical hardware is provided by server and storage units.

3.10.2 Virtual Infrastructures

- A virtual infrastructure allows you to utilise the IT capabilities of physical resources as software that can be used across multiple platforms. These resources are shared across multiple virtual machines (VMs) and applications for maximum efficiency, creating a virtual infrastructure.
- Virtual Infrastructure consists of the following components :
  a) Bare-metal hypervisors to enable full virtualisation of each x86 computer.
  b) Virtual infrastructure services such as resource management and consolidated backup to optimise available resources among virtual machines.
  c) Automation solutions that provide special capabilities to optimise a particular IT process such as provisioning or disaster recovery.
- Cloud computing provides virtual infrastructures which provide facility for data storage and computing power without direct management by users.

3.10.3 CPU Virtualization

- Certain processors such as Intel VT provide hardware assistance for CPU virtualization.
- When using this assistance, the guest can use a separate mode of execution called guest mode. The guest code, whether application code or privileged code, runs in the guest mode.
- On certain events, the processor exits out of guest mode and enters root mode. The hypervisor executes in the root mode, determines the reason for the exit, takes any required actions, and restarts the guest in guest mode.
- When you use hardware assistance for virtualization, there is no need to translate the code. As a result, system calls or trap-intensive workloads run very close to native speed.
- Some workloads, such as those involving updates to page tables, lead to a large number of exits from guest mode to root mode. Depending on the number of such exits and total time spent in exits, this can slow down execution significantly.
- CPU virtualization features enable faithful abstraction of the full prowess of Intel CPU to a virtual machine.
- All software in the VM can run without any performance, as if it was running natively on a dedicated CPU. Live migration from one Intel CPU generation to another, as well as nested virtualization, is possible.

3.10.4 Network and Storage Virtualization

- Network virtualization refers to the technology that enables partitioning or aggregating a collection of network resources and presenting them to various users in a way that each user experiences an isolated and unique view of the physical network.
- Network virtualization creates virtual networks whereby each application sees its own logical network independent of the physical network.
- A virtual LAN (VLAN) is an example of network virtualization that provides an easy, flexible, and less expensive way to manage networks.
- VLANs make large networks more manageable by enabling a centralized configuration of devices located in physically diverse locations.
- Fig. 3.10.2 shows network virtualization.

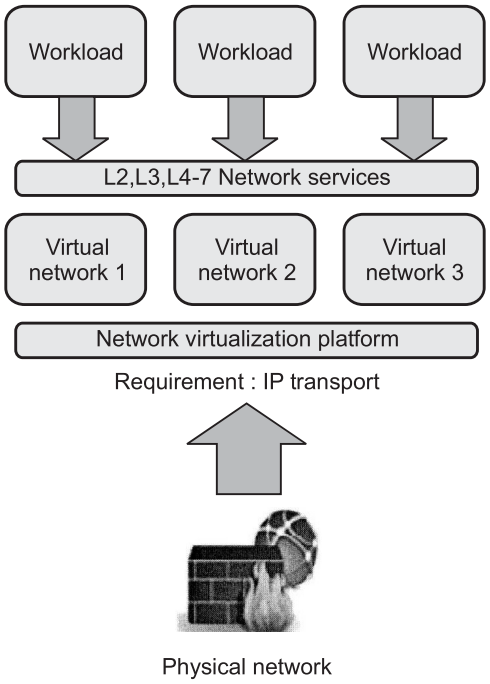


Fig. 3.10.2 Network virtualization

- Consider a company in which the users of a department are separated over a metropolitan area with their resources centrally located at one office.
- In a typical network, each location has its own network connected to the others through routers. When network packets cross routers, latency influences network performance.

- With VLANs, users with similar access requirements can be grouped together into the same virtual network. This setup eliminates the need for network routing.
- As a result, although users are physically located at disparate locations, they appear to be at the same location accessing resources locally.
- In addition to improving network performance, VLANs also provide enhanced security by isolating sensitive data from the other networks and by restricting access to the resources located within the networks.
- Network virtualization decouples the roles of the traditional Internet Service Providers (ISPs) into Infrastructure Providers (InPs) and Service Providers (SPs)
- Benefits :
  1. Reduces the number of physical devices needed.
  2. Easily segment networks.
  3. Permits rapid change/scalability and agile deployment.
  4. Security from destruction of physical devices.

3.11 Multiple Choice Questions

Q.1 Which of the following type of virtualization is also characteristic of cloud computing ?

a Storage b Application
c CPU d All of the mentioned

Q.2 Which of the following network resources can be load balanced ?

a Connection through intelligent switches b DNS
c Storage resources d All of these

Q.3 Each guest OS is managed by a virtual machine monitor also known as _____ .

a server b hypervisor
c storage d none

Q.4 _____ is the process of making logical components of resources over the existing physical resources.

a Virtualization b Cloud computing
c Storage d Loading

Q.5 Which of the following are types of server virtualization ?

a Full virtualization
b Para - virtualization
c OS level virtualization
d All of these

Q.6 Which of the following type of virtualization is also characteristic of cloud computing ?

a Storage
b Application
c CPU
d All of these

Q.7 What is the solution for full virtualization ?

a Processor
b Application
c Desktop
d Hardware

Q.8 The creation of a virtual machine using a combination of hardware and software is referred to as _____ virtualization.

a system
b CPU
c machine
d platform

Q.9 Library-level virtualization is also known as user-level _____ interface.

a software
b user
c application
d application binary

Answer Keys for Multiple Choice Questions :

Q.1	c	**Q.2**	d	**Q.3**	b	**Q.4**	a
Q.5	d	**Q.6**	d	**Q.7**	d	**Q.8**	d
Q.9	d						

❑❑❑

Notes

UNIT - IV

4 Cloud Platforms and Cloud Applications

Syllabus

Amazon Web Services (AWS) : *Amazon Web Services and Components, Amazon Simple DB, Elastic Cloud Computing (EC2), Amazon Storage System, Amazon Database services (Dynamo DB).* ***Microsoft Cloud Services :*** *Azure core concepts, SQL Azure, Windows Azure Platform Appliance.* ***Cloud Computing Applications :*** *Healthcare : ECG Analysis in the Cloud, Biology : Protein Structure Prediction, Geosciences : Satellite Image Processing, Business and Consumer Applications : CRM and ERP, Social Networking, Google Cloud Application : Google App Engine. Overview of OpenStack architecture.*

Contents