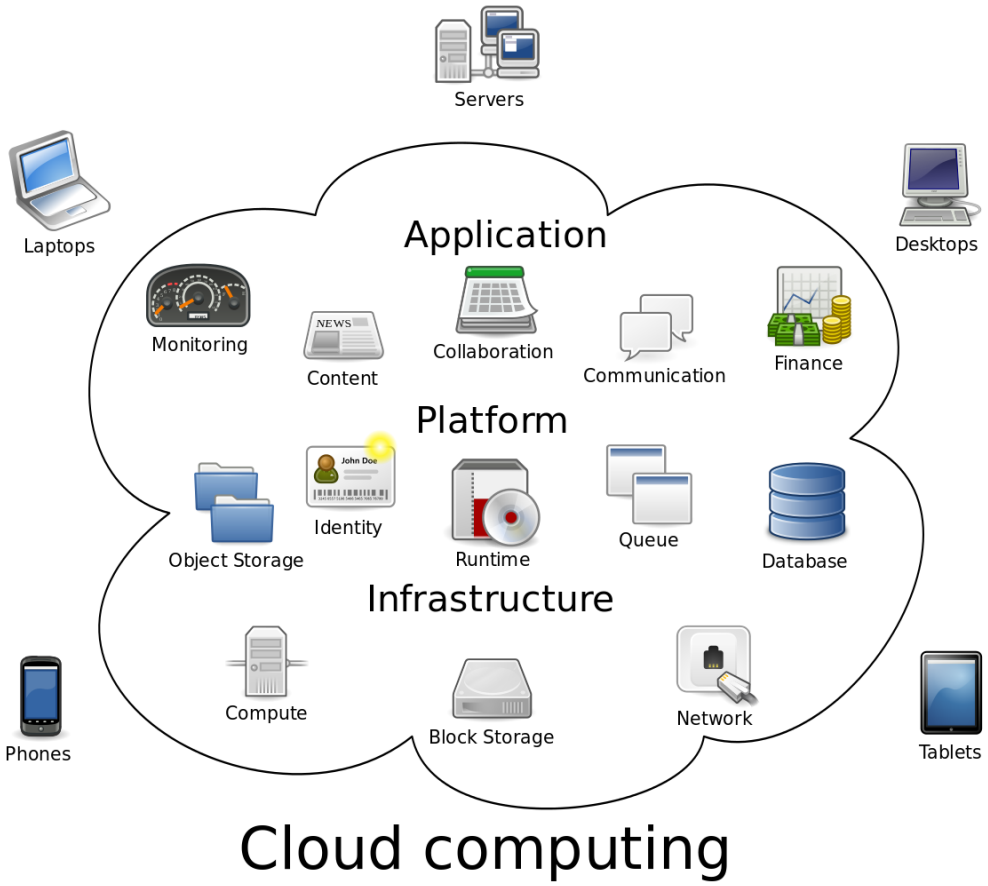
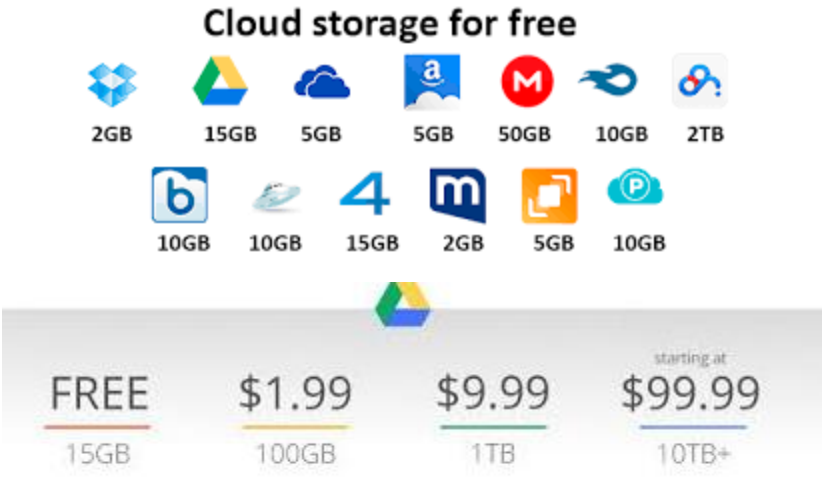
UNIT-1
JOURNEY TO THE CLOUD

Cloud Computing

The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. The cloud metaphor to the internet, in which cloud computing providing computing resources like server, storage, databases, networking, software, analytics etc. over the internet. The companies providing these services are called cloud providers and typically charged based on usage, like how paying for electricity and water bill at home.

We need not install any piece of software on our local PC and this shows how cloud computing overcome platform dependency.





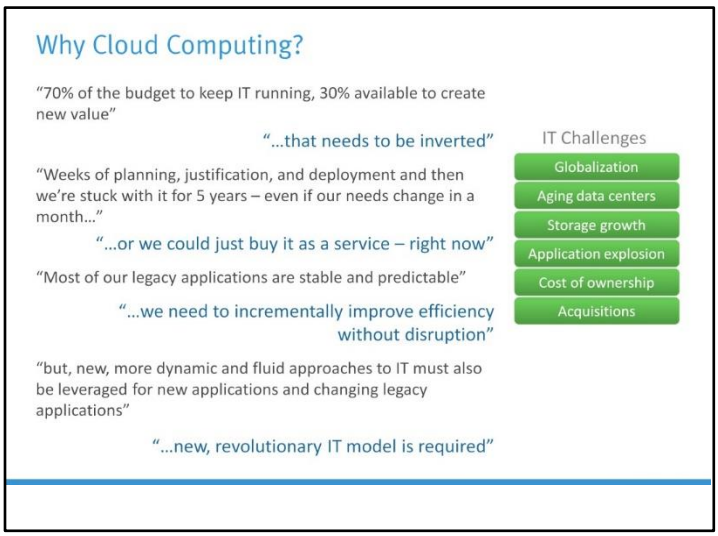
History:

The concept of cloud computing came into existence in 1950 with implementation of mainframe computers, accessible via thin/thick clients. Since then, cloud computing has been evolved from static to dynamic clients from software to services.

Why Cloud computing:

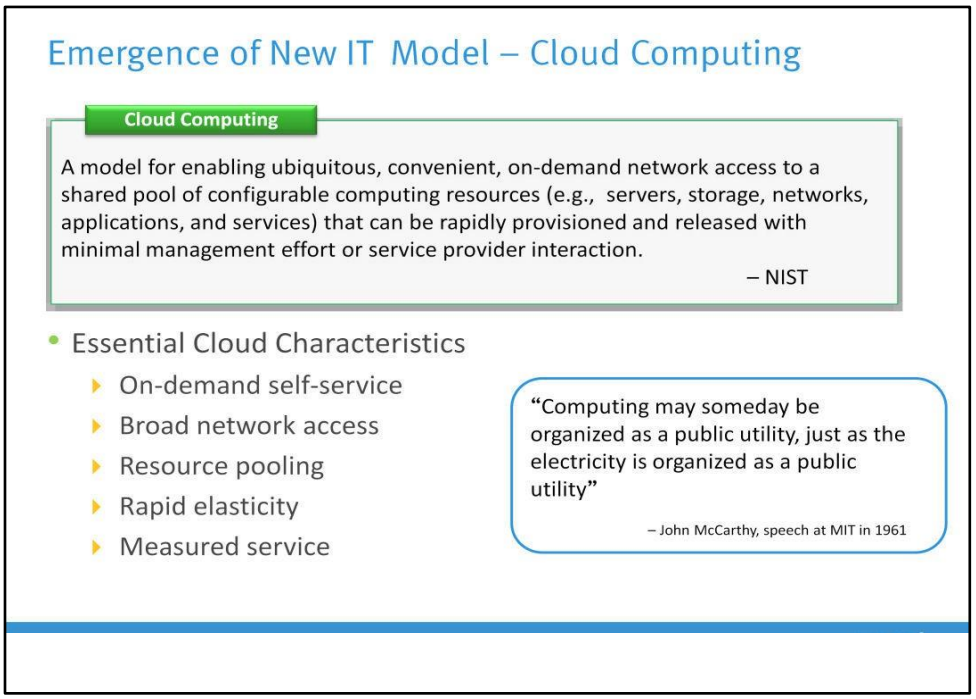
Cloud computing as a service has seen its phenomenal growth in recent years the primary motivation for this growth has been the promise of reduced capital and operating expenses and the easy of dynamically scaling and applying new series without maintaining a dedicated compute infrastructure. Hence cloud computing has to rapidly transform the way organisation view their IT resources.

Cloud computing as a service has seen its phenomenal growth in recent years the primary motivation for this growth has been the promise of reduced capital and operating expenses and the easy of dynamically scaling and applying new series without maintaining a dedicated compute infrastructure. Hence cloud computing has to rapidly transform the way organisation view their IT resources.

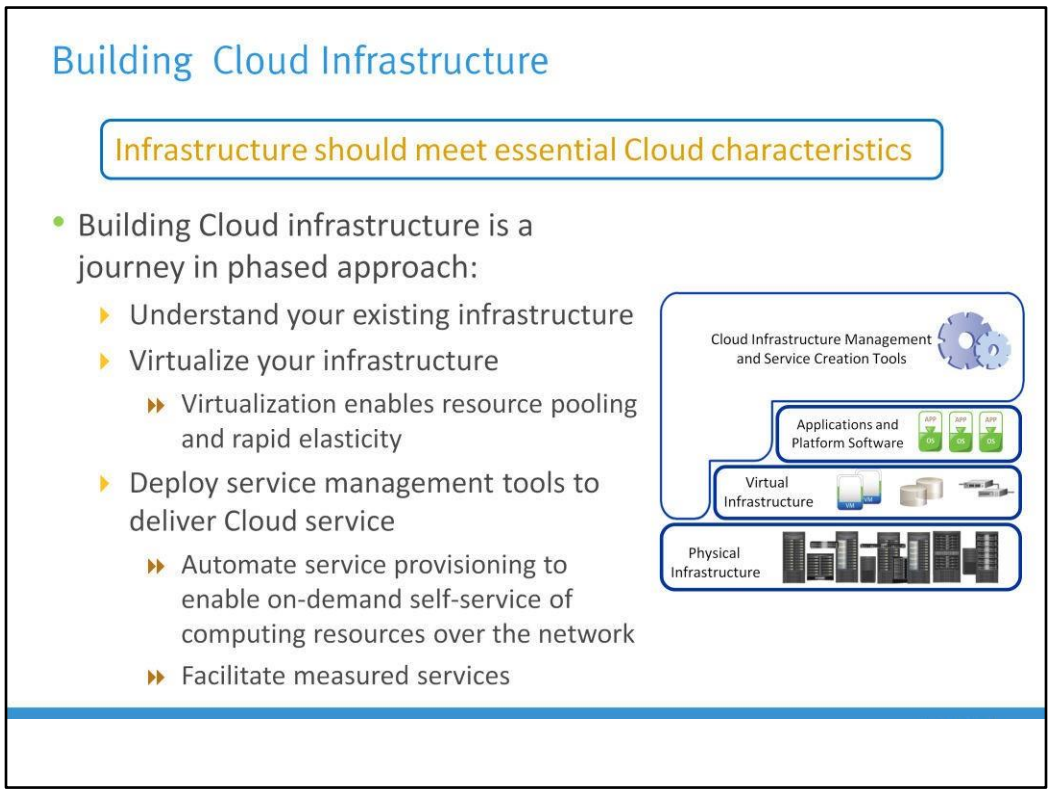


With all of the interest surrounding Cloud, it is helpful to understand what is driving this need to change. Each IT organization has its own unique drivers, but they generally fall into some general categories: cost, availability, time-to-market, etc. There are pressures outside IT from the organization's highest level executives who are looking for more flexibility, doing more with less cost, and using information as a competitive advantage. IT organizations want to respond by transforming IT into something with greater business agility. The IT challenges listed below have made organizations think about the Cloud Computing model to provide better service to their customers:

- **Globalization**: IT must meet the business needs to serve customers world-wide, round the clock – 24x7x365.
- **Aging Data Centers**: Migration, upgrading technology to replace old technology.
- **Storage Growth**: Explosion of storage consumption and usage.
- **Application Explosion**: New applications need to be deployed and their usage may scale rapidly. The current data center infrastructures are not planned to accommodate for such rapid growth.
- **Cost of ownership**: Due to increasing business demand, the cost of buying new equipment's, power, cooling, support, licenses, etc., increases the Total Cost of Ownership (TCO.)
- **Acquisitions**: When companies are acquired, the IT infrastructures of the acquired company and the acquiring company are often different. These differences in the IT infrastructures demand significant effort to make them interoperable.



For organizations to be competitive in today's fast-paced, online, and highly interconnected global economy, they must be agile, flexible, and able to respond rapidly to the changing market conditions. Cloud, a next generation style of computing, provides highly scalable and flexible computing that is available on demand. Cloud Computing empowers self-service requesting through a fully automated request-fulfillment process in the background. Cloud Computing promises real costs savings and agility to organizations. Through Cloud Computing, an organization can rapidly deploy applications where the underlying technology components can scale-up and scale-down, based on the business requirements.



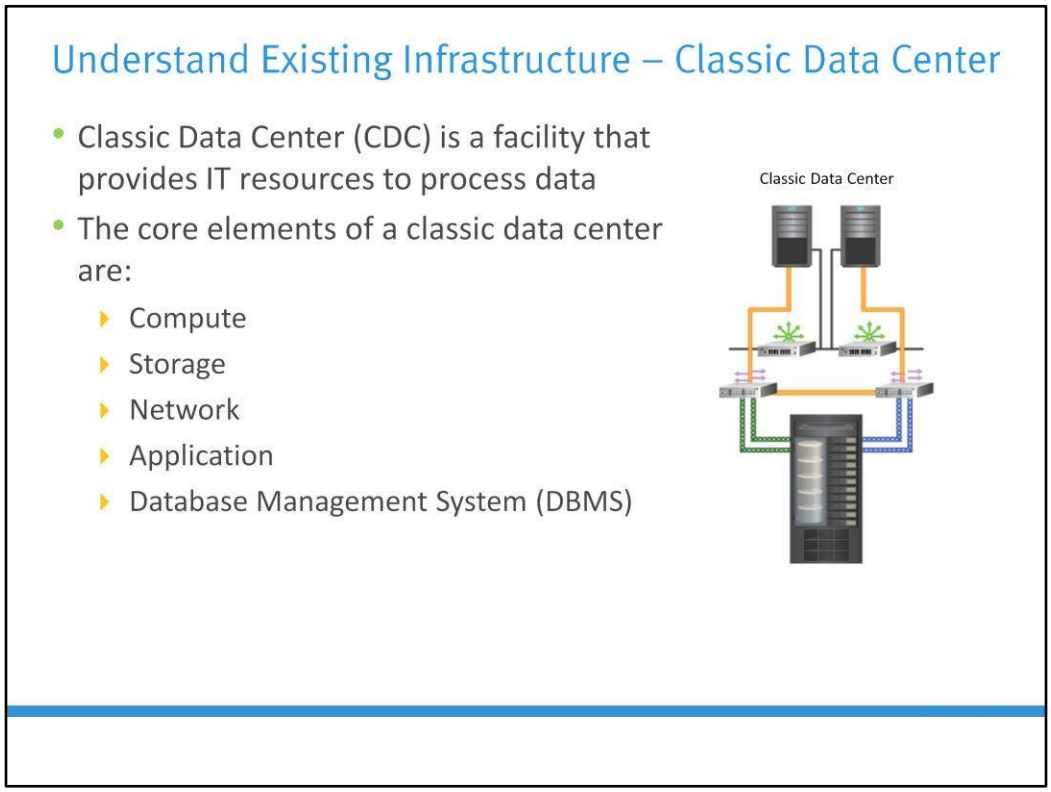
An infrastructure should fulfill the essential characteristics to support Cloud services. It can be built using a shared pool of computing resources, such as compute, storage, and network. The infrastructure should be flexible to meet the rapidly-changing demands of its consumers and allow them to provision resources on-demand over a network. The infrastructure should also enable monitoring , control and optimization of resource usage.

Building a Cloud infrastructure is a phased approach. The journey begins with understanding the existing physical infrastructure, its elements, and processes. The next step is to focus on aggregating the existing infrastructure resources using virtualization technologies. These resource pools facilitate centralized management of resources and enables faster resource provisioning.

The next step is to deploy service management tools that enable automation of

processes and management to minimize human intervention. Service management tools also include measured services which enable consumption based metering. With the service management in place, on-demand provisioning of IT resources become more dynamic and allow IT to be delivered as a service.

Although virtualization is a key step towards building Cloud, it is possible to use highly automated physical infrastructure to provide Cloud services. However, it may not be optimized.

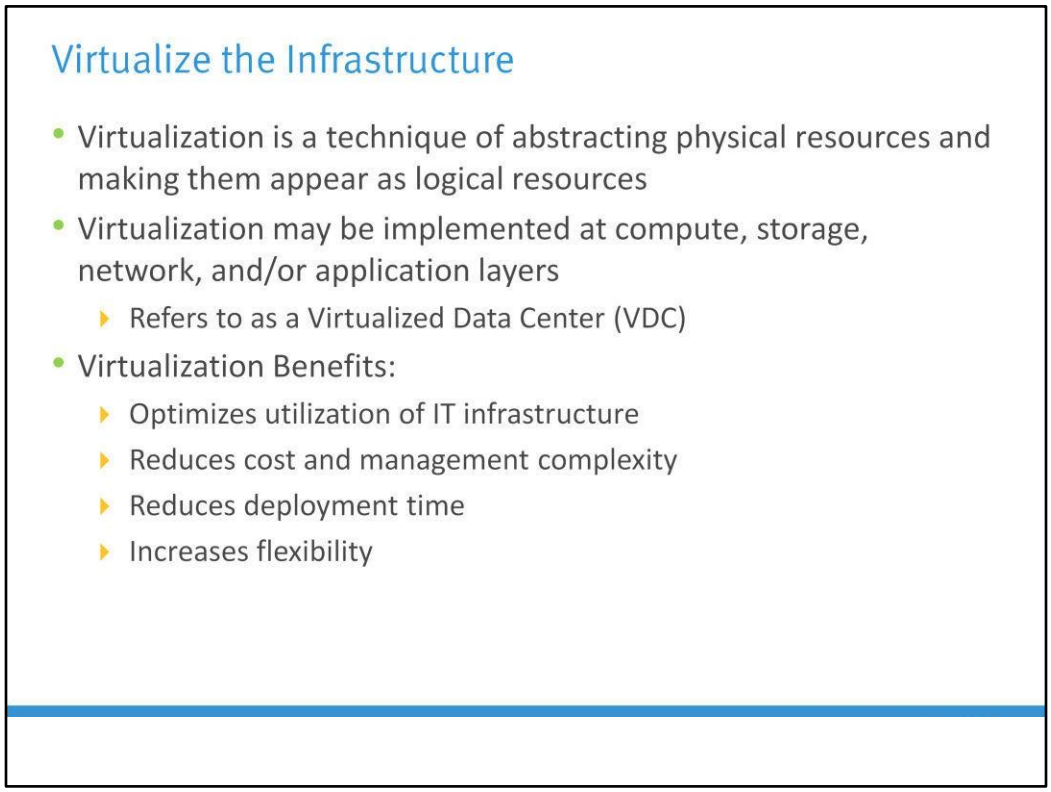


A Classic Data Center (CDC) is a facility that enables IT resources to process data. The core elements of CDC are compute, storage, network, application, and Database Management System (DBMS).

- **Application** is a computer program that provides the logic for computing operations. Applications may use a DBMS, which uses operating system services, to perform store/retrieve operations on storage devices.
- **DBMS** provides a structured way to store data in logically organized tables that are interrelated. A DBMS optimizes the storage and retrieval of data.
- **Compute** is a resource that runs applications with the help of underlying computing components.
- **Storage** is a resource that stores data persistently for subsequent use.
- **Network** is a data path that facilitates communication between compute systems or between compute systems and storage.

These IT resources are typically viewed and managed as separate entities. But all

these elements must work together to address data processing requirements. Other elements of a CDC are power supplies and environmental controls such as air conditioning and fire suppression.

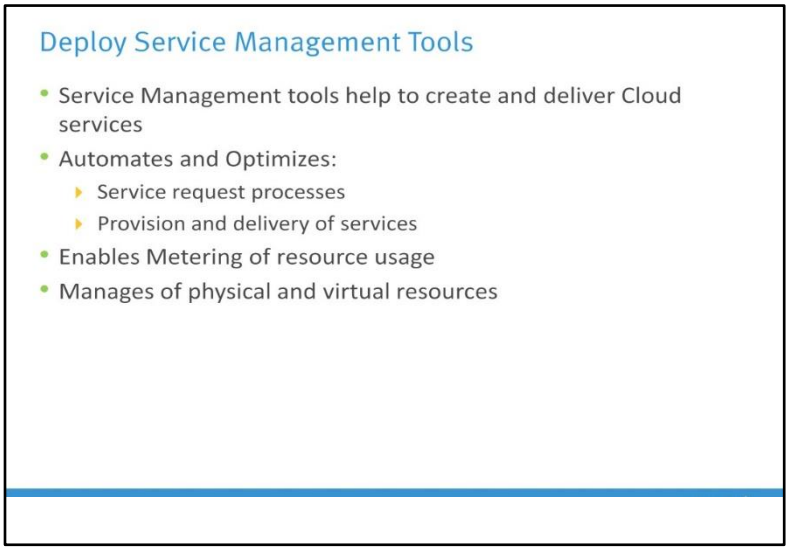


Virtualization abstracts physical resources, such as compute, storage, and network, to function as logical resources. It creates an abstraction layer to hide the physical characteristics of resources from users. For example, in compute system virtualization, a physical machine appears as multiple logical machines (virtual machines), each running an operating system concurrently.

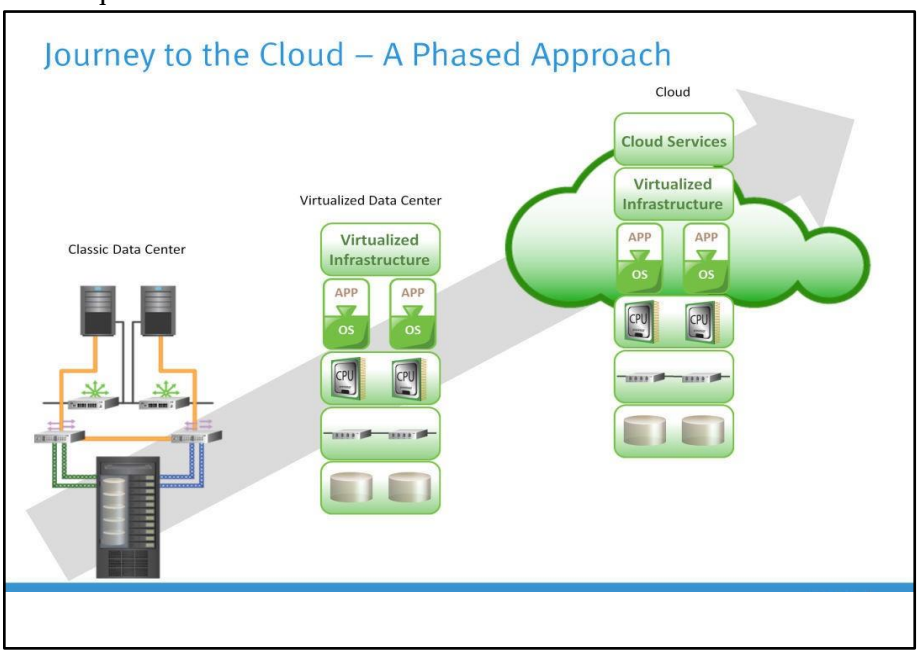
A VDC is a data center in which compute, storage, network, and/or applications are virtualized. Compute virtualization enables running multiple operating systems concurrently on a compute system. This improves compute system utilization. Storage virtualization provides a logical view of storage and presents it to the compute system. In network virtualization, multiple logical networks are created on a physical network. Each of these virtualization technologies is explained in detail in the forthcoming modules.

By consolidating IT resources using virtualization techniques, organizations can optimize their infrastructure utilization. By improving the utilization of IT assets, organizations can reduce the costs associated with purchasing new hardware. They also reduce space and energy costs associated with maintaining the resources. Moreover, less people are required to administer these resources, which further lowers the cost. Virtual resources are created using software that enables faster deployment, compared to deploying physical resources.

Virtualization increases flexibility by allowing to create and reclaim the logical resources based on business requirements.



A service management tool enables creation and optimization of Cloud services to meet business objectives and to provide value to the consumers. The services built are listed in a service catalog that allows consumers to choose the desired services. Service management automates service creation and provisioning without any manual intervention. It also helps the monitoring and metering services in measuring resource usage and chargeback. Service management tools are also responsible for managing both physical and virtual resources that are used to create Cloud services. Examples of management activities are capacity management, configuration management, change management, etc. These management processes enable meeting service assurance and compliance requirements.



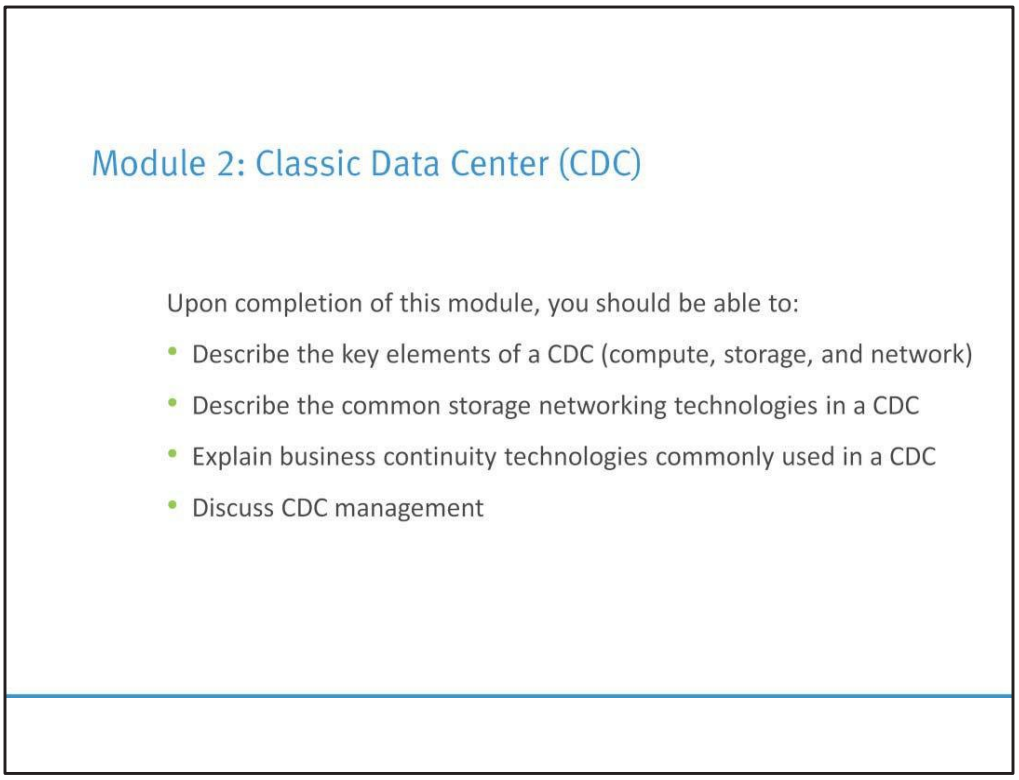
As discussed, Cloud adoption for an organization is a journey. Organizations have to perform various steps to elevate their existing data centers, to provide Cloud services.

Data centers provide centralized digital data-processing capabilities required to support an organization's business. A typical data center includes compute, storage, and network, which enable storing and processing large amounts of data. These data centers are also referred as Classic Data Centers (CDCs). In a Classic data center, resources are typically dedicated for each of the business units or applications. This leads to complex management and underutilization of resources. The limitations of CDC resulted in the emergence of Virtualized Data Centers (VDCs).

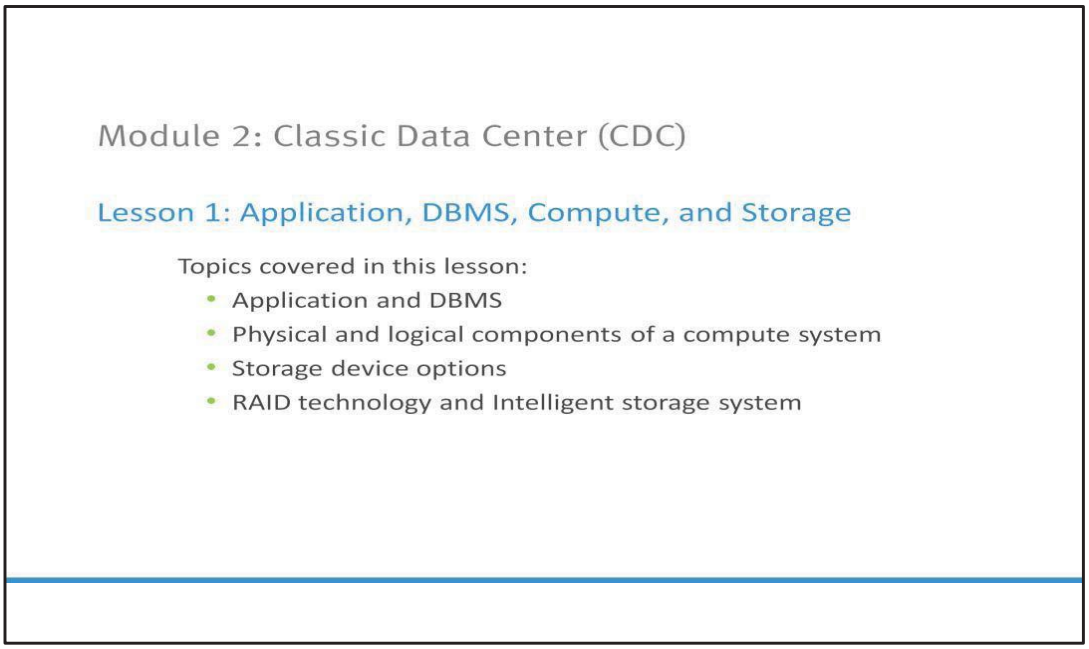
Continuous cost pressure on IT and on-demand data processing requirement of businesses have resulted in the emergence of Cloud computing. The virtualized data center forms the basis for understanding further discussion on Cloud infrastructure, service management, security and migration.

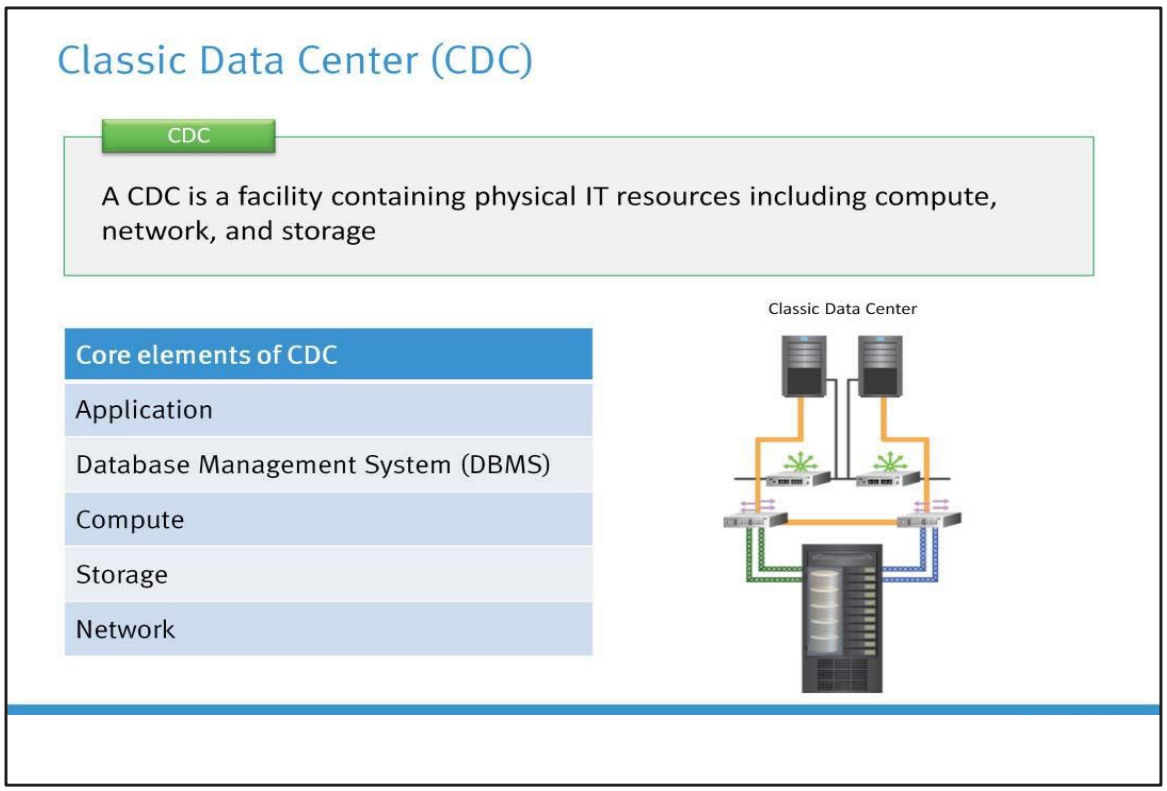
UNIT-2

Classic Data Center



It is important to understand the Classic Data Center (CDC) environment before discussing Virtual Data Center and Cloud. This module focuses on the key elements of CDC – compute, storage, and network, with focus on storage networking, business continuity, and data center management. This module does not cover classic compute and network in detail, based on the assumption that audience is already familiar with them.





The core elements of a CDC include compute (server), storage, network, application, and DBMS.

Application: An application is a computer program that provides the logic for computing operations. Applications may use a DBMS, which uses operating system services to perform store/retrieve operations on storage devices.

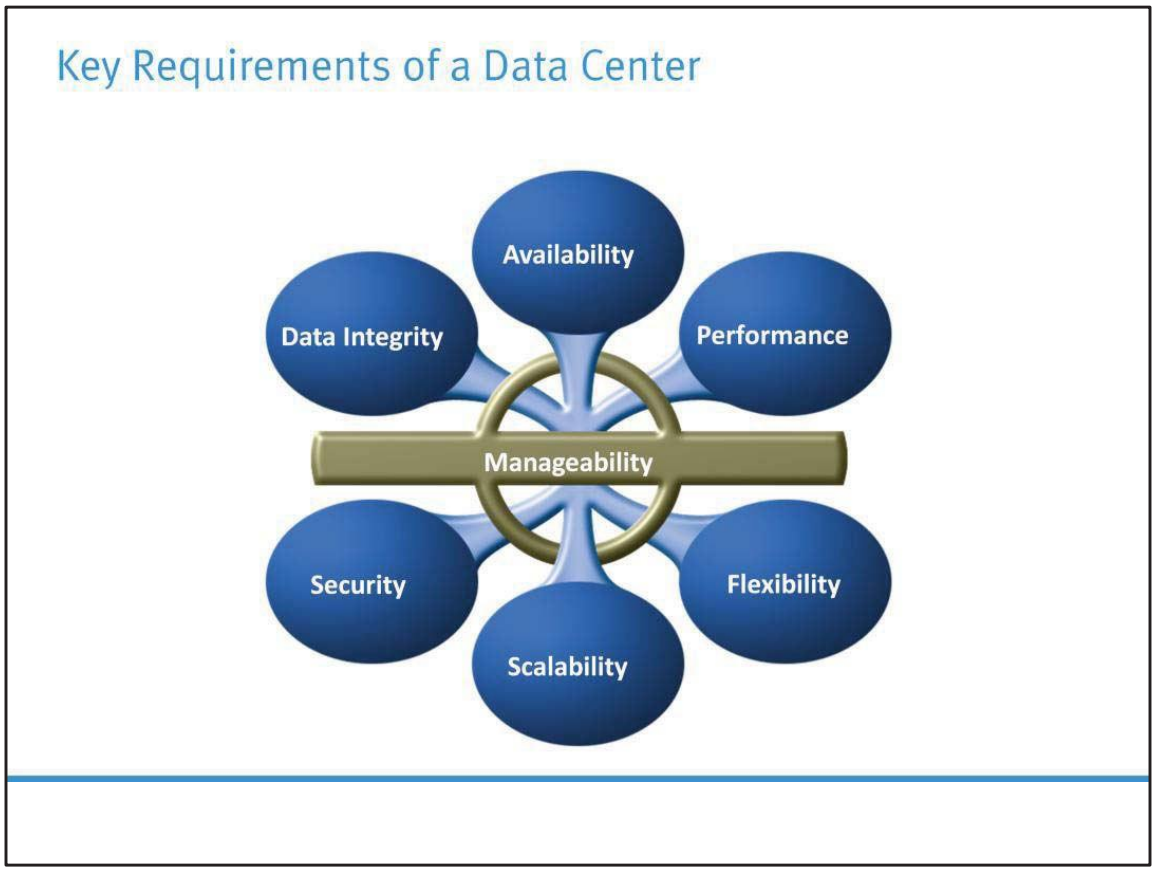
DBMS: DBMS provides a structured way to store data in logically organized tables that are interrelated. A DBMS optimizes the storage and retrieval of data.

Compute: Compute is a physical computing machine that runs operating systems, applications, and databases.

Storage: Storage refers to a device that stores data persistently for subsequent use.

Network: Network is a data path that facilitates communication between clients and compute systems or between compute systems and storage.

These core elements are typically viewed and managed as separate entities. But, all these elements must work together to address data processing requirements. Other elements of a CDC are power supplies and environmental controls, such as air conditioning and fire suppression.



A data center infrastructure should meet the following attributes to ensure that data is accessible to its users effectively and efficiently all the time:

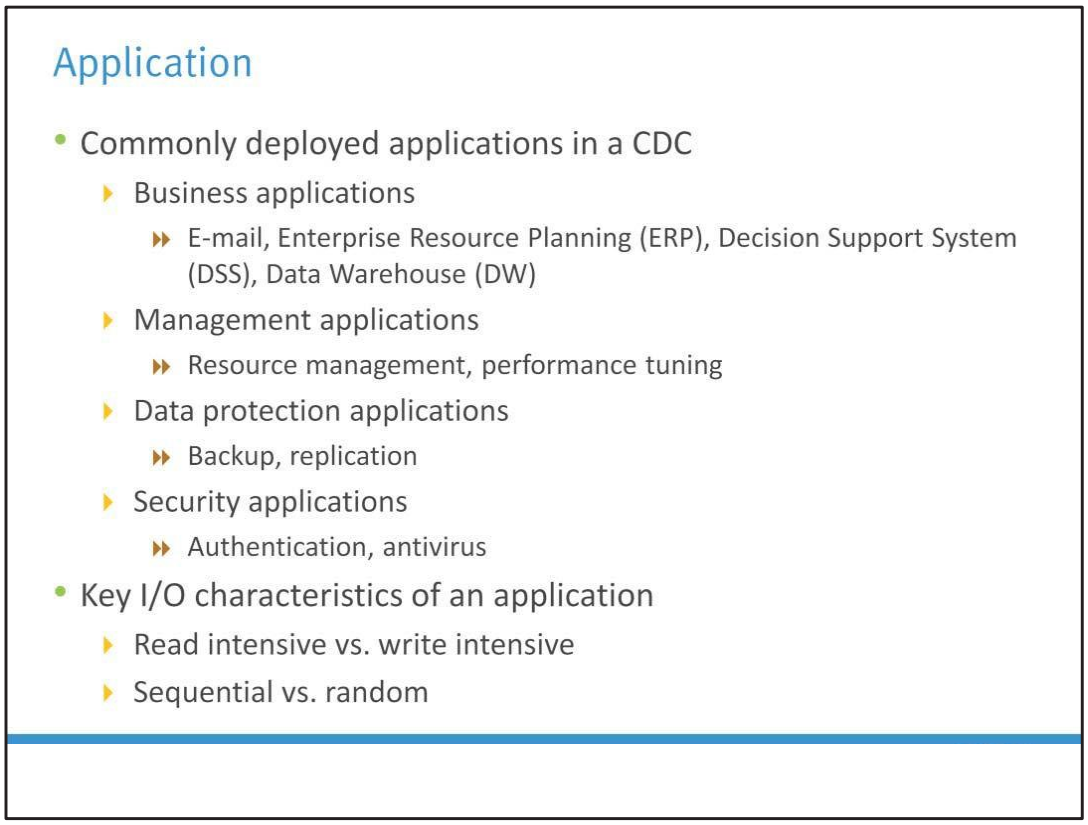
Availability: All data center elements should be designed to ensure accessibility. The inability of users to access data can have a significant negative impact on a business.

Performance: All the elements of the data center should provide optimal performance and service all processing requests at high speed.

Scalability/flexibility: Data center operations should be able to allocate additional processing capabilities or storage space on demand, without interrupting business operations. Business growth often requires deploying more servers, new applications, and additional databases.
The infrastructure should be flexible enough to meet the changing business requirements. Data center should be able to provide additional resources on demand without interrupting availability, or, at the very least, with minimal disruption.

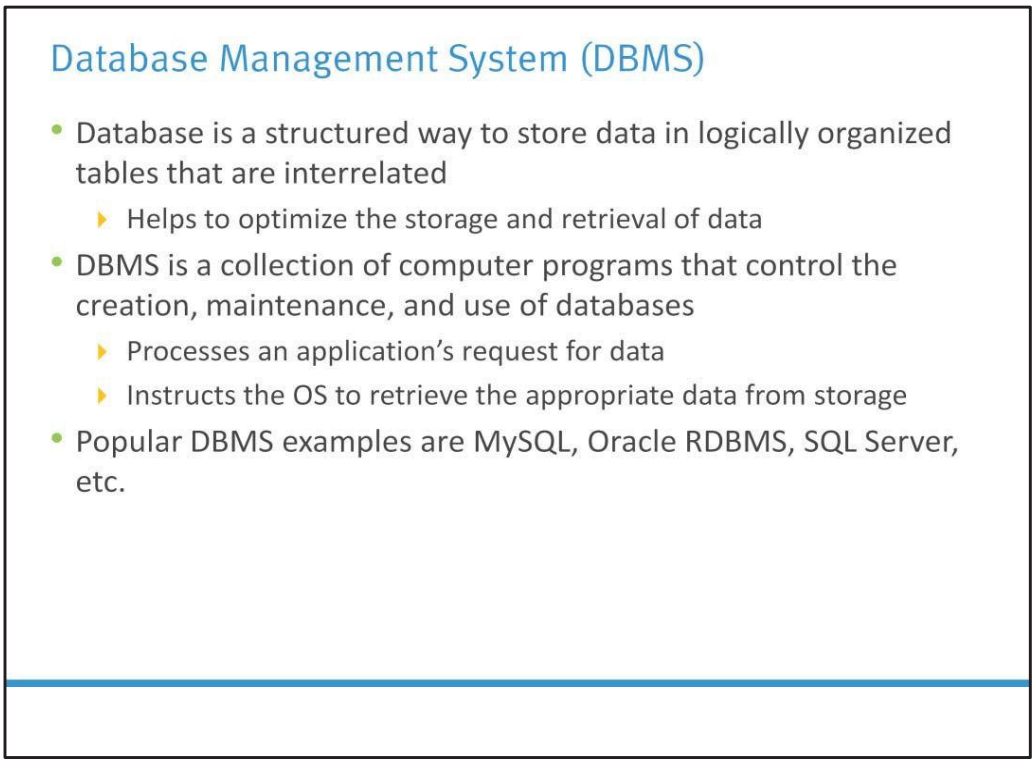
Security/Data integrity: It is important to establish policies, procedures, and proper integration of the data center elements to prevent unauthorized access to information. Data integrity ensures that data is unaltered. Any variation in data during its retrieval implies corruption, which may affect the operations of the organization.

Manageability: Manageability can be achieved through automation and reduction of manual intervention in common tasks. The additional resource requirements should primarily be managed by reallocating or improving utilization of existing resources, rather than by adding new resources. The cost of data center management is one of the key concerns.
Organizations are looking towards optimizing their IT expenditure on data center Maintenance, so that they can invest on innovation and new application deployment.

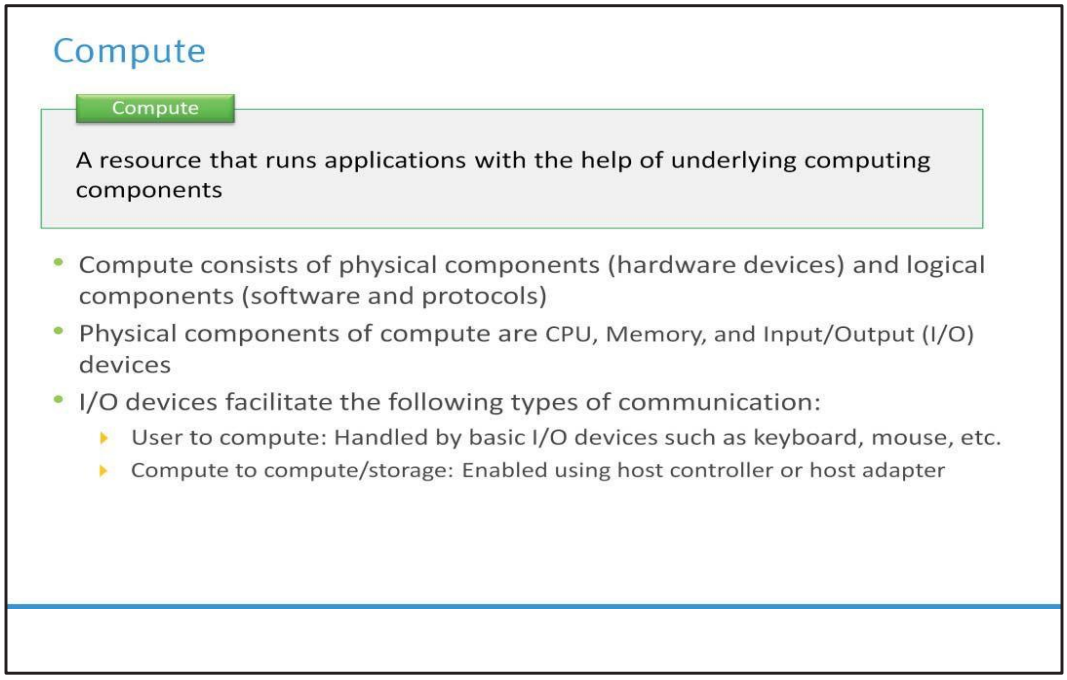


This slide lists the commonly-deployed applications in a data center. An application provides an interface between the user and the host and among multiple hosts. Typical business applications use databases that have a three-tiered architecture – the application user interface is the front-end tier; the computing logic or the application itself is the middle tier; and the underlying databases that organize the data is the back-end tier. The application sends requests to the underlying operating system to perform read/write (R/W) operations on the storage devices. Applications can be layered on the database, which in turn, uses OS services to perform R/W operations to storage devices. These R/W operations (I/O operations) enable transactions between the front-end and back-end tiers.

Application I/O (Input / Output) characteristic is an important parameter and influences the overall performance of the underlying IT solution.



A database is a structured way to store data in logically organized tables that are interrelated. A database helps to optimize the storage and retrieval of data. A Database Management System (DBMS) is a collection of computer programs that control the creation, maintenance, and use of a database. The DBMS processes an application's request for data and instructs the operating system to access the appropriate data. The DBMS manages incoming data, organizes, and facilitates the data to be modified or extracted by users or other programs.
Some popular DBMS solutions are MySQL, Oracle, SQL Server etc.

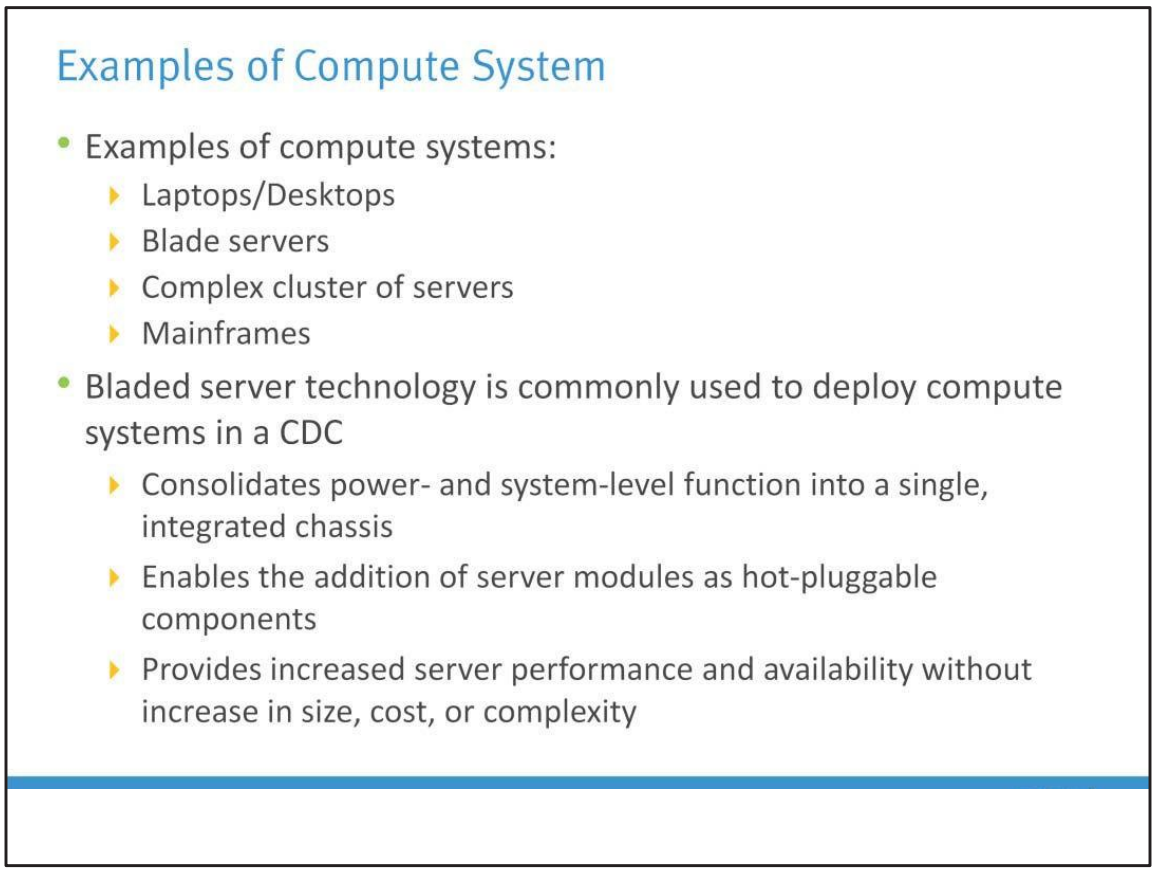


Compute consists of physical components (hardware devices) that communicate with one another using logical components (software and protocols). Compute has three key physical

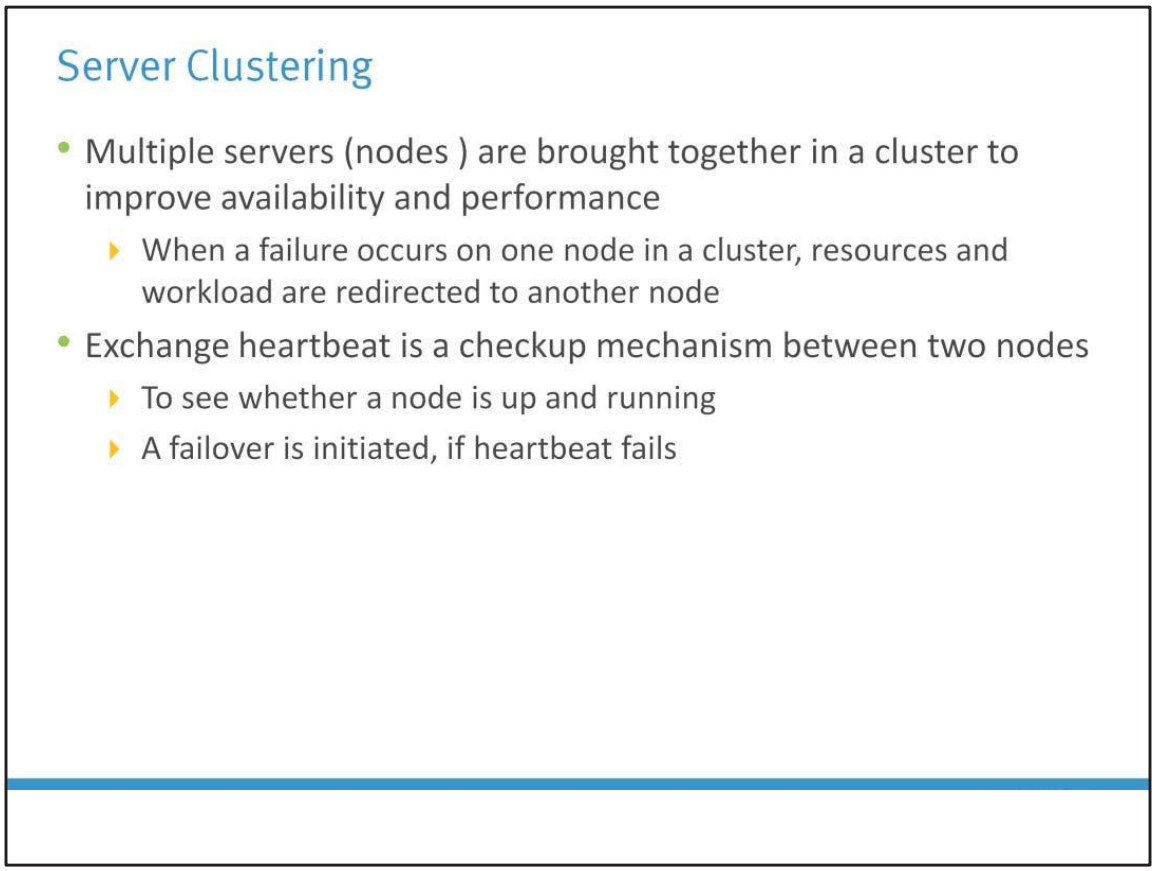
components: Central Processing Unit (CPU), Memory, and Input/Output (I/O) devices.

Memory is used to store data, either persistently or temporarily. Random Access Memory(RAM) and Read-Only Memory (ROM) are the two types of memory generally present in a compute system. I/O devices enable sending and receiving data to and from a compute system. This communication may be one of the following types:

- User to compute communications**:** Handled by basic I/O devices, such as the keyboard, mouse, and monitor. These devices enable users to enter data and view the results of operations.
- Compute to compute/ storage communications: Handled by the host controller or host adapter that connects a compute system to another compute system/ storage device. Host Bus Adaptor (HBA) is an example of a host controller that connects compute systems to Fibre channel storage devices. HBA is an Application-Specific Integrated Circuit (ASIC) board that performs I/O interface functions between the compute system and the storage, relieving the CPU from additional I/O processing workload. HBAs also provide connectivity outlets, known as Ports, to connect the compute systems to the storage device.



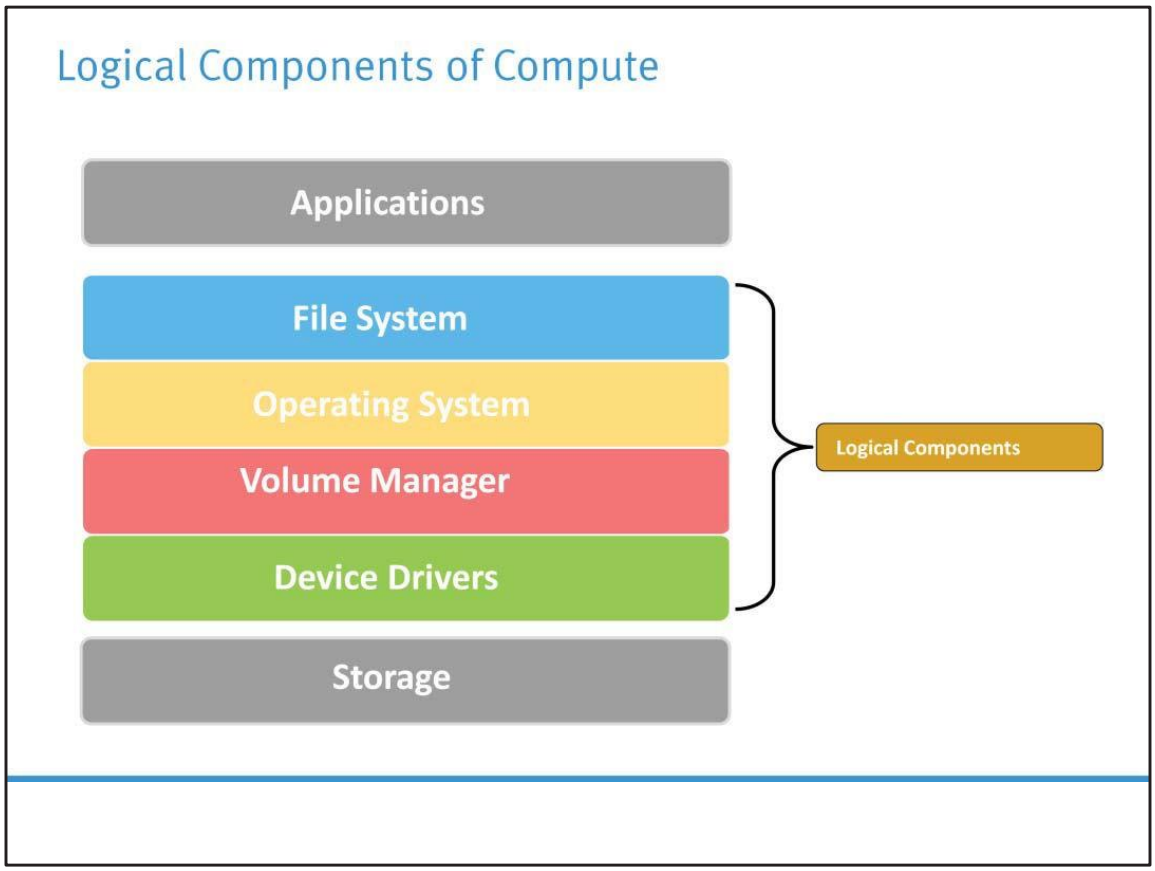
Compute systems may be a simple laptop, standalone servers/blade servers, mainframe computers, etc. Blade servers were developed in response to the growing need of a computing power that does not consume a large floor space. They bring down the cost and complexity in the datacenter. Blade servers consolidate power- and system-level functions into a single, integrated chassis and enable the addition of server modules as hot-pluggable components. Blade server technology greatly increases server density, lowers power and cooling costs, eases server expansion and simplifies datacenter management.



Server Clustering is the method of grouping two or more servers (also known as Nodes) and making them work together as a single system. This ensures that mission-critical applications and resources are highly available. When a failure occurs on one node in a cluster, resources are redirected and the workload is re-assigned to another node in the cluster. The cluster service is the software that connects the nodes in the cluster and provides a single-system view to the clients that are using the cluster.

Multiple servers (nodes) are brought together in a cluster to improve availability and performance. These nodes communicate with each other over a private network.
Communication between the nodes of a cluster enables the cluster service to detect node failures and status changes, and manages the cluster as a single entity.

Heartbeat is a checkup mechanism arranged between two nodes through private network to see whether a node is up and running. A failover is initiated if heartbeat is not functioning. In such cases, another node in the cluster will take over the workload of the failed node. Server clustering offers several benefits besides high availability. It offers scalability by enabling clusters to expand non-disruptively. It also provides load balancing by distributing the application load evenly among multiple servers within the cluster. Ease of management is another advantage of clustering because it allows non-disruptive maintenance of server resources.



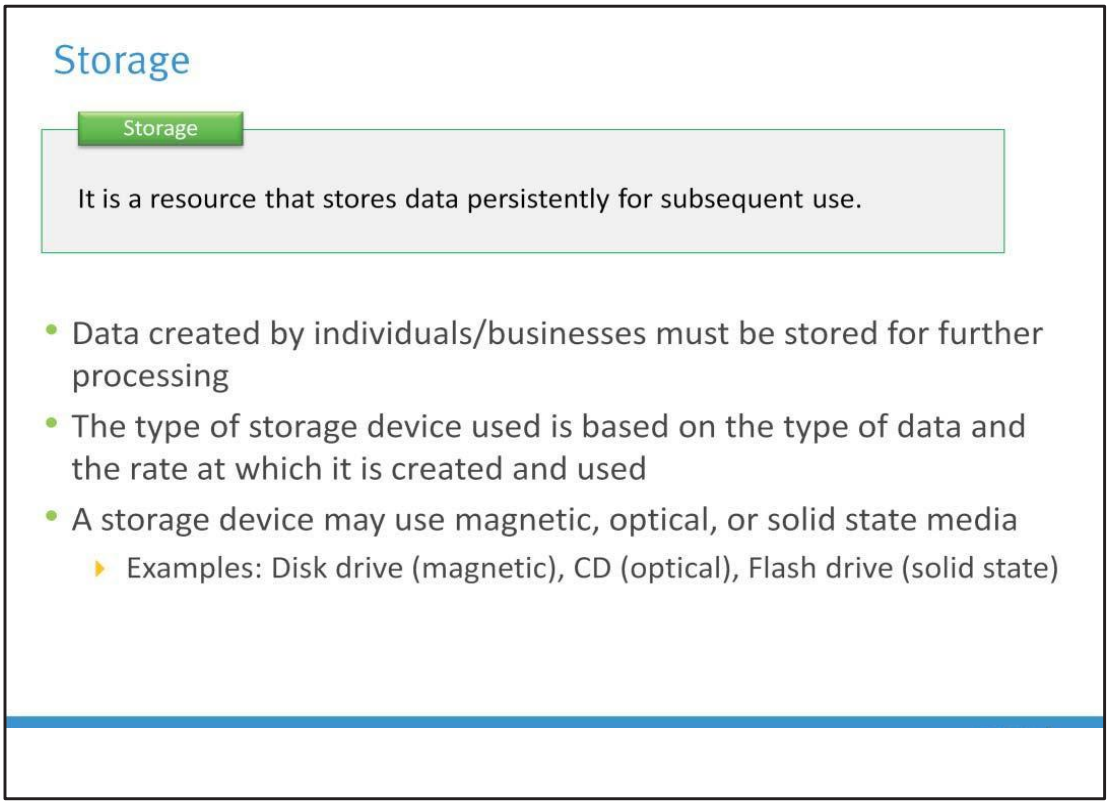
The following are the logical components of a compute system:

File System: A file is a collection of related records or data stored as a unit with a name. A file system is a hierarchical structure of files. File systems enable easy access to data files residing within a disk drive, a disk partition, or a logical volume. A file system needs compute-based logical structures and software routines that control access to files. It provides users with the functionality to create, modify, delete, and access files. Access to the files on the disks is controlled by the permissions given to the file by the owner. These permissions are also maintained by the file system.

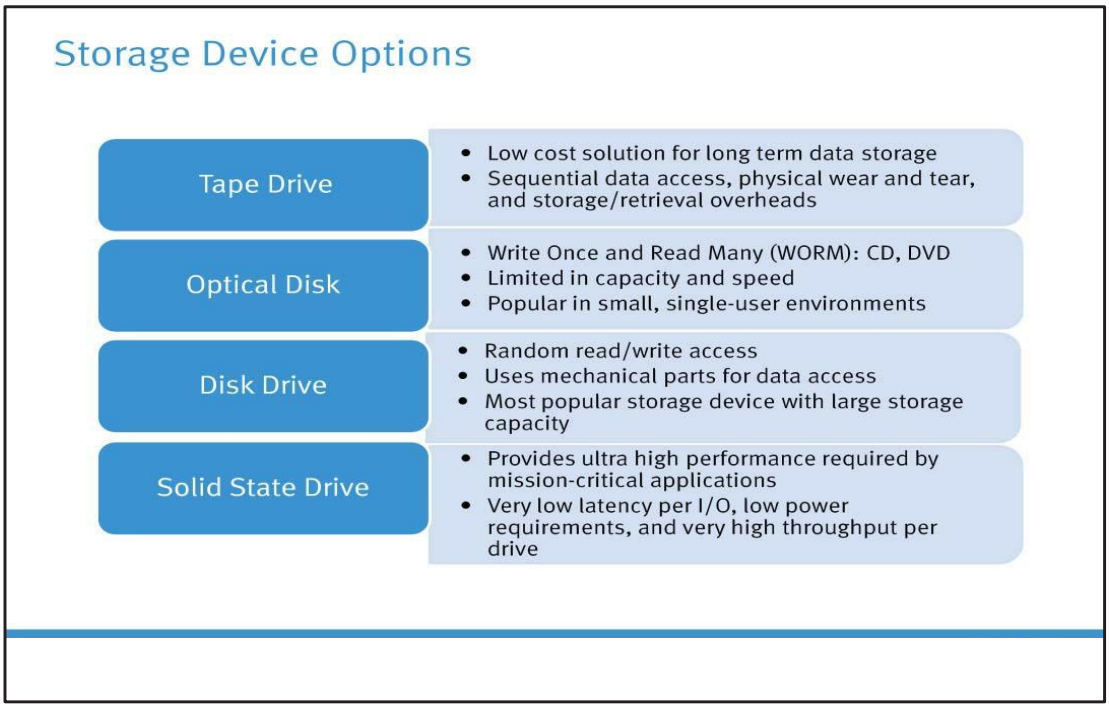
Operating system: An operating system controls all aspects of a computing environment. It works between the application and physical components of the compute system. One of the services it provides to the application is data access. The operating system also monitors and responds to user actions and the environment. It organizes and controls hardware components and manages the allocation of hardware resources. It provides basic security for the access and usage of all managed resources. An operating system also performs basic storage management tasks while managing other underlying components, such as the file system, volume manager, and device drivers.

Volume Manager: Logical Volume Managers (LVMs) introduce a logical layer between the operating system and physical storage. LVMs have the ability to define logical storage structures that can span multiple physical devices. The logical storage structures appear contiguous to the operating system and applications. The Logical Volume Manager provides a set of operating system commands, library subroutines, and other tools that enable the creation and control of logical storage.

Device Drivers: A device driver is special software that permits the operating system to interact with a specific device, such as a printer, a mouse, or a hard drive. A device driver enables the operating system to recognize the device and to use a standard interface (provided as an application programming interface or API) to access and control devices.



Data created by individuals or businesses must be stored so that it is easily accessible for further processing. Devices designed for storing data are termed as Storage Devices or simply Storage. The type of storage used varies based on the type of data and the rate at which it is created and used. Devices such as memory in a cell phone or digital camera, DVDs, CD- ROMs, and disk drives in personal computers are examples of storage devices. A storage device uses magnetic, optic or solid state media. Disks, tapes, and diskettes use magnetic media while CD/DVD uses optical media for storage. Removable flash memory or Flash drives are examples of solid state media.



Tapes are a popular storage option for backup purposes because of their relatively low cost and

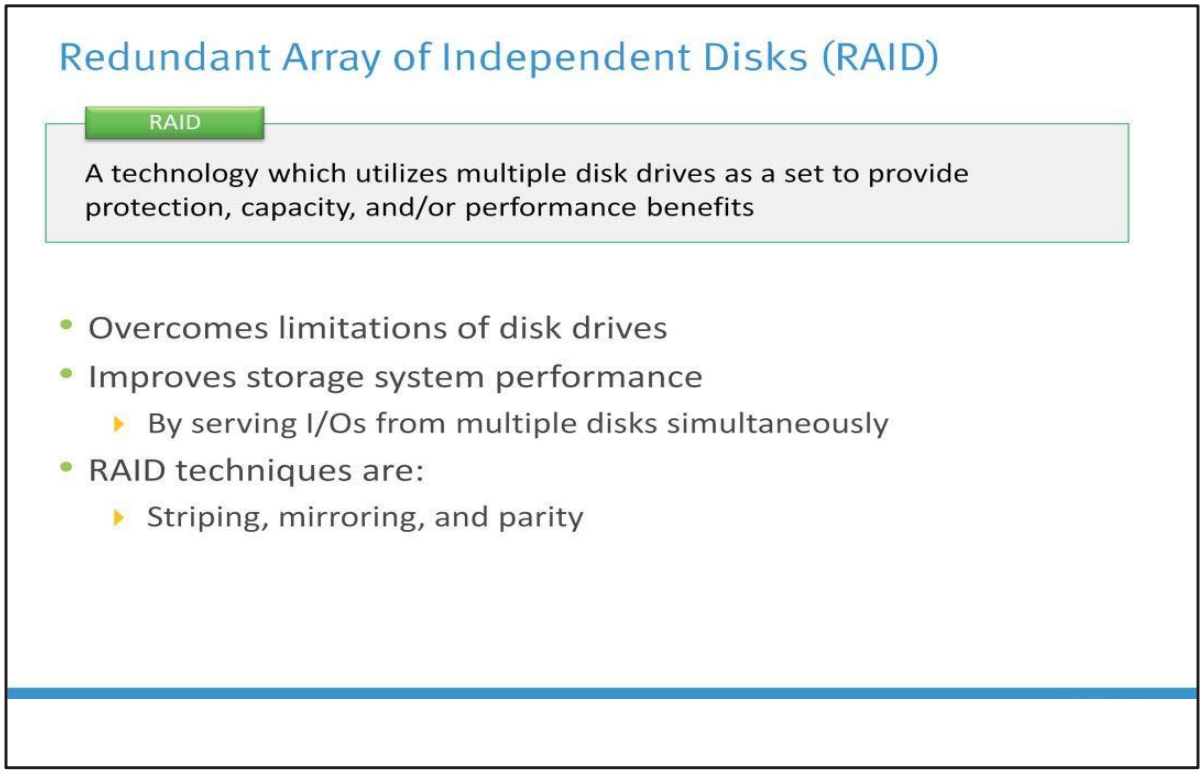
transportability. However, tape has its limitations; data is stored on the tape linearly along the length of the tape. Search and retrieval of data is done sequentially, taking several seconds to access the data. On a tape drive, the read/write head touches the tape surface, which causes the tape to degrade or wear out after repeated use. Also, the overhead associated with managing tape media is significant.

Optical disk storage is popular in small, single-user computing environments. Optical disks have limited capacity and speed, which limits the use of optical media as a business data storage solution. The capability to Write Once and Read Many (WORM) is an advantage of optical disk storage. Optical disks, to some degree, guarantee that the content has not been altered. Hence, they can be used as low-cost alternatives for long-term storage of relatively small amounts of fixed content.

Disk drives are the most popular storage medium for storing and accessing data for performance- intensive, online applications. A disk drive uses a rapidly moving arm to read and write data across a flat platter coated with magnetic particles. Data is transferred from the magnetic platter through the R/W head to the computer. Several platters are assembled together with the R/W head and controller, and this is referred to as a Disk Drive. Disks support rapid access to random data locations. This means that data can be written or retrieved quickly for a large number of simultaneous users or applications. In addition, disk drives provide large storage capacity.

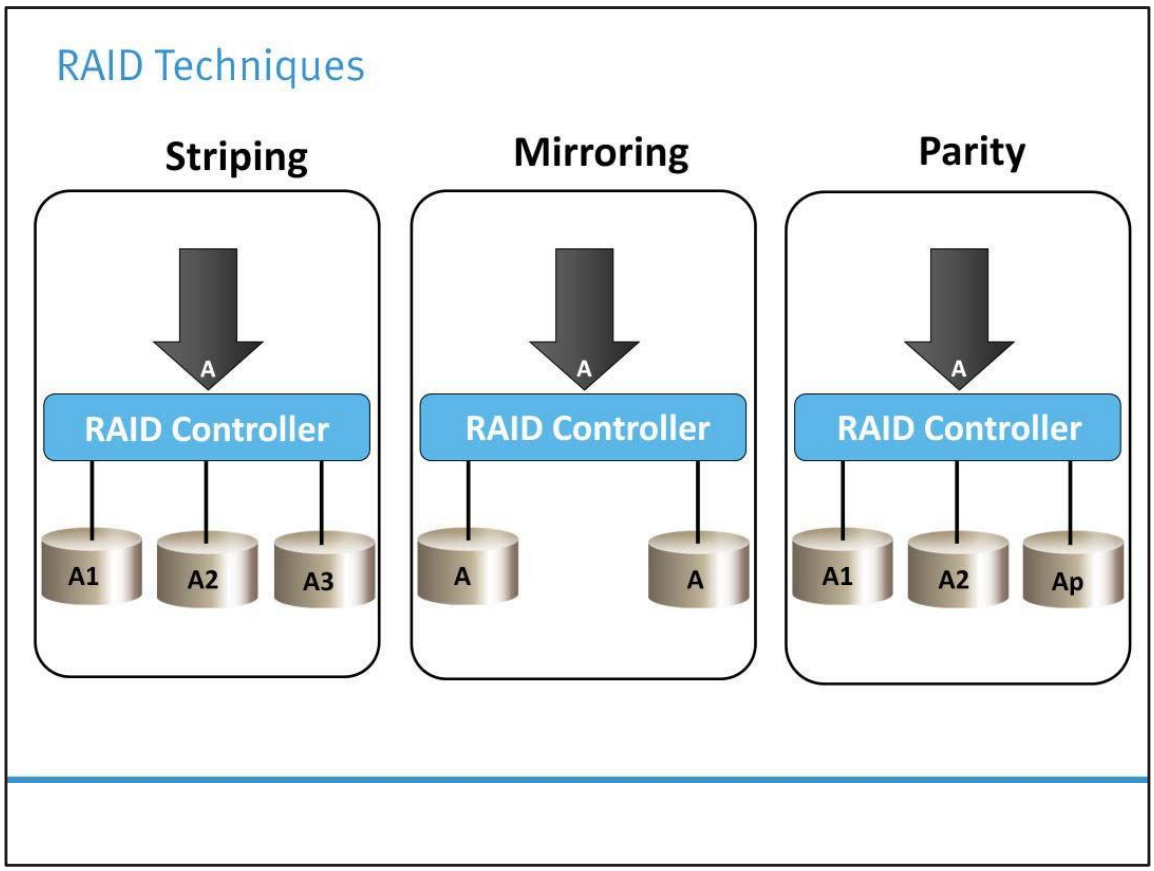
Solid State Drives (SSDs), also referred as Flash Drives, are new generation drives that deliver ultra-high performance required by mission-critical applications. Flash drives use semiconductor based solid state memory (flash memory) to store and retrieve data.

Note: The compute system and the storage device communicate with each other by using predefined protocols such as IDE/ATA, SATA, SAS, and FC. These protocols are implemented on the disk-controller interface. As a result, a disk drive is also known by the name of the protocol it supports. For example, if a disk supports SATA protocol, then it is called as SATA disk.



Disk drives are susceptible to failures due to mechanical wear and tear and other environmental factors. A disk drive failure could result in data loss. A disk drive has a projected life expectancy before it fails. Today, data centers house thousands of disk drives in their storage infrastructures. Greater the number of disk drives in a storage array, greater the probability of a disk failure in the array. RAID technology was developed to mitigate this problem. RAID is an enabling technology

that leverages multiple disks as part of a set. The technology also provides data protection against drive failures. In general, RAID implementations also improve storage system performance by serving I/Os from multiple disks simultaneously. Storage systems with flash drives are also benefited in terms of protection and performance by implementing RAID. Striping, Mirroring, and Parity are RAID techniques that form the basis for defining various RAID levels. These techniques determine the data availability and performance of a RAID set. RAID controller helps implementing these RAID techniques.



Striping is a technique of spreading data across multiple drives in order to use the drives in parallel. All read-write heads work simultaneously. This allows more data to be processed in a shorter time. Consequently, performance increases, when compared to writing/retrieving data to/from one disk at a time.

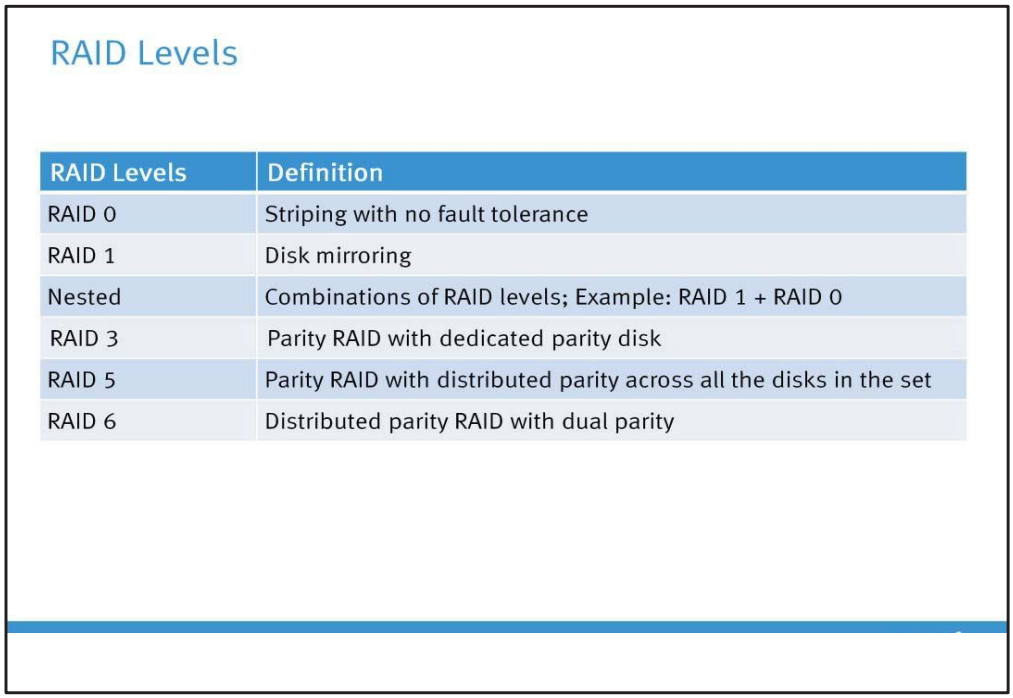
Mirroring is a technique where data is stored on two different disk drives, yielding two copies of data. In the event of one drive failure, the data is intact on the surviving drive, and the controller continues to service the compute system's data requests from the surviving disk of the mirrored pair. Mirroring improves read performance because read requests are serviced by both disks. However, write performance deteriorates because each write request manifests as two writes on the disk drives.

Mirroring is expensive because it involves duplication of data — the amount of storage capacity required is twice the amount of data being stored. Mirroring can be implemented with striped RAID, by mirroring entire stripes of disk set to stripes on the other disk set. This is known as nested RAID.

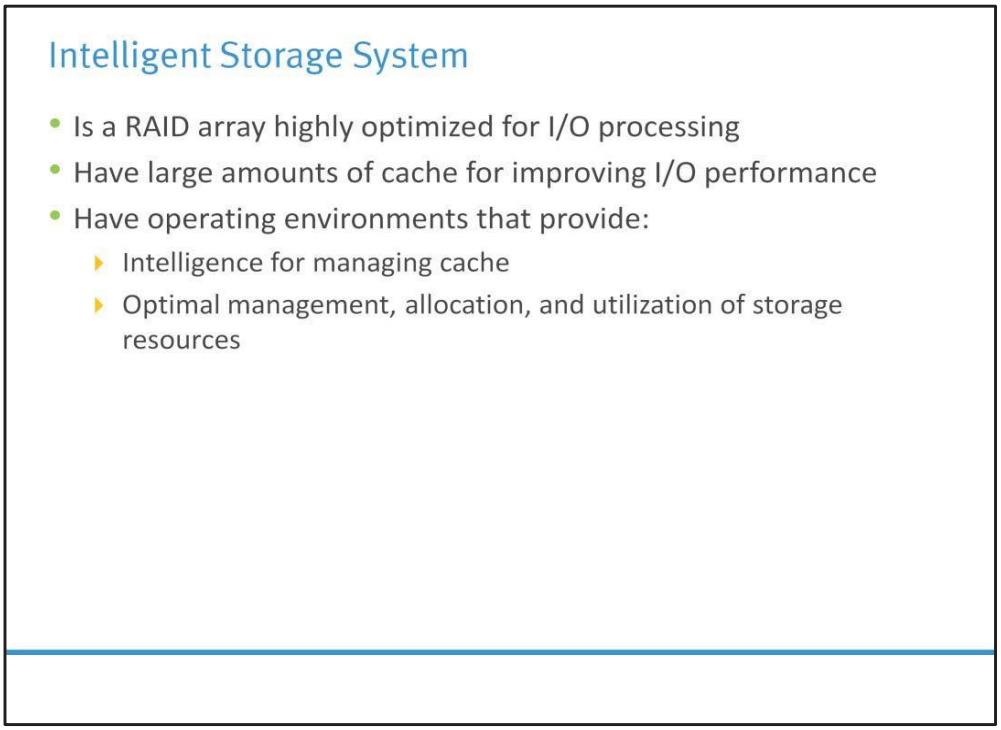
Parity is a method of protecting striped data from disk failure without the cost of mirroring. An additional disk drive is added in the strip set to hold parity, a mathematical construct that allows re-creation of the missing data. Parity RAID is less expensive than mirroring because parity overhead is only a fraction of the total capacity. Parity information can be stored on separate, dedicated disk drives or distributed across all the drives in a RAID set. Parity calculation is a bitwise XOR operation. Calculation of parity is a function of the RAID controller. If one of the disks fails in a RAID set, the

value of its data is calculated by using the parity information and the data on the surviving disks. The value is calculated using XOR operation.

However, there are some disadvantages of using parity. Parity information is generated from data on the data disks. As a result, parity is recalculated every time there is a change in data. This recalculation takes time and affects the performance during write operation.



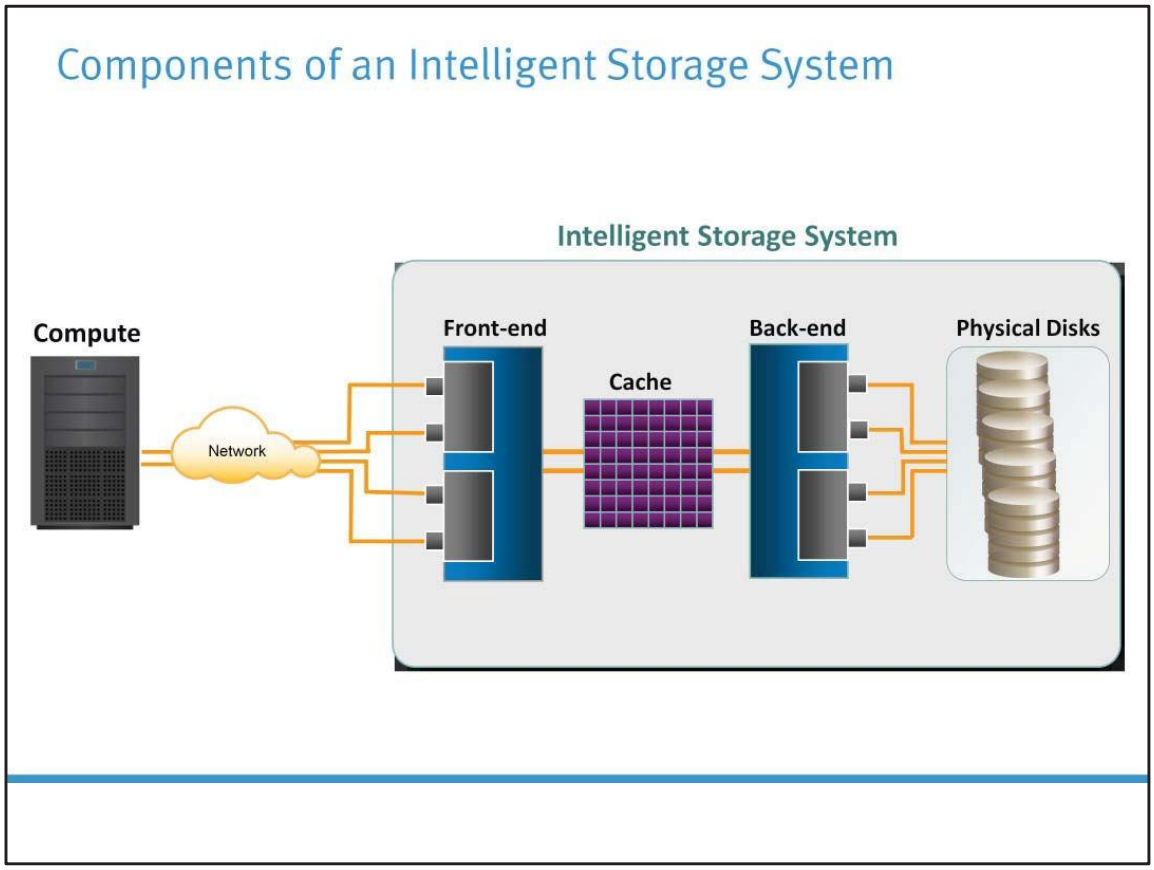
Application performance and data availability requirements determine the RAID level selection. These RAID levels are defined on the basis of the RAID technique(s) implemented. Some RAID levels use one technique, whereas others use a combination of techniques. Commonly used RAID levels are shown in the table.



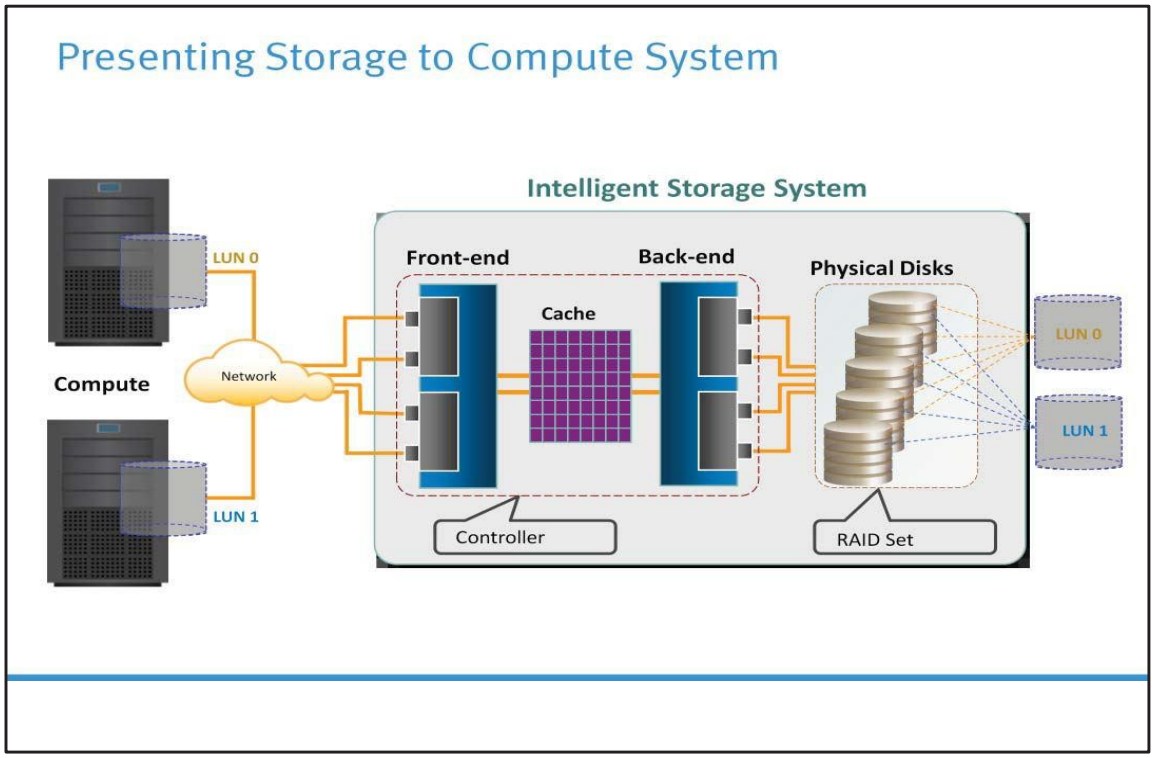
Business-critical applications require high levels of performance, availability, security, and scalability. A disk drive is a core element of storage that governs the performance of any storage

system. Some of the older disk array technologies could not overcome performance constraints due to the limitations of a disk drive and its mechanical components. RAID technology made an important contribution to enhance storage performance and reliability, but disk drives, even with a RAID implementation could not meet performance requirements of today's applications.

With advancements in technology, a new breed of storage solutions, known as an Intelligent Storage System, has evolved. These intelligent storage systems are feature-rich RAID arrays that provide highly optimized I/O processing capabilities. These storage systems are configured with a large amount of memory (called cache) and multiple I/O paths and use sophisticated algorithms to meet the requirements of performance-sensitive applications.
These arrays have an operating environment that intelligently and optimally handles the Management, allocation, and utilization of storage resources.



An intelligent storage system consists of four key components: Front-end, Cache, Back-end, and Physical Disks. The figure on the slide illustrates these components and their interconnections. An I/O request received from the compute system at the front-end port is processed through cache and the back end to enable storage and retrieval of data from the physical disk. A read request can be serviced directly from cache if the requested data is found in cache.



In an intelligent storage system, physical disks are logically grouped together to form a set, called RAID set, on which a required RAID level is applied. The number of drives in the RAID set and the RAID level determine the availability, capacity, and performance of the RAID set. It is highly recommended that the RAID set be created from the same type, speed, and capacity drives to ensure the maximum usable capacity, reliability, and consistent performance. For example, if drives of different capacities are mixed in a RAID set, then the capacity of the smallest drive will be used from each disk in the set to make up for the RAID set's overall capacity. The remaining capacity of the larger drives will remain unused. Likewise, mixing higher Revolutions per minute(RPM) drives with lower RPM drives lowers the overall RAID set's performance.

RAID sets usually have large capacities because they combine the total capacity of individual drives in the set. Logical Units are created from the RAID sets by partitioning (seen as slices of RAID set) the available capacity into smaller units. These units are then assigned to the compute system for their storage requirements.

Logical units are spread across all the physical disks, which belong to that set. Each logical unit created from the RAID set is assigned a unique ID called Logical Unit Number (LUN). LUNs hide the organization and composition of RAID set from the compute systems. The diagram on the slide shows a RAID set consisting of five disks that have been sliced or partitioned into two LUNs: LUN 0 and LUN 1. These LUNs are then assigned to compute system 1 and compute system 2 for their storage requirements.

It is also possible to control access of LUNs by a compute system. This is done with the help of "LUN masking". LUN masking is a process that provides data access control by defining which LUNs a compute system can access. LUN masking function is implemented on the storage processor/controller. This ensures that the volume access by servers is controlled appropriately, preventing unauthorized or accidental use in a shared environment.

Module 2: Classic Data Center (CDC)

Lesson 2: Storage Networking Technologies -1

Topics covered in this lesson:

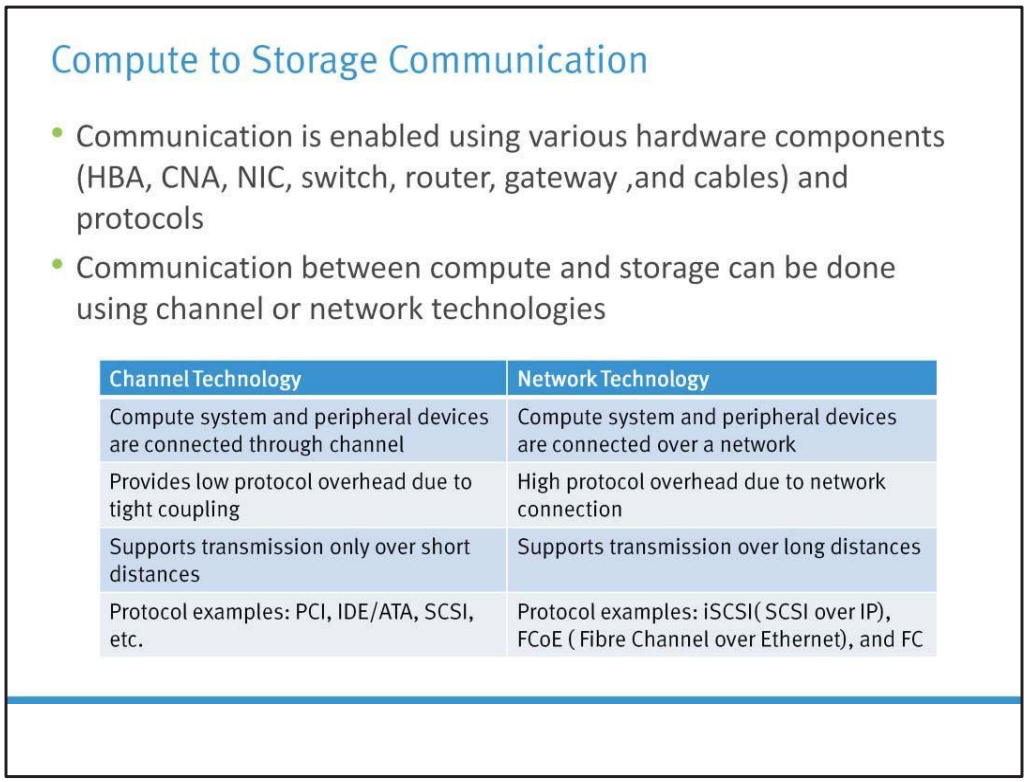
- Compute to compute communication
- Compute to storage communication
  - Direct Attached Storage (DAS)
  - Fibre Channel SAN (FC-SAN)

Compute to Compute Communication

- Typically uses Ethernet or TCP/IP protocol
  - LAN, MAN, and WAN
- Communication is enabled using various components:
  - Network Interface Card (NIC)
    - Has unique MAC address
  - Switches and routers
    - Switch provides scalability and interconnection between multiple compute systems
    - Routers allow different networks to communicate with each other
  - Cables
    - Twisted pair, co-axial cable, optical fiber

Compute to compute communication typically uses Ethernet for Local Area network (LAN) and TCP/IP protocol for Metropolitan Area Network (MAN) and Wide Area Network (WAN). Each compute system is connected to the network through a Network Interface Card (NIC). Each NIC card has a unique address called Media Access Control (MAC) address, which uniquely identifies nodes in the network. Switches and routers are the commonly used interconnect devices. Switches allow

different nodes of a network to communicate directly with each other and to provide scalability to the network. A router is a device or a software that determines the next network point to which a packet should be forwarded to reach its destination. Router allows different networks to communicate with each other. Commonly used physical media/cables are twisted pair, co-axial cable, optical fiber, and so on.

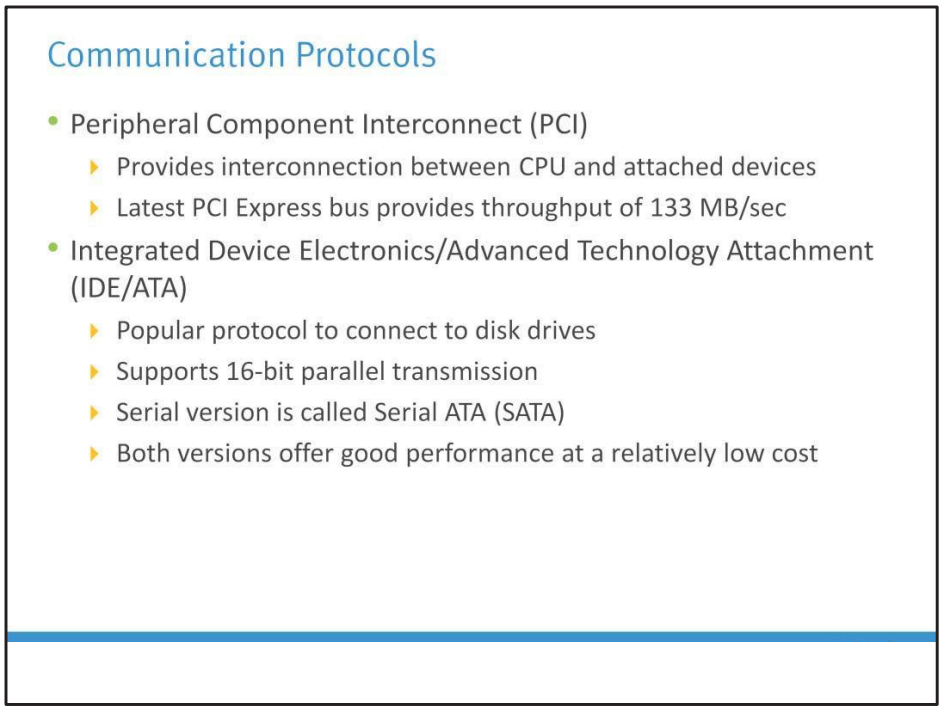


Compute to storage communication is enabled by various hardware components such as HBA, CNA, NIC, switch, router, cables, and protocols.

- Host bus adaptor (HBA): Is an application-specific integrated circuit (ASCI) board that performs I/O interface functions between the host and the storage, relieving the CPU from additional I/O processing workload.
- Converged network adaptor (CNA): Is a multi function adapter which consolidates the functionality of an NIC card and a Fibre Channel HBA onto a single adapter.

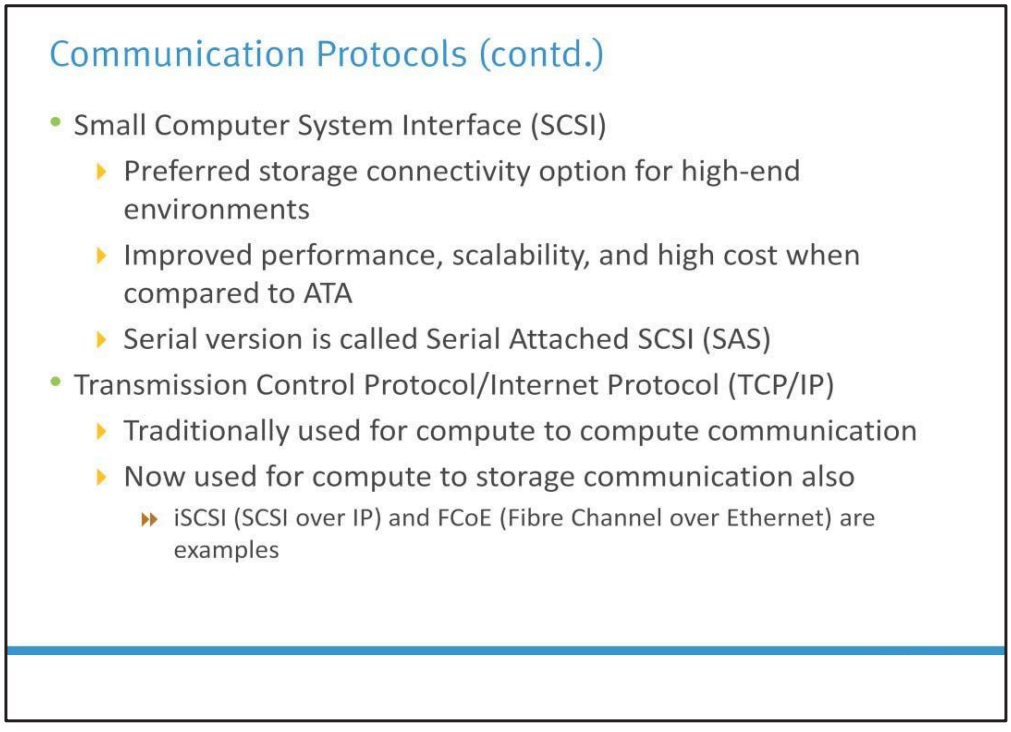
Compute systems communicate with storage devices using channel or network technologies. Channel technologies provide fixed connections between compute systems and their peripheral devices and support communication over short distances. When using channels, static connections are defined to the operating system in advance. Tight integration between the transmission protocol and the physical interface minimizes the overhead required to establish communication and to transport large amounts of data to the statically defined devices. Network technologies are more flexible than channel technologies, and provide greater distance capabilities. Network communication involves sharing bandwidth among multiple systems, and this results in greater protocol overhead and reduced performance.

Channel and network technologies use distinct hardware components, such as interconnect devices, cables, buses, ports etc, and protocols for communication. Switches, routers, modems, etc are some of the commonly used interconnect devices. Port is a specialized outlet that enables connectivity between a device and other external devices. Cables connect compute systems to internal or external storage devices using copper or fiber optic media. Twisted pair, coaxial, optical fiber etc are some of the commonly used cables.



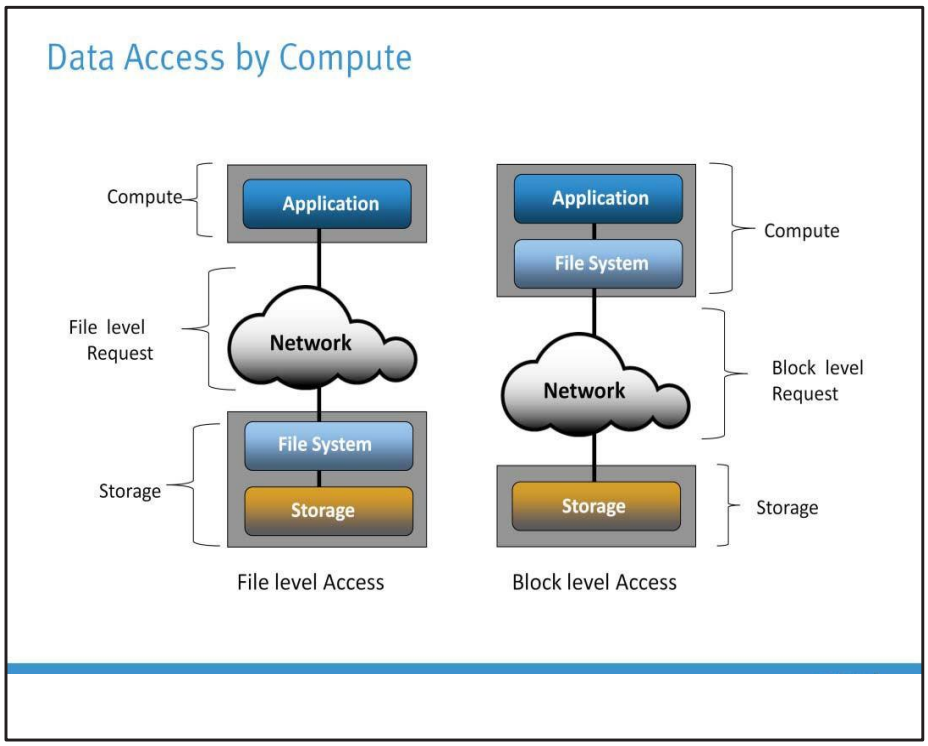
PCI is a specification that standardizes how PCI expansion cards, such as network cards or modems, exchange information with the CPU. PCI provides the interconnection between the CPU and the attached devices. The plug-and-play functionality of PCI enables the compute system to easily recognize and configure new cards and devices. The width of a PCI bus can be 32 bits or 64 bits. A 32-bit PCI bus can provide a throughput of 133 MB/s. PCI Express is an enhanced version of the PCI bus with higher throughput and clock speed.

IDE/ATA is a popular interface protocol used for connecting storage devices, such as disk drive and CD-ROM drive. This protocol supports 16-bit parallel transmission, and therefore, is also known as Parallel ATA (PATA) or simply ATA. A serial version of this protocol supports a single-bit serial transmission and is known as Serial ATA (SATA). Both ATA and SATA offer good performance at a relatively low cost.



SCSI has emerged as a preferred connectivity protocol in high-end computers. This protocol supports parallel transmission and offers improved performance, scalability, and compatibility compared to ATA. However, the higher cost associated with SCSI limits the popularity among home or business desktop users. Over the years, SCSI has been enhanced and now includes a wide variety of related technologies and standards. Serial attached SCSI (SAS) is a serial version of SCSI. SAS 600 supports data transfer rates up to 6Gb/s.

TCP/IP is also referred to as the Internet Protocol (IP). It is actually two standardized protocols that are "stacked" together. TCP (Transmission Control Protocol) is the higher-level protocol; it creates packets for the data and messages that need to be transmitted. IP is the lower-level protocol; it is responsible for adding addresses to the packets to ensure that the packets reach the proper destination. It is conventionally used for compute to compute communication. However, because of the advancements in technology, these networks are used for compute to storage communication. This gives cost benefit to the organizations because they can leverage the existing networks to provide compute to storage communication. The two protocols that provide compute to storage communication using IP networks are iSCSI (SCSI over IP) and FCIP(Fibre Channel over IP).



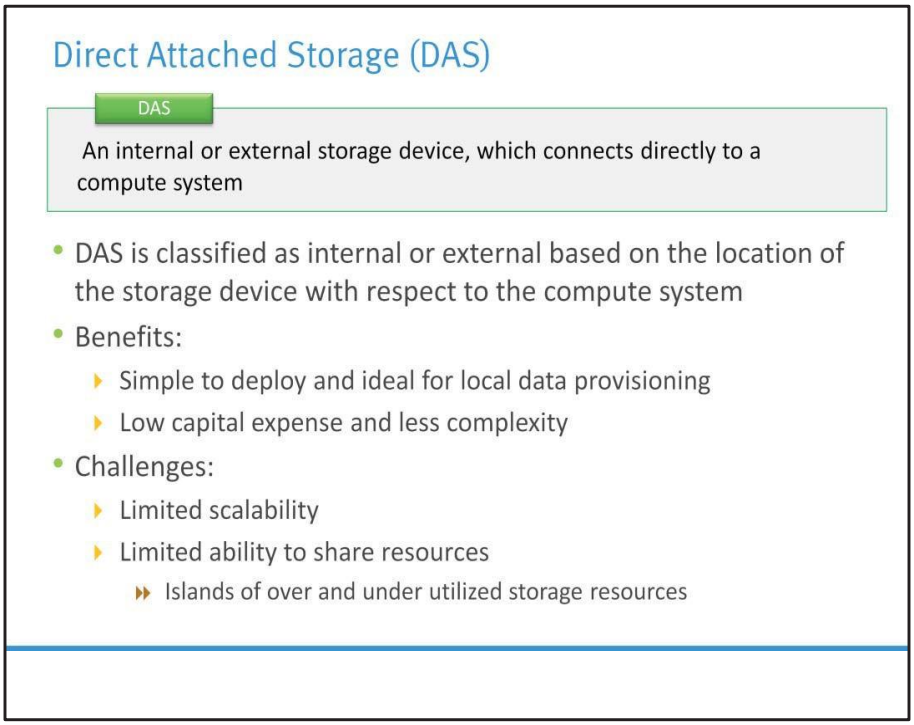
Data is accessed and stored by an application using the underlying infrastructure, as shown in the diagram. The key components of this infrastructure are operating system (or File system), Connectivity (network) and Storage itself. The storage device can be internal and (or) external to the compute system. In either case, the host controller card inside the compute systems accesses the storage devices using pre-defined protocols such as IDE/ATA, SCSI, or Fibre Channel. IDE/ATA and SCSI are popularly used in small and personal computing environment for accessing internal storage. Fibre Channel and iSCSI protocols are used for accessing data from an external storage device (or subsystems). External storage devices can be connected to the compute systems directly or through a storage network. Data can be accessed over a network in one of the following ways – Block level or file level.

In general, an application requests data from the file system (or operating system) by specifying the file name and location. The file system maps the file attributes to the logical block address of the data and sends it to the storage device. The storage device converts LBA to CHS address and fetches

the data.

In block-level access, the file system is created on a compute system, and the data is accessed on a network at the block level, as shown in diagram. In this case, raw disks or logical volumes are assigned to the compute for creating a file system. In a file level access, the file system is created on a network or at the storage side and the file level request is sent over a network. Because data is accessed at a file level, this method has higher overhead, compared to the data accessed at the block level.

Note: *Earlier disk drives used physical addresses consisting of Cylinder, Head, and Sector (CHS) numbers to refer to specific locations on the drive. Logical Block Addressing (LBA) simplifies addressing by using a linear address to access physical blocks of data.*



Direct Attached Storage (DAS) is an architecture where storage is connected directly to compute systems. Applications access data from DAS at a block level. DAS is classified as internal or external based on the location of the storage device with respect to the compute system.

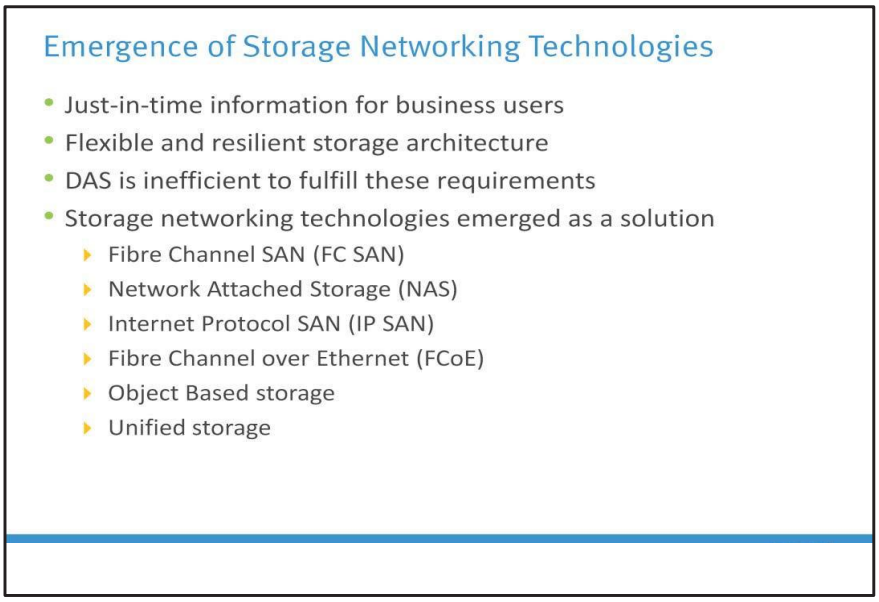
Internal DAS: In internal DAS architectures, the storage device is located within the compute system and is connected to the compute system by a serial or parallel bus. The internal disk drive of a compute system is an example of internal DAS.

External DAS: In external DAS architectures, the storage device is located outside the compute system. In most cases, communication between the compute system and the storage device takes place over an SCSI or FC protocol. Tape libraries and directly connected external disk drive packs are some examples of external DAS. External DAS overcomes the distance and device count limitations of internal DAS.

DAS requires a relatively lower initial investment when compared to other storage networking technologies. DAS configuration is simple and can be deployed easily and rapidly. Setup is managed using compute based tools, such as the compute system OS, which makes storage management tasks easy for small and medium enterprises. DAS is the simplest solution when compared to other storage networking models, and requires fewer management tasks and less hardware and software elements to set up and operate.

DAS Challenges:

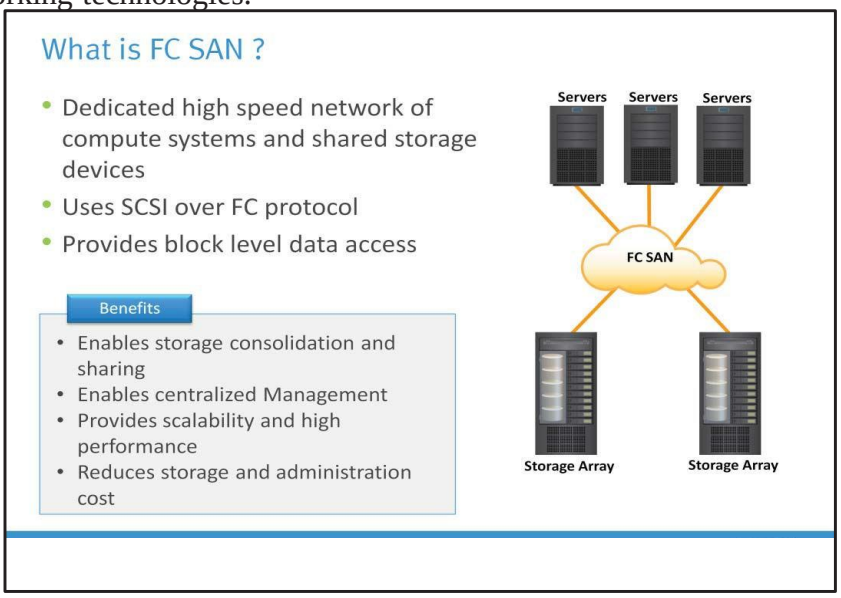
DAS does not scale well. A storage device has a limited number of ports, which restricts the number of compute systems that can directly connect to the storage. A limited bandwidth in DAS restricts the available I/O processing capability. DAS does not make optimal use of resources due to its limited ability to share front end ports. In DAS environments, unused resources cannot be easily re-allocated, resulting in islands of over-utilized and under-utilized storage pools.



Just-in-time information for business users: Information must be available to business users when they need it. The explosive growth in online storage, proliferation of new servers and applications, spread of mission-critical data throughout enterprises, and demand for 24 × 7 data availability are some of the challenges that need to be addressed.

Flexible and resilient storage architecture: The storage infrastructure must provide flexibility and resilience that aligns with changing business requirements. Storage should scale without compromising performance requirements of the applications. At the same time, the total cost of managing information must be low.

DAS is inefficient to fulfill these requirements, and this was the motivation for the evolution of storage networking technologies.

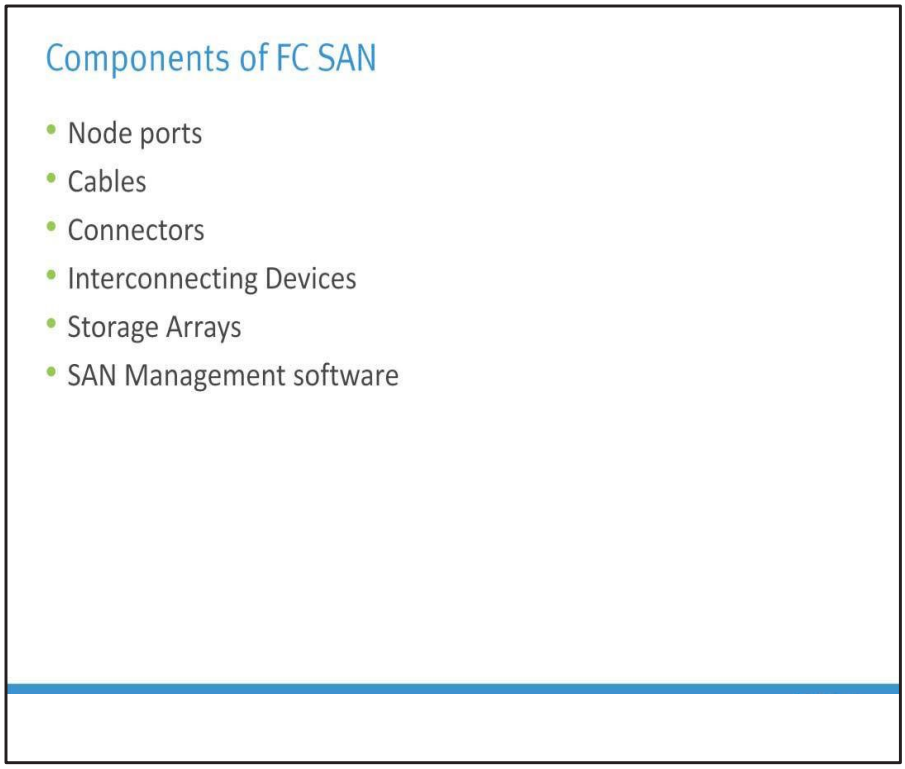


FC SAN is a high-speed, dedicated network of compute systems and shared storage devices. FC SAN uses SCSI over FC protocol to transfer data between compute systems and storage devices. It provides block-level access to storage devices.

FC SAN enables storage consolidation and allows a storage system to be shared by multiple compute systems. This improves utilization of storage resources, compared to DAS architecture. Because compute systems share a common pool of storage, the total amount of storage an organization needs to purchase and manage is reduced. As consolidation occurs, storage management becomes centralized and less complex, which significantly reduces the people cost for managing data. SAN architecture is highly scalable. Theoretically, it can be scaled up to approximately 15 million devices.

FC SAN ensures high availability of critical applications by providing multiple connectivity paths between compute system and storage devices. FC SAN provides hot-plugging capabilities that enable organizations to install, configure, and bring storage online without experiencing server downtime.

FC provides a serial data transfer interface that can operate over copper wire (for back end connectivity) and/or optical fiber (for front end connectivity). 16 GFC Fibre Channel interface enables data rates up to 16 Gbps and offers throughput of 3200 MB/s. FC communication uses Fibre Channel Protocol (FCP), which is SCSI data, encapsulated and transported within Fibre Channel frames.



A SAN consists of three basic components: servers(compute systems), network infrastructure, and storage. Servers and storage are the end points or terminal devices (called ‘nodes’) which provide ports for communication. Network infrastructure consists of cables and connectors, interconnecting devices (such as FC switches or hubs), and SAN management software.

SAN implementations use optical fiber cabling. Copper may be used for shorter distances, example back-end connectivity, because it provides a better signal-to-noise ratio for distances up to 30 meters.

A connector is attached at the end of a cable to enable swift connection and disconnection of the cable to and from a port. Standard connector (SC) and a Lucent connector (LC) are two commonly

used connectors for fiber optic cables.

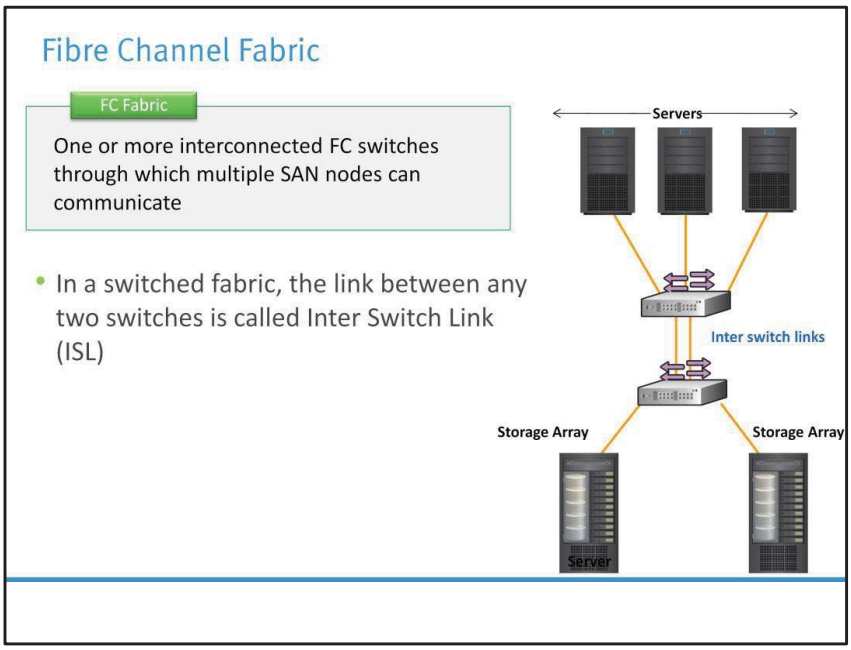
FC Hubs, departmental switches, and enterprise directors are the interconnect devices commonly used in FC SAN. Hubs are used as communication devices in Fibre Channel Arbitrated Loop (FC-AL) implementations. Hubs physically connect nodes in a logical loop or a physical star topology.

Because low cost and high performance switches are readily available, hubs are rarely used in FC SANs. Switches are more intelligent than hubs and directly route data from one physical port to another. Therefore, nodes do not share the bandwidth; instead, each node has a dedicated communication path.

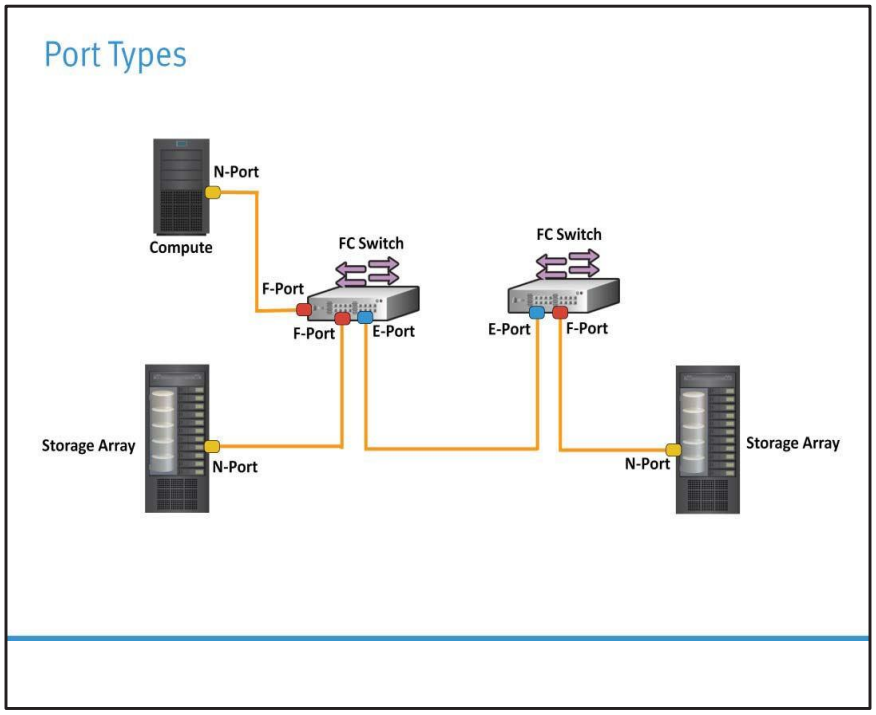
Directors are high end switches with a higher port count and better fault tolerance capabilities.

The fundamental purpose of a SAN is to provide compute access to storage resources. The large storage capacities offered by modern storage arrays have been exploited in SAN environments for storage consolidation and centralization.

A SAN management software manages the interfaces between compute systems, interconnect devices, and storage arrays. The software provides a view of the SAN environment and enables management of various resources from one central console.

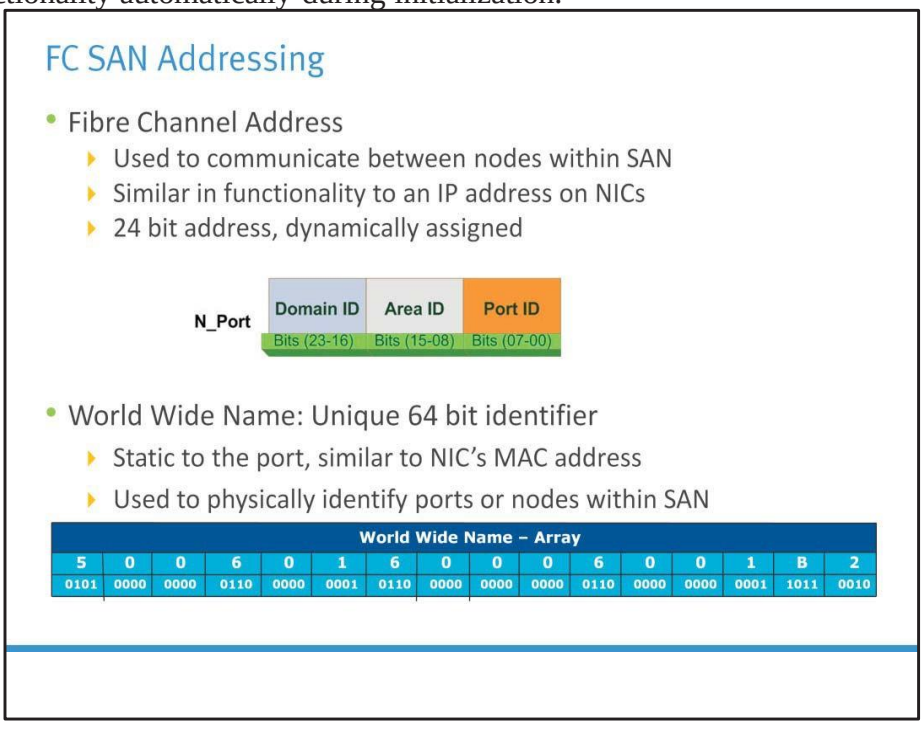


The Fibre Channel Fabric is a logical space in which all nodes communicate with one another through an FC switch or multiple interconnected FC switches. If an FC Fabric involves multiple switches, they are linked together through an FC cable. In a switched fabric, the link between any two switches is called Inter Switch Link (ISL). ISLs allow for two or more Fibre Channel switches to be connected together to form a single, but larger, fabric. ISLs are used to transfer from servers to storage arrays and the fabric management traffic from one switch to another. By using inter-switch links, a switched fabric can be expanded to connect a large number of nodes.



Compute systems communicate with storage devices through specialized outlets called ports. Ports are the basic building blocks of an FC network. Ports in the network can be one of the following types:

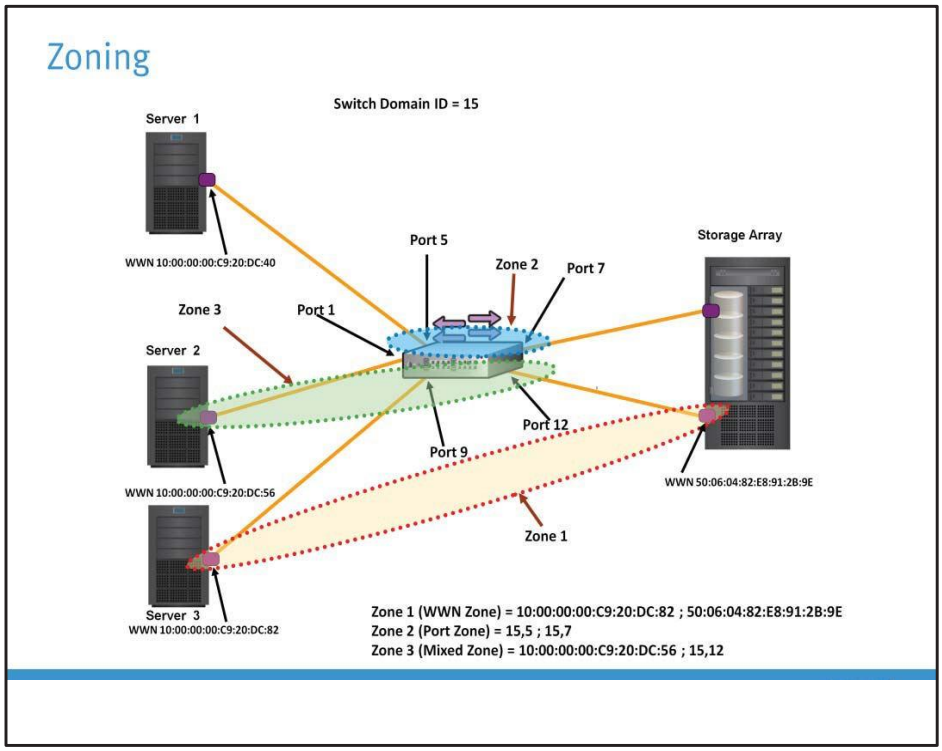
- N_port: An end point in the fabric. This port is also known as the node port. Typically, it is a compute system port (HBA) or a storage array port that is connected to a switch in a switched fabric.
- E_port: An FC port that forms the connection between two FC switches. This port is also known as the expansion port. The E_port on an FC switch connects to the E_port of another FC switch in the fabric through an ISL.
- F_port: A port on a switch that connects an N_port. It is also known as a fabric port.
- G_port: A generic port that can operate as an E_port or an F_port and determines its functionality automatically during initialization.



Two types of addresses are used for communication in an FC SAN environment – Channel Address and World Wide Name.

A Fibre Channel address is dynamically assigned when a node port logs on to FC SAN. The FC address is a 24 bit addressing scheme with three parts. The first field in the FC address contains the domain ID of the switch. A domain ID is a unique identification number provided to each switch in FC SAN. The area ID is used to identify a group of switch ports used to connect nodes. An example of a group of ports with a common area ID would be a port card on the switch. The last field identifies the port within the group.

Each device in the FC environment is assigned a 64-bit unique identifier called the World Wide Name (WWN). The Fibre Channel environment uses two types of WWNs: World Wide Node Name (WWNN) and World Wide Port Name (WWPN). Unlike an FC address, which is assigned dynamically, a WWN is a static name for each device on an FC network. WWNs are similar to the Media Access Control (MAC) addresses used in IP networking. WWNs are burned into the hardware or assigned through software. Several configuration definitions in a SAN use WWN for identifying storage devices and HBAs.



Zoning is an FC switch function that enables nodes within the fabric to be logically segmented into groups that can communicate with each other. When a device (compute system or storage array) logs onto a fabric, it is registered with the name server. When a port logs onto the fabric, it goes through a device discovery process with other devices registered in the name server. The zoning function controls this process by allowing only the members in the same zone to establish these link-level services.

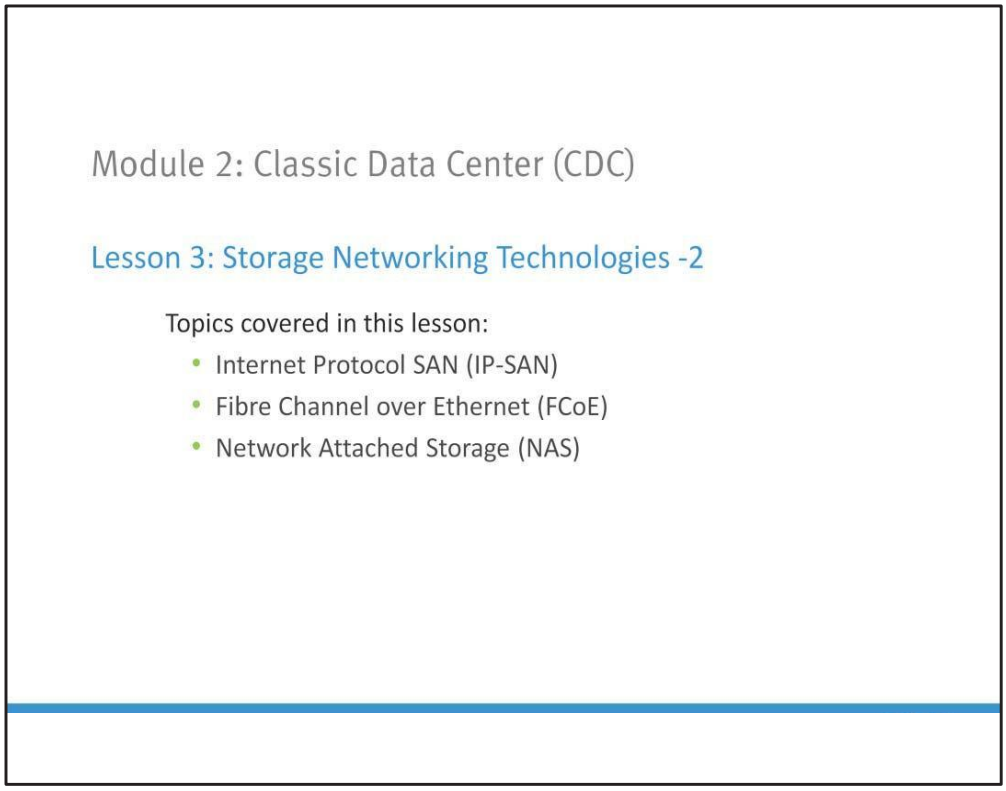
Zoning can be categorized into three types:

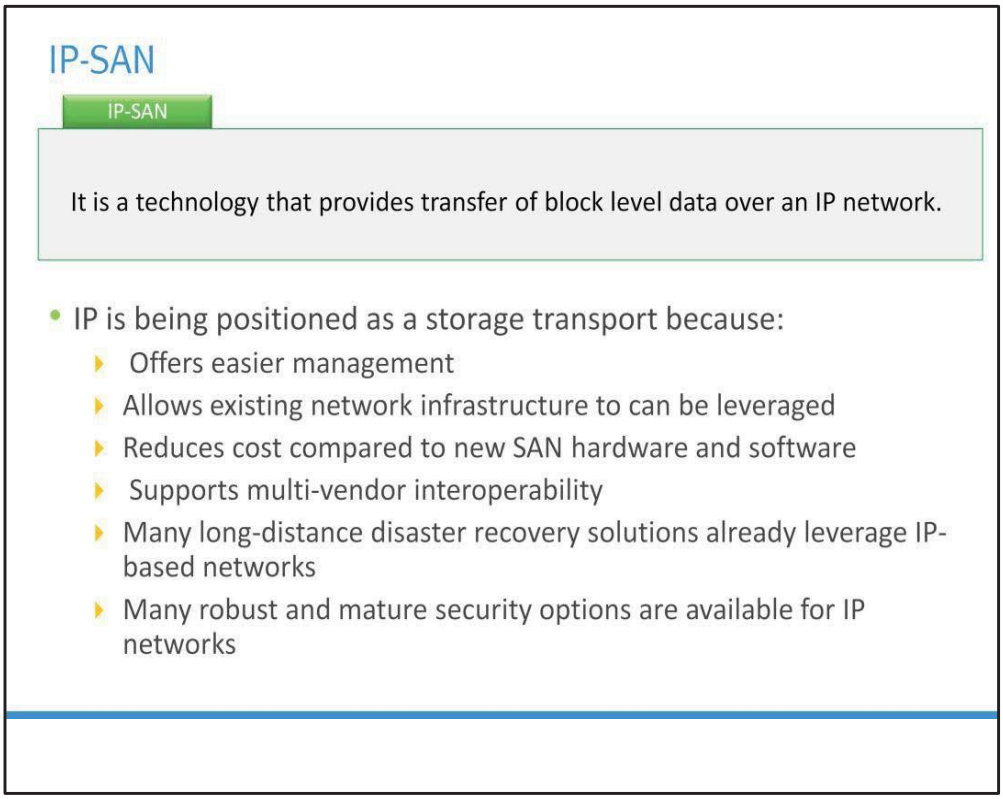
- Port zoning: It uses the FC addresses of the physical ports to define zones. In port zoning, access to data is determined by the physical switch port to which a node is connected. Port zoning is also called *hard zoning*.
- WWN zoning: It uses World Wide Names to define zones. WWN zoning is also referred to as *soft zoning*. A major advantage of WWN zoning is its flexibility. It allows the SAN to be

re-cabled without reconfiguring the zone information.

- Mixed zoning: It combines the qualities of both WWN zoning and port zoning. Using mixed zoning enables a specific port to be tied to the WWN of a node.

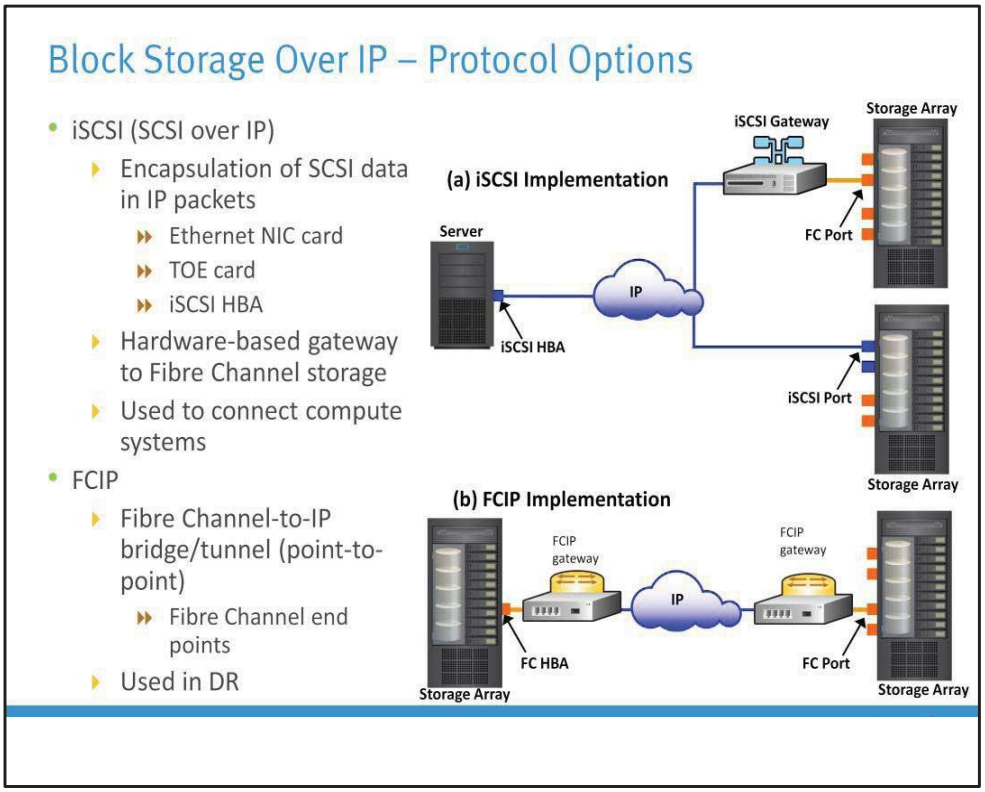
Zoning is used in conjunction with LUN masking to control server access to storage. However, these are two different activities. Zoning takes place at the fabric level and LUN masking is done at the array level.





Traditional SAN environments allow block I/O over Fibre Channel, whereas NAS environments allow file I/O over IP-based networks. Organizations require the performance and scalability of SAN plus the ease of use and lower Total Cost of Ownership of NAS solutions. The emergence of IP technology that supports block I/O over IP has positioned IP for storage solutions.

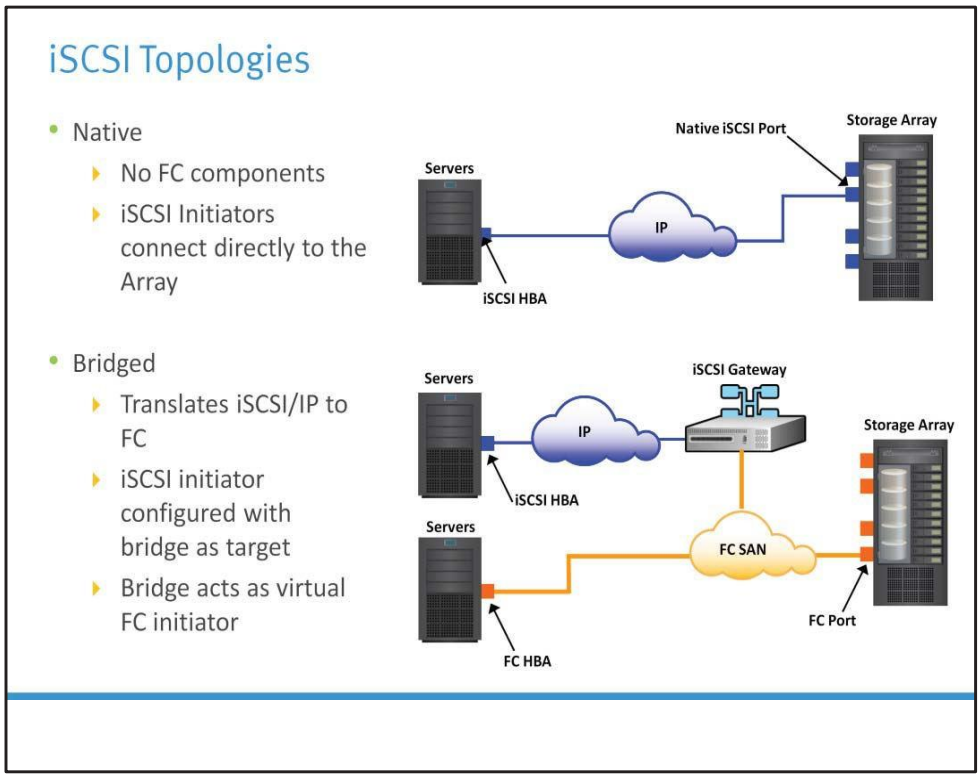
IP offers easier management and better interoperability. When block I/O is run over IP, the existing network infrastructure can be leveraged. This is more economical than investing in a new SAN hardware and software. Many long-distance, disaster recovery (DR) solutions are already leveraging IP-based networks. In addition, many robust and mature security options are now available for IP networks. With the advent of block storage technology that leverages IP networks (the result is often referred to as IP -SAN), organizations can extend the geographical reach of their storage infrastructure.



Two primary protocols that leverage IP as the transport mechanism for block level data transmission are iSCSI and Fibre Channel over IP (FCIP).

iSCSI is the compute-based encapsulation of SCSI I/O over IP using an Ethernet NIC card, TCP/IP offload engine (TOE) card, or an iSCSI HBA in the compute system. As illustrated in figure, IP traffic is routed over a network either to a gateway device that extracts the SCSI I/O from the IP packets or to an iSCSI storage array. The gateway can then send the SCSI I/O to an FC-based external storage array, whereas an iSCSI storage array can handle the extraction and I/O natively.

FCIP uses a pair of bridges (FCIP gateways) communicating over TCP/IP, which is the transport protocol. FCIP is used for extension of FC networks over distances using an existing IP-based infrastructure, as illustrated in the slide. Today, iSCSI is widely adopted for connecting compute systems to storage because it is relatively inexpensive and easy to implement, especially in environments where an FC SAN does not exist. FCIP is extensively used in Disaster Recovery(DR) implementations, where data is duplicated to an alternate site.

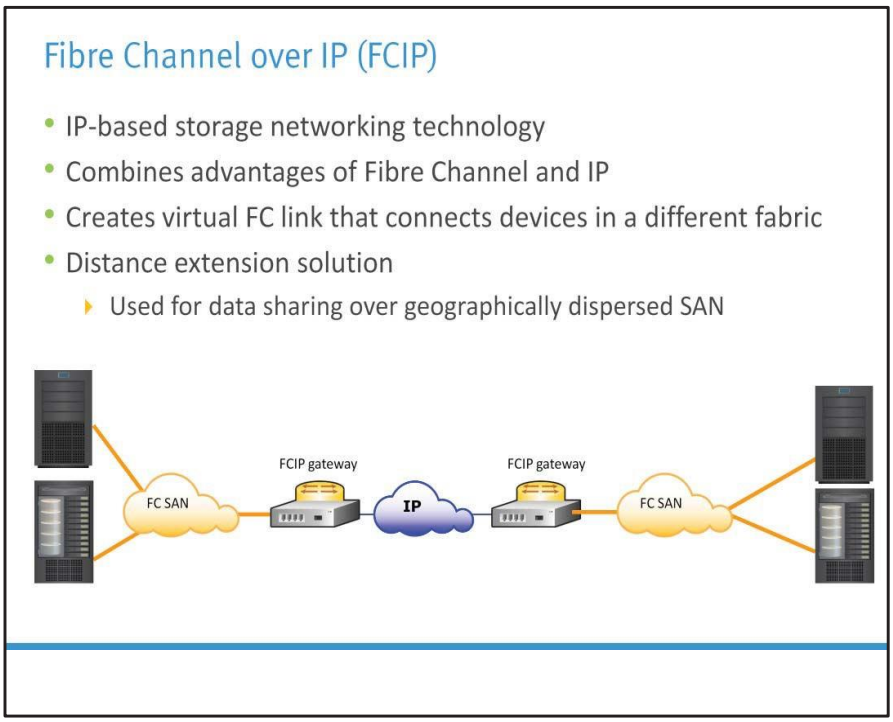


The topologies used for implementation of iSCSI can be categorized into two classes: Native and Bridged.

Native topologies do not have any FC components; they perform all communication over IP. The initiators (a system component that originates an I/O command over an I/O bus or network) may be either directly attached to targets or connected using standard IP routers and switches. If an iSCSI-enabled array is deployed, FC components are not required for iSCSI connectivity in the native topology. In the example shown in figure, the array has one or more Ethernet NICs that are connected to a standard Ethernet switch and configured with an IP address and listening port. Once a Client/Initiator is configured with the appropriate target information, it connects to the array and requests a list of available LUNs. A single array port can service multiple compute systems or initiators as long as the array can handle the amount of storage traffic that the compute system generates.

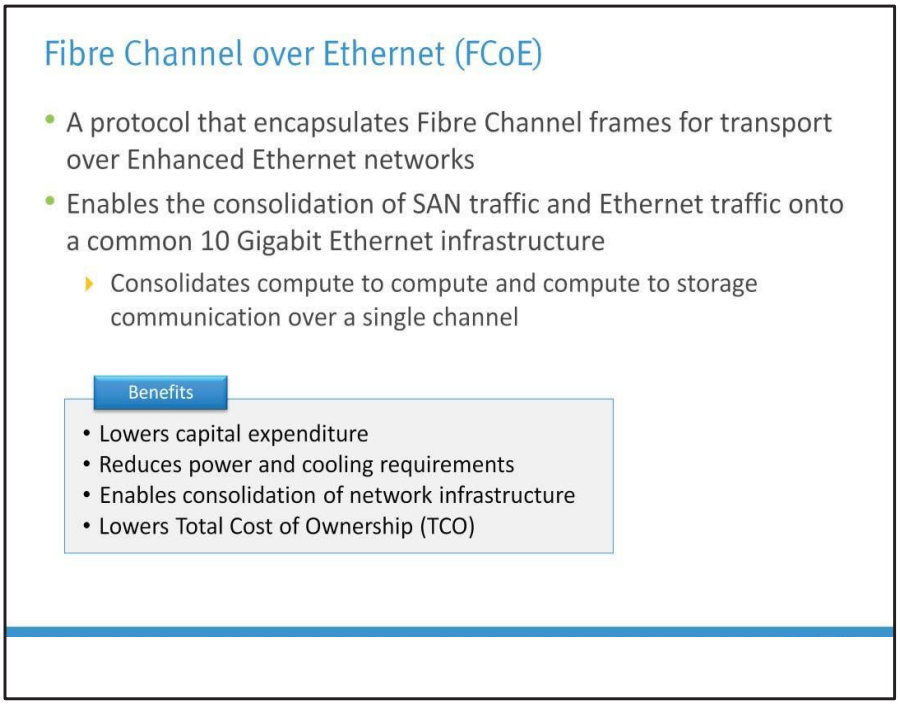
Many arrays provide more than one interface so that they can be configured in a highly available design or have multiple targets configured on the initiator. Some NAS devices are also capable of functioning as iSCSI targets, enabling file-level and block-level access to a centralized storage.

A bridged iSCSI implementation includes FC components in its configuration. The figure illustrates an existing FC storage array used to service compute systems connected through iSCSI. The array does not have any native iSCSI capabilities—that is, it does not have any iSCSI ports. As a result, an external device, called a bridge, router, gateway, or a multi-protocol router, must be used to bridge the communication from the IP network to the FC SAN. These devices can be a stand-alone unit, or in many cases, integrated with an existing FC switch. In this configuration, the bridge device has Ethernet ports connected to the IP network, and FC ports connected to the storage. These ports are assigned IP addresses, in the same way as on an iSCSI-enabled array. The iSCSI initiator/compute system is configured with the bridge's IP address as its target destination. The bridge is also configured with an FC initiator or multiple initiators. These are called virtual initiators because there is no physical device, such as an HBA, to generate the initiator record.



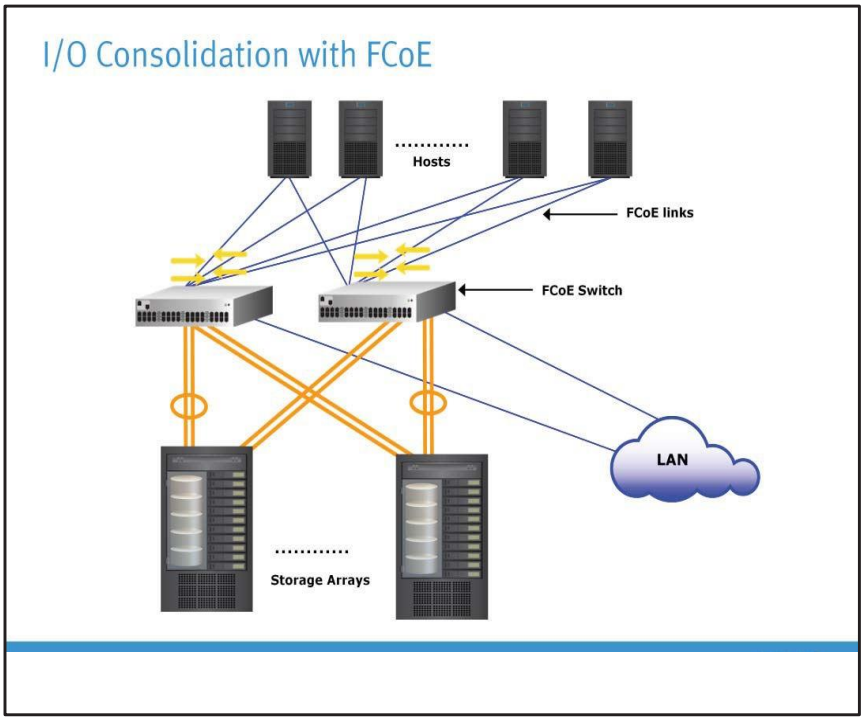
Organizations are now looking for new ways to transport data throughout the enterprise, locally over the SAN as well as over longer distances, to ensure that data reaches all the users who need it. One of the best ways to achieve this goal is to interconnect geographically dispersed SANs through reliable, high-speed links. This approach involves transporting FC block data over the existing IP infrastructure used throughout the enterprise. The FCIP standard has rapidly gained acceptance as a manageable, cost effective way to blend the best of the two worlds: FC block-data storage and the proven, widely deployed, IP infrastructure.

FCIP is a tunneling protocol that enables FC data to be sent over IP connection. It enables distributed FC SAN islands to be transparently interconnected over existing IP- based local, metropolitan, and wide-area networks. As a result, organizations now have a better way to protect, store, and move their data while leveraging investments in existing technology.

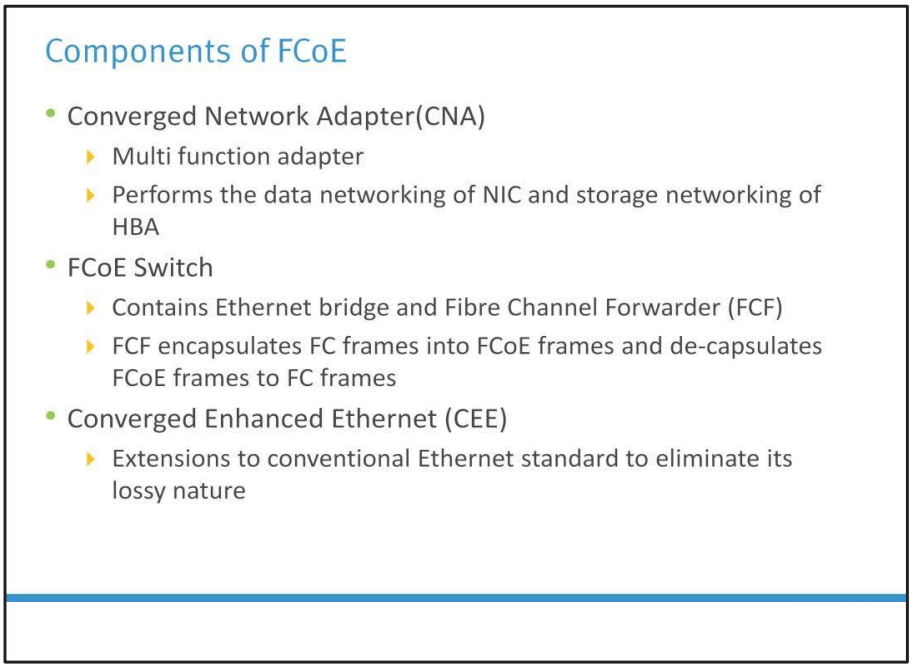


FCoE enables SAN traffic to be natively transported over Ethernet networks, while protecting and extending the investment that enterprises have made in storage networks. FCoE basically enables organizations to continue to run Fibre Channel over the same wires as their data networks. Unlike other storage networking protocols that use Ethernet, FCoE utilizes a new version of the Ethernet standard that makes it more reliable. The new "Enhanced" Ethernet is known as Converged Enhanced Ethernet (or CEE). FCoE combined with 10 Gigabit Ethernet (10 Gbps) fabrics grant organizations the ability to consolidate their I/O, cables, and adapters, while at the same time, increase the utilization of their servers. It combines LAN and SAN traffic over a single 10Gb Ethernet connection.

The benefits of FCoE include lower capital and operating costs and lower power and cooling requirements. This results in a lower total cost of ownership. FCoE enables input/output consolidation by allowing LAN and SAN traffic to converge on a single cable or link. It reduces the number of server cables, adapters, and switch ports in the data center and greatly simplifies the physical infrastructure. It also reduces the administrative overhead and complexity associated with managing the data center.



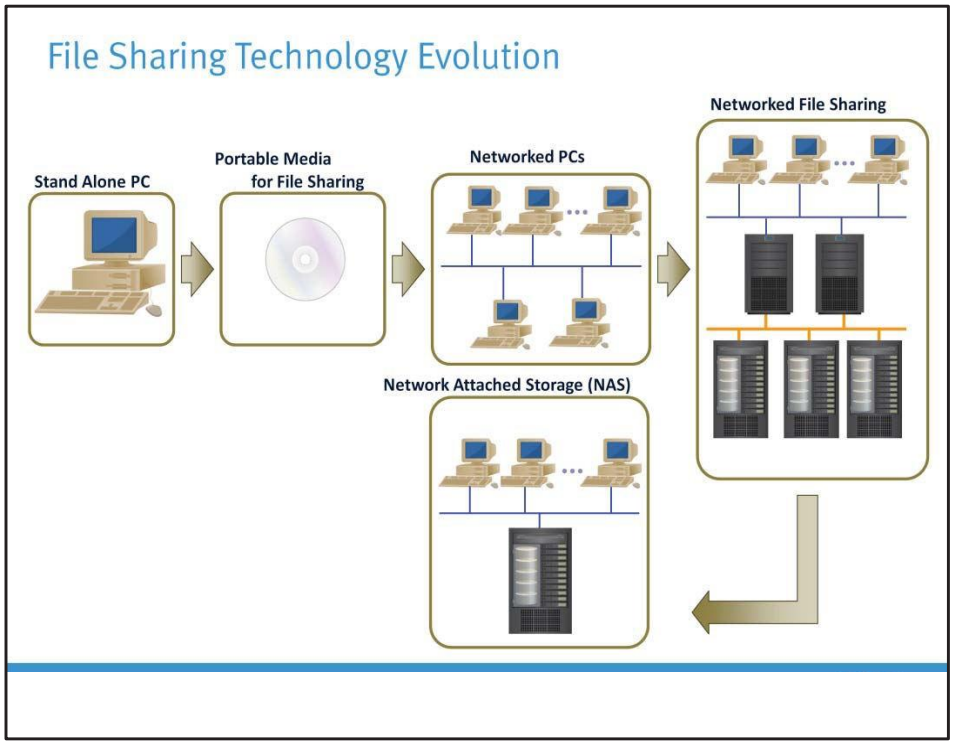
Shown is the I/O consolidation with FCoE using FCoE switches and CNAs. The FCoE switch passes the Fibre Channel traffic to SAN, and the Ethernet traffic to an attached Ethernet network. With FCoE, the cable requirements from the host to FCoE switches can be substantially reduced, which, in turn, reduces the cooling costs, management requirements, and the overall operational cost.



The CNA (Converged Network Adapter) is a multi function adapter which consolidates the data networking of an NIC card and the storage networking of a Fibre Channel HBA onto a single adapter. It eliminates the need of separate interface cards for FC and IP network. It consolidates the I/O into a single 10 Gigabit Ethernet link. An FCoE switch contains both an Ethernet switch function (Ethernet Bridge) and a Fibre channel switch function (Fibre Channel Forwarder). The main function of the Fibre Channel Forwarder (FCF) is that it encapsulates FC frames into FCoE frames and de-capsulates FCoE frames to FC frames. If the destination of an FCoE frame is reachable through the Ethernet port, the frame is sent to the Ethernet bridge, which forwards it to the appropriate Ethernet port. If the destination is reachable through the FC port, then the frame is sent to the FCF which decapsulates the frame and sends it to the appropriate FC port.

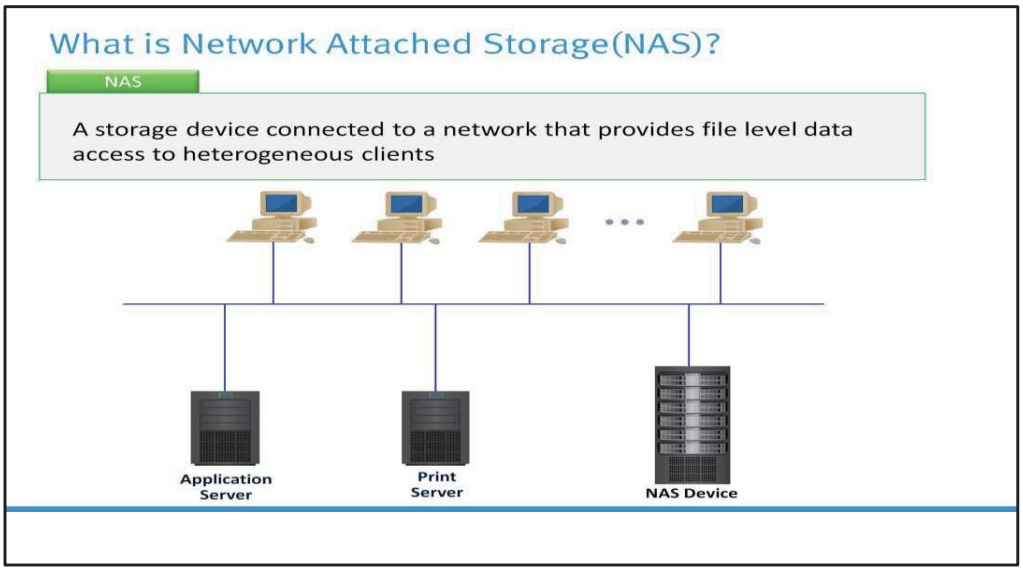
There are currently two options available for FCoE cabling: copper based Twinax and standard optical. A Twinax cable is composed of two pairs of copper cables that are covered with a shielded casing. Twinax cables require less power and are less expensive, but they can be used only for very short distances. The SFP+ connector is the primary connector used for FCoE links and can be used with both optical and copper cables.

Conventional Ethernet is lossy in nature, which means it will lose frames during transmission. Converged Enhanced Ethernet (CEE), lossless Ethernet, or Data Center Ethernet specifies new extensions to the existing Ethernet standards. These specifications eliminate the lossy nature of Ethernet and makes 10Gb Ethernet a viable storage networking option, similar to FC. There are many functionalities required for lossless Ethernet/Converged Enhanced Ethernet (CEE). These functionalities are defined and maintained by data center bridging (DCB) Task Group (TG), which is a part of the IEEE 802.1 working group.



In the past, floppy drives with capacities in mere KBs were widely used to share data files. Over time, the need for a larger capacity has emerged due to the growing need for data to be shared across organizations. Removable storage media, such as flash drives, capable of storing gigabytes (GB) of data, have now complemented the traditional removable media drives.

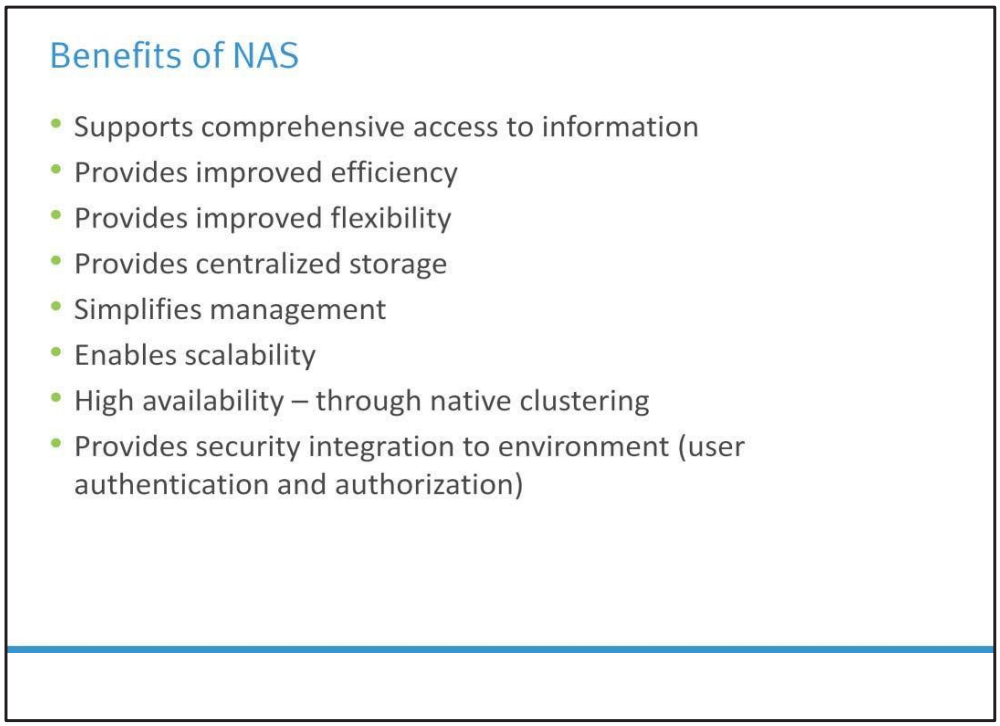
Businesses not only need the capacity to handle huge data storage requirements, but also need to share their data. This has made Network Attached Storage (NAS) an attractive option. NAS systems use external storage for the servers, which adds more flexibility for network storage.



NAS is a dedicated high performance file server with storage system. It provides file-level data access and sharing. NAS is a preferred storage solution that enables clients to share files quickly and directly with minimum storage management overhead. NAS also helps to eliminate bottlenecks that users face when accessing files from a general-purpose server.

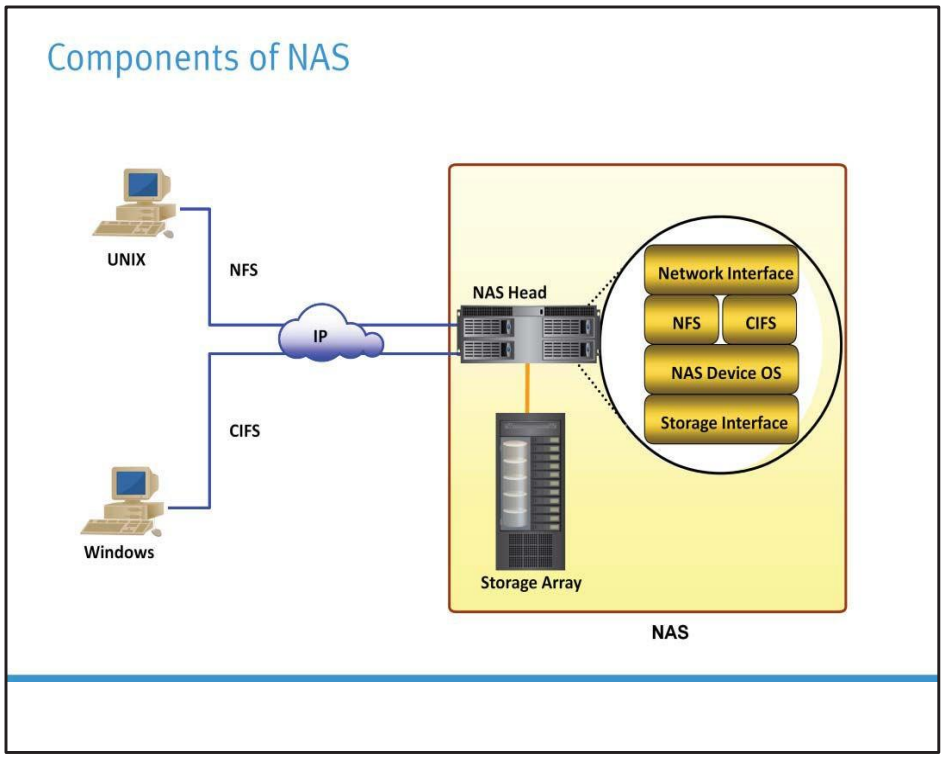
NAS uses network and file sharing protocols, which include TCP/IP for data transfer and CIFS and NFS for remote file services. Recent advancements in networking technology have enabled NAS

to scale up to enterprise requirements for improved performance and reliability in accessing data. NAS serves a mix of clients and servers over an IP network. A NAS device uses its own operating system and integrated hardware and software components to meet specific file serving needs. As a result, a NAS device can serve more clients than traditional file servers, and can provide the benefit of server consolidation.



NAS offers the following benefits:

- Supports comprehensive access to information: Enables efficient file sharing and supports many-to-one and one-to-many configurations. The many-to-one configuration enables a NAS device to serve many clients simultaneously. The one-to-many configuration enables a single client to connect with many NAS devices simultaneously.
- Improved efficiency: Eliminates bottlenecks that occur during file access from a general-purpose file server because NAS uses an operating system specialized for file serving. It improves the utilization of general purpose servers by relieving them of file server operations.
- Improved flexibility: Compatible for clients on both UNIX and Windows platforms using industry-standard protocols.
- Centralized storage: Centralizes data storage to minimize duplication of data on client workstations.
- Simplified management: Provides a centralized console that makes it possible to manage file systems efficiently.
- Scalability: Scales well in accordance with different utilization profiles and types of business applications because of the high performance and low-latency design.
- High availability: Offers efficient replication and recovery options, enabling high data availability. NAS uses redundant networking components that provide maximum connectivity options. A NAS device can use clustering technology for failover.
- Security: User authentication and file locking in conjunction with industry- standard security schemas.



NAS has the following components:

- NAS head (CPU and Memory)
- One or more network interface cards (NICs) that provide connectivity to the network
- An optimized operating system for managing NAS functionality
- Network File System (NFS) and Common Internet File System (CIFS) protocols for file sharing. NFS is predominantly used in UNIX based operating environments; CIFS is used in Microsoft Windows based operating environments. These file sharing protocols enable users to share file data across different operating environments.
- Industry-standard storage protocols to connect and manage physical disk resources, such as ATA, SCSI, FC and so on.
- Storage Array

The NAS environment includes clients accessing a NAS device over an IP network using standard protocols.

Module 2: Classic Data Center (CDC)

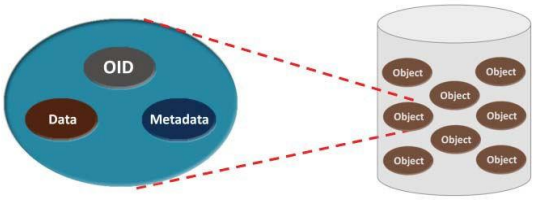
Lesson 4: Object Based and Unified Storage Technologies

Topics covered in this lesson:

- Object Based Storage
- Unified Storage

Object Based Storage

- Object Based Storage combines data with rich metadata to create an "object"
- Object Based Storage stores data in a flat address space
  - There are no hierarchies or nested directories
- Each object is identified by an unique ID (Object ID)
  - Generated by a hashing function



Object Based Storage combines data with rich metadata to create an "object." For example, when an MRI scan is stored as a file in a NAS system, the metadata is basic and may include information such as file name, creation data, creator, and file type. When stored as an Object, on the other hand, the MRI scan can have the basic metadata plus additional metadata, such as the patient's name, the patient's ID, the procedure date, the attending physician's name, and provide pointers to files that contain the physician's notes. Object Based Storage stores data in a flat address space. There are no hierarchies or nested directories. As a result, there are no limits on the number of files that can be stored. The storage capacity can be easily scaled from terabytes to petabytes. Each object in Object Storage is identified by a unique ID called object ID. This ID is generated using a hash function and guarantees that every object is uniquely identified. It also serves as a unique pointer to an object similar to the way URLs point to unique files in the Internet.

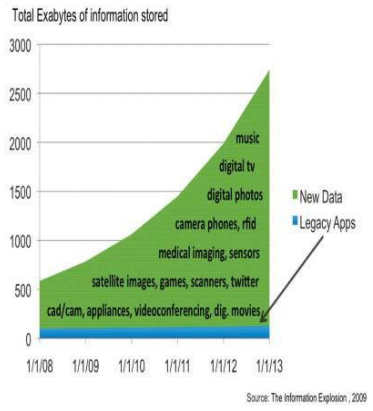
Object Based Storage (contd.)

- Object Based Storage uses HTTP communication as its standard interface
  - SOAP and REST are the protocols commonly used in object based communication in Cloud
- Simple Object Access Protocol (SOAP) is used for communication between peers in a distributed environment
  - Uses Extensible Markup Language (XML) framework
- Representational State Transfer (REST) is used to retrieve information from a Website by reading Web pages

Object Storage uses HTTP communication as its standard interface. This makes it ideal for communicating data transfers with storage locations via an Intranet and the Internet. SOAP is a way of exchanging messages between peers on a distributed network, such as the Internet. SOAP provides a set of XML elements and attributes, which are used to construct a "SOAP" message. The REST protocol is used to retrieve information from a Website by reading Web pages that include the XML file. XML statements describe the information that is available on a Web page. This information can then be accessed by typing the Uniform Resource Locator(URL) or of the Web page. Both SOAP and REST place their messages within an HTTP message before transmitting them.

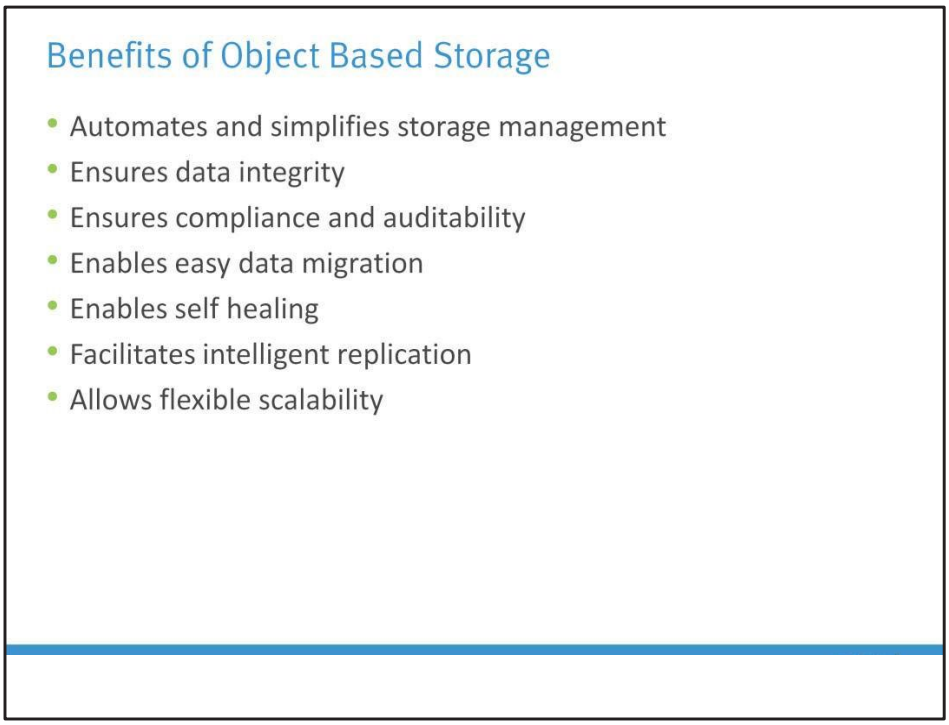
Why Object Based Storage ?

- Increasing amount of unstructured data
  - SAN is highly scalable and supports data access at a block level
    - Not a good option for data sharing
  - NAS is a good option for applications which need to share data
    - Limited scalability due to hierarchical structure
  - Object Based approach potentially eliminates SAN and NAS limitations
    - Highly scalable with data sharing capabilities



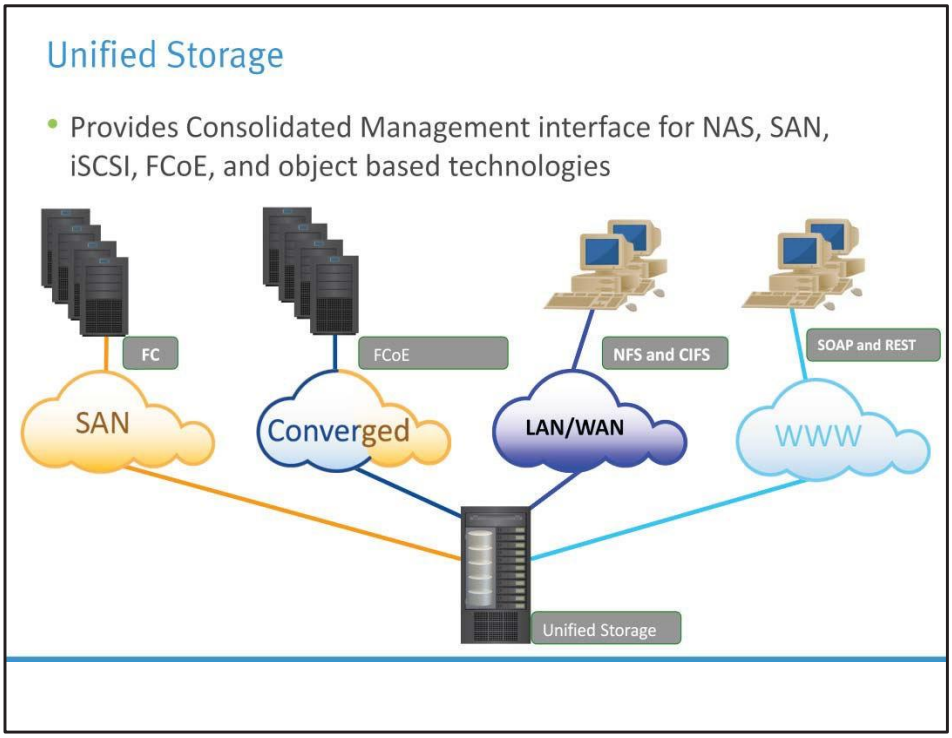
Unstructured data is data that has no specific structure or schema that dictates how it is stored,

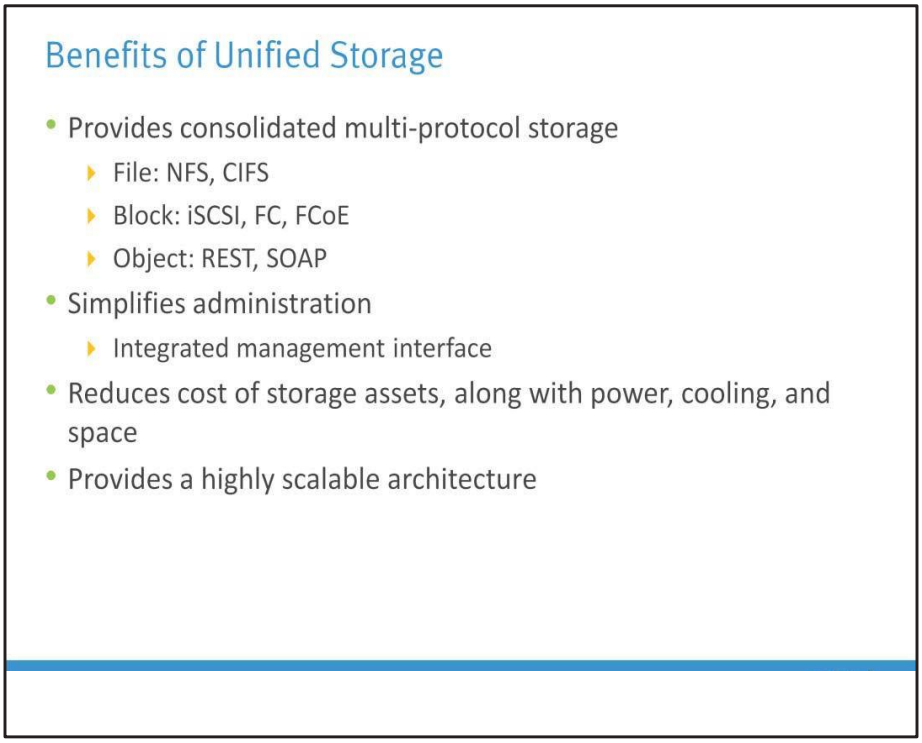
formatted, accessed, or organized within the system. Examples of this include text documents, contracts, pictures, movies, music, blogs etc. The majority of information created today falls into this category. The SAN architecture is highly scalable, but is not a recommended option for applications which need to share data. Historically, unstructured data has been stored in NAS file systems. File systems organize data into a hierarchal structure of directories or folders and files. This system works well for smaller data sets, but becomes problematic and costly as the amount of data grows. Typical problems include performance, scaling, data portability, and ongoing management costs. These were the main reasons for the emergence of object based storage technologies. They have a highly scalable architecture with good data sharing capabilities.



Reading the rich metadata attached to each object, storage administrators can automatically apply policies for storage management according to the file contents. Further, Object Based Storage has a flat address space, which negates the need for managing LUNs and RAID groups. A hash function is used to generate object ID for each object, which serves as a powerful mechanism for compliance and auditability. This is because the resulting digest of a hash function, its digital signature, will change if the data file is changed; consequently an unchanged signature provides proof of data integrity as well as its authenticity. This makes Object based Storage useful for protecting archived data, meeting regulatory requirements, and for data that has high legal or compliance risk. Leveraging each object's unique ID, storage administrators can easily migrate files from one storage system to another, as required. This is because all that is required to retrieve an object is its ID; its particular location is not relevant. By monitoring the state of each signature, an Object based Storage system can tell if a file has been corrupted through a change in its signature. With this knowledge, the system can "self heal" and replace any corrupted files. Instead of RAID systems for backup protection, storage administrators can leverage Object Based Storage to easily create multiple replicas, as necessary, with each replica identified with the same object storage ID. These replicas can further be distributed geographically, increasing the protection provided by replication. Through its use of a flat address space, Object Based Storage enables flexible scalability in terms of capacity and numbers of objects. This benefit is particularly valuable in cases where the exact storage requirements, for example like those in Cloud Storage, are unknown. Some scenarios where object based storage systems may be extensively used are multimedia content rich Web applications,

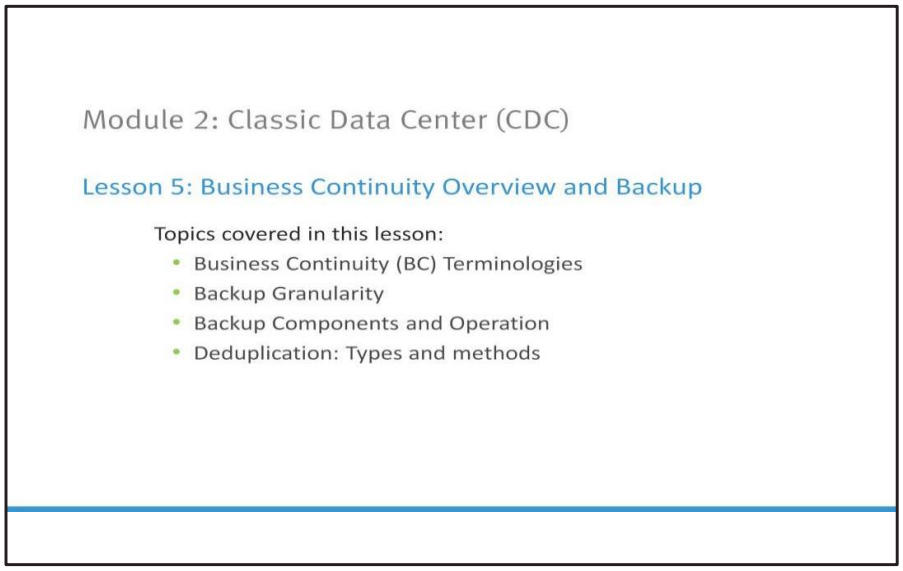
archives, Cloud and so on.

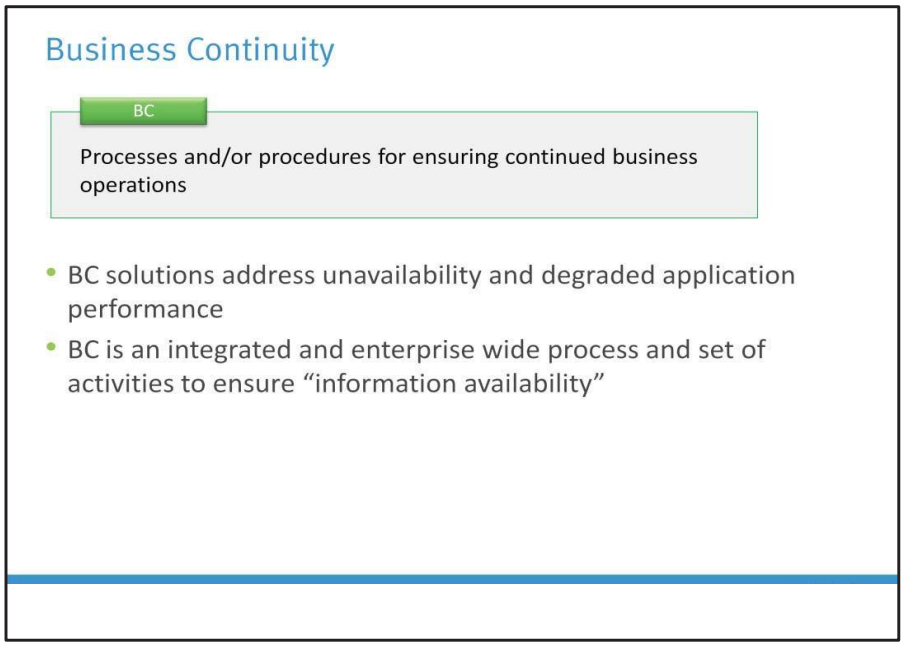




Unified storage consolidates NAS-based, SAN-based, and object based access within one unified platform. It supports NAS protocols (CIFS for Windows and NFS for UNIX/Linux), iSCSI, Fibre Channel, FCoE, REST, and SOAP protocols. The ability to serve multiple protocols from the same storage system helps to freely mix and match workloads to greatly improve utilization. Having a single data model and toolset for unified storage enables a consistent management framework across many applications and workloads. This greatly simplifies administration and creates a hierarchy of

values from management of physical storage to application-level integration. Unified storage provides storage consolidation. This reduces the storage requirements of the organization, which, in turn, lowers the cost of acquiring storage assets, power, cooling, and space. Unified storage provides a highly scalable architecture that can be scaled from workgroup to full enterprise systems.





In today's world, continuous access to information is a must for the smooth functioning of business operations. The cost of unavailability of information is greater than ever, the outages in key industries costing millions of dollars per hour. As a result, it is critical for businesses to define appropriate strategies that can help them overcome these crises.
Business continuity is an important process to define and implement these plans.

BC is an integrated and enterprise-wide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime. BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. The goal of a business continuity solution is to ensure the "information availability" required to conduct vital business operations.

BC Terminologies

BC Terminologies	Description
Disaster Recovery(DR)	Coordinated process of restoring systems, data, and infrastructure required to support ongoing business operations in the event of a disaster
Hot Site	• A site where an enterprise's operations can be moved in the event of disaster • DR site infrastructure is up and running all the times
Cold Site	The IT infrastructure required to support DR is not activated
Cluster	A group of servers and other necessary resources, coupled to operate as a single system

DR is the coordinated process of restoring systems, data, and the infrastructure, required to support key ongoing business operations in the event of a disaster. It is the process of restoring and/or resuming business operations from a consistent copy of the data. After all recoveries are completed, the data is validated to ensure that it is correct.

Hot site**:** A site to where an enterprise's operations can be moved, in the event of a disaster. It is a site equipped with all the required hardware, operating system, application, and network support that help perform business operations, and where the equipment is available and running at all times.

Cold site: A site to where an enterprise's operations can be moved, in the event of disaster. It has minimum IT infrastructure and environmental facilities in place, but are not activated.

Cluster: A group of servers and other necessary resources, coupled to operate as a single system. Clusters ensure high availability and load balancing. Typically, in failover clusters, one server runs an application and updates the data, and the other is kept as a standby to take over completely, when required. In more sophisticated clusters, multiple servers may access data, while typically, one server is kept as a standby.

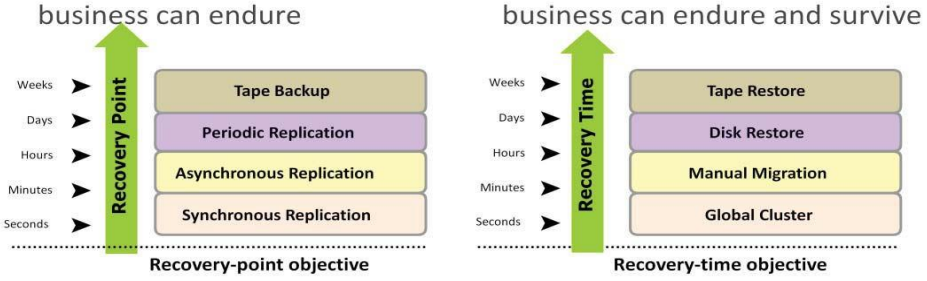
RTO and RPO

Recovery Point Objective (RPO)

- Point in time to which systems and data must be recovered after an outage
- Amount of data loss that a business can endure

Recovery Time Objective (RTO)

- Time within which systems, applications, or functions must be recovered after an outage
- Amount of downtime that a business can endure and survive



Recovery-Point Objective (RPO): This is the point in time to which systems and data must be recovered after an outage. It defines the amount of data loss that a business can endure. A large RPO signifies high tolerance to information loss in a business. Based on the RPO, organizations plan for the minimum frequency with which a backup or replica must be made. For example, if the RPO is six hours, backups or replicas must be made at least once in 6 hours. The figure shows various RPOs and their corresponding ideal recovery strategies. An organization may plan for an appropriate BC technology solution on the basis of the RPO it sets. For example, if RPO is 24 hours, that means that backups are created on an offsite tape drive every midnight. The corresponding recovery strategy is to restore data from the set of last backup tapes. Similarly, for zero RPO, data is mirrored synchronously to a remote site.

Recovery-Time Objective (RTO): The time within which systems, applications, or functions must be recovered after an outage. It defines the amount of downtime that a business can endure and survive. Businesses can optimize disaster recovery plans after defining the RTO for a given data center or network. For example, if the RTO is two hours, then use a disk backup because it enables a faster restore than a tape backup. However, for an RTO of one week, tape backup will most likely meet requirements. Few examples of RTOs and the recovery strategies to ensure data availability are listed below:

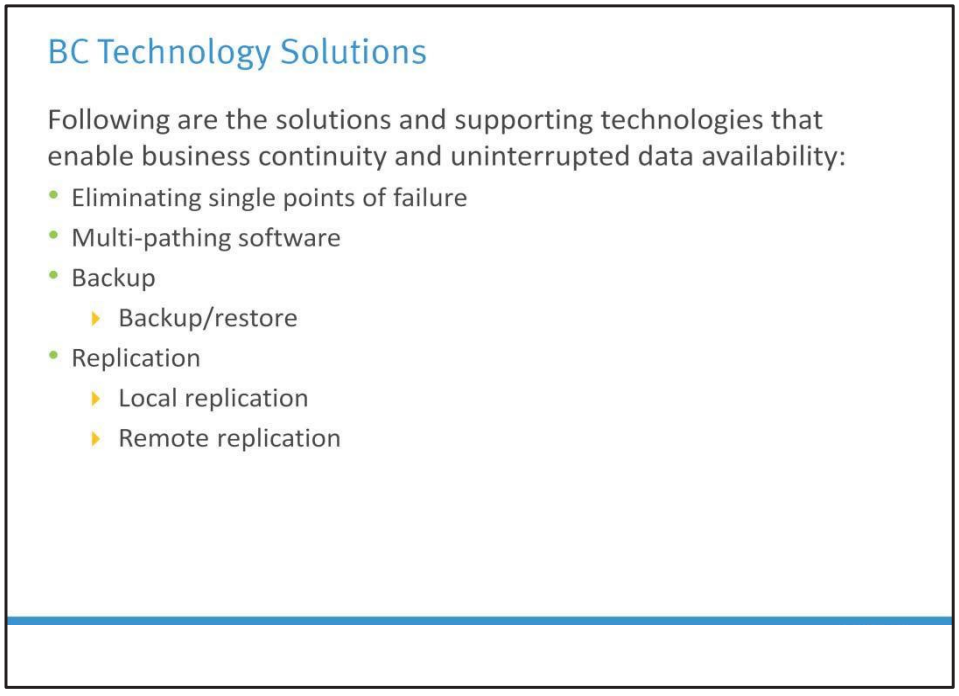
RTO of 72 hours: Restore from backup tapes at a cold site.

RTO of 12 hours: Restore from tapes at a hot site.

RTO of 4 hours: Use a data vault to a hot site.

RTO of 1 hour: Cluster production servers with controller-based disk mirroring.

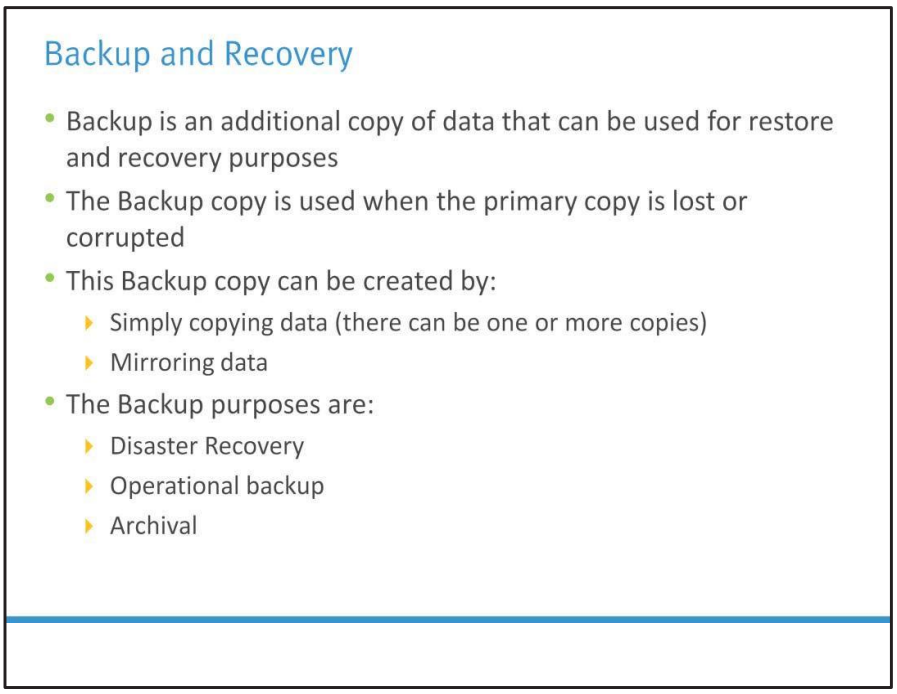
RTO of a few seconds: Cluster production servers with bi-directional mirroring, enabling the applications to run at both sites simultaneously.



A single point of failure refers to the failure of a single component, which can terminate the availability of the entire system or an IT service. In a setup where each component must function as required to ensure data availability, the failure of a single component causes the failure of the entire data center or an application, resulting in the disruption of business operations. There are several single points of failure in a CDC. The single HBA on the compute, the compute itself, the IP network,

the FC switch, the storage array ports, or even the storage array could become potential single points of failure. To mitigate a single point of failure, systems are designed with redundancy, such that the system fails only if all the components in the redundancy group fail. This ensures that the failure of a single component does not affect data availability. For example, implementing server clustering to mitigate the impact of a compute failure.

Configuration of multiple paths increases the data availability through path failover. If servers are configured with one I/O path to the data, there will be no access to the data if that path fails. Redundant paths eliminate the path from becoming single points of failure. Multiple paths to access data also improves I/O performance through load sharing and maximizes server, storage, and data path utilization. In practice, merely configuring multiple paths does not serve the purpose. Even with multiple paths, if one path fails, I/O will not be rerouted unless the system recognizes that it has an alternate path. Multipathing software provides the functionality to recognize and utilize alternate I/O path to data. Multipathing software also manages load balancing by distributing I/Os to all available, active paths.



Backup is a copy of the production data, created and retained for the sole purpose of recovering deleted or corrupted data. With growing business and regulatory demands for data storage, retention, and availability, organizations are faced with the task of backing up an ever-increasing amount of data. This task becomes more challenging as demand for consistent backup and quick restore of data increases throughout the enterprise – which may be spread over multiple sites. Moreover, organizations need to accomplish backup at a lower cost with minimum resources.

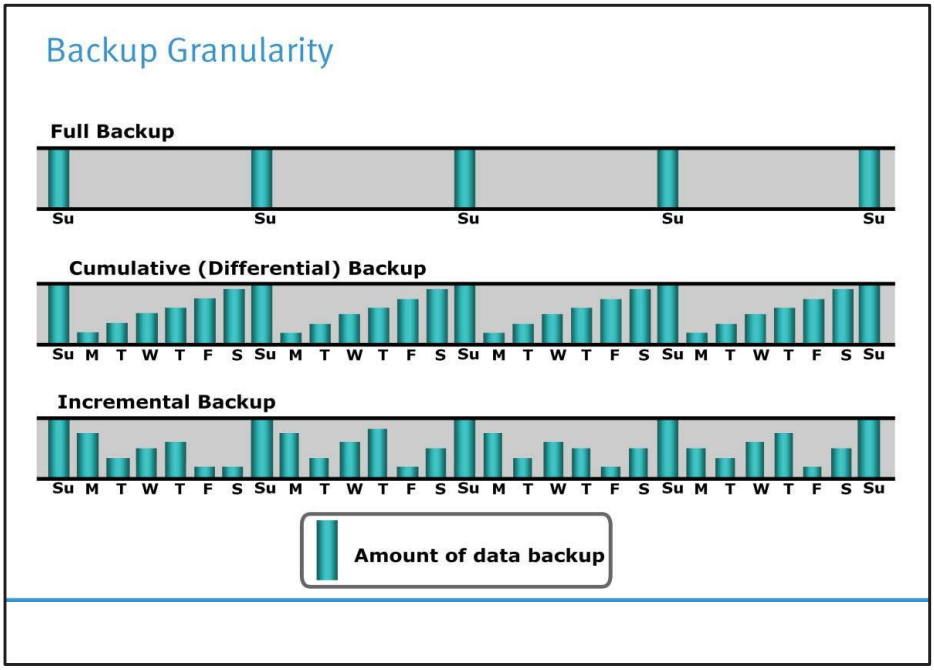
Organizations must ensure that the right data is in the right place at the right time. Evaluating backup technologies, recovery, and retention requirements for data and applications is an essential step to ensure successful implementation of the backup and recovery solution. The solution must facilitate easy recovery and retrieval from backups, as required by the business.

Backups are performed for three primary purposes: Disaster Recovery, Operational Restores, and Archival.

Disaster recovery addresses the requirement to be able to restore all, or a large part, of an IT infrastructure in the event of a major disaster. The backup copies are used for restoring data at an alternate site when the primary site is incapacitated due to a disaster. Based on RPO and RTO requirements, organizations use different backup strategies for disaster recovery.

Data in the production environment changes with every business transaction and operation. Operational backup is a backup of data at a point in time and is used to restore data in the event of data loss or logical corruptions that may occur during routine process. The majority of restore requests in most organizations fall in this category. For example, it is common for a user to accidentally delete an important e-mail or for a file to become corrupted. In such cases data can restored from the operational backup.

Archival is a common requirement used to preserve transaction records, email, and other business work products for regulatory compliance. The regulations could be internal, governmental, or perhaps derived from specific industry requirements. Backups are also performed to address archival requirements.



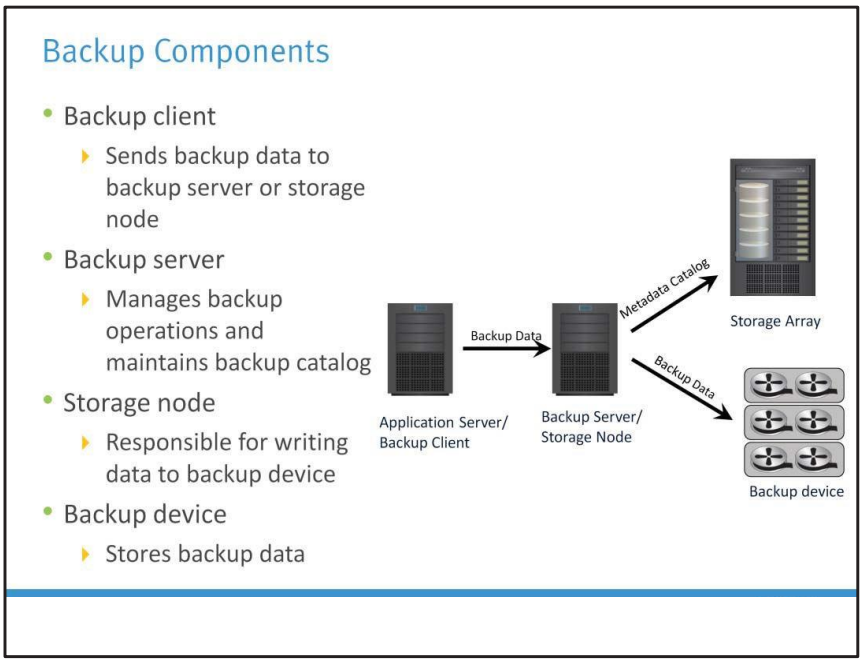
Backup granularity depends on business needs and the required RTO/RPO. Based on granularity, backups can be categorized as full, cumulative, and incremental. Most organizations use a combination of these three backup types to meet their backup and recovery requirements. The figure depicts the categories of backup granularity.

Full backup is a backup of the complete data on the production volumes at a certain point in time. A full backup copy is created by copying the data on the production volumes to a secondary storage device. Incremental backup copies the data that has changed since the last full or incremental backup, whichever has occurred more recently. This is much faster (because the volume of data backed up is restricted to changed data), but takes longer to restore.

Cumulative (or differential) backup copies the data that has changed since the last full backup.

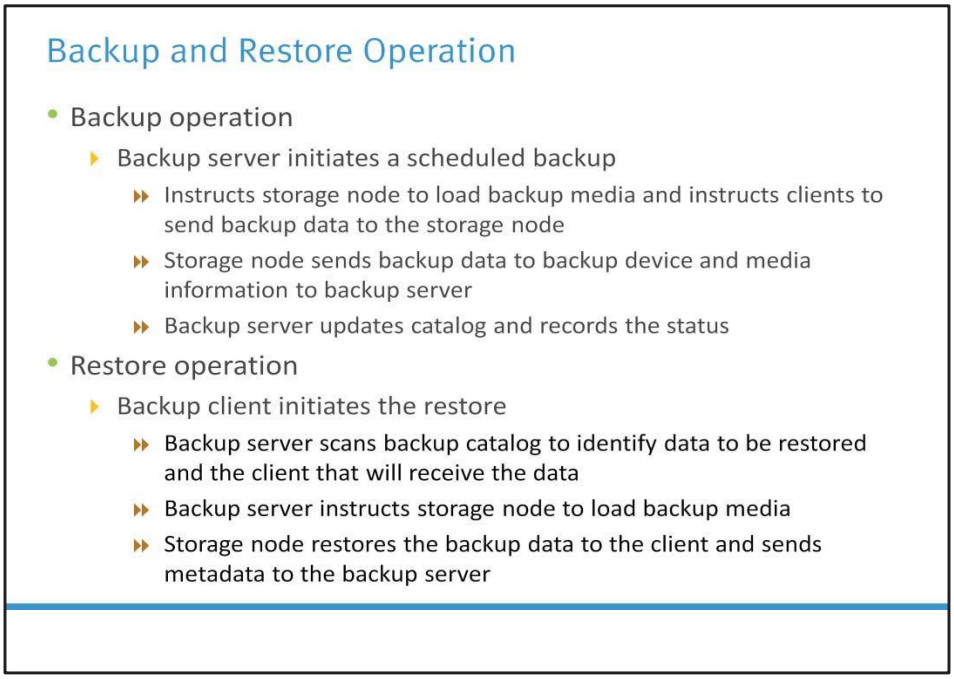
This method takes longer than incremental backup, but is faster to restore.

Synthetic (or constructed) full backup is another type of backup that is used in implementations where the production volume resources cannot be exclusively reserved for a backup process for extended periods. It is usually created from the most recent full backup and all the incremental backups performed thereafter. A synthetic full backup enables a full backup copy to be created offline without disrupting the I/O operation on the production volume. This also frees up network resources from the backup process, making them available for other production uses.



A backup system uses client/server architecture with a backup server and multiple backup clients. The backup server manages the backup operations and maintains the backup catalog, which contains information about the backup process and backup metadata. The backup server depends on the backup clients to gather the data to be backed up. The backup clients can be local to the server or can reside on another server, presumably to back up the data visible to that server. The backup server receives the backup metadata from the backup clients to perform its activities. The metadata is stored either locally within the backup server or externally in a storage array

The figure on the slide illustrates the backup process. The storage node is responsible for writing data to the backup device (in a backup environment, a storage node is a compute system that controls backup devices). Typically, the storage node is integrated with the backup server and both are hosted on the same physical platform. A backup device is attached directly to the storage node. Some backup architecture refers to the storage node as the media server because it connects to the storage device.

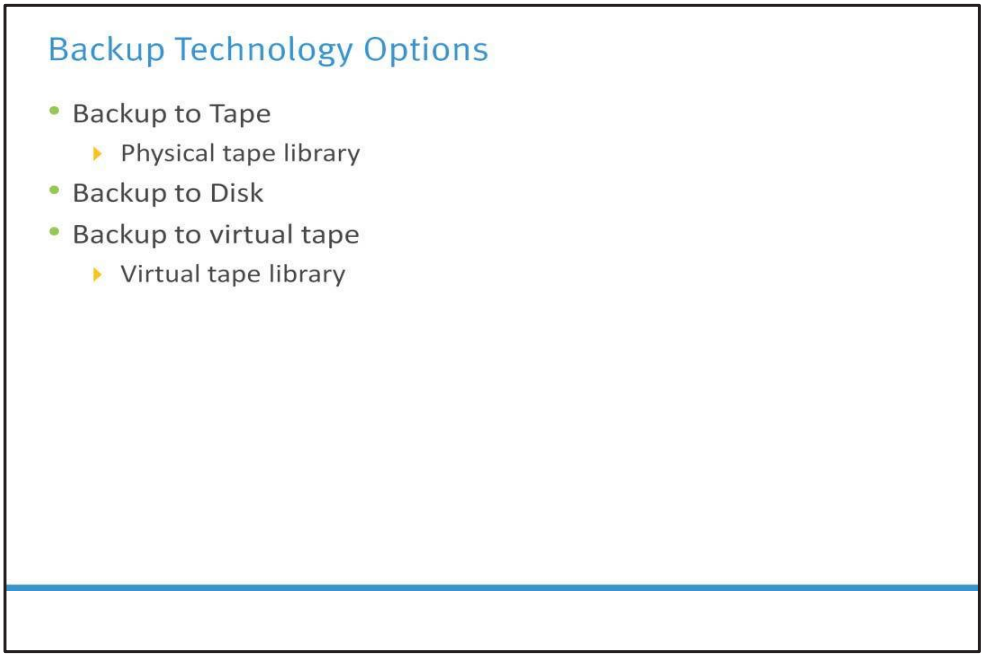


In a scheduled backup operation, the backup server initiates the backup process for different clients based on the backup schedule configured for them. The backup server coordinates the backup process with all the components in a backup configuration. The backup server maintains the information about the backup clients to be contacted and the storage nodes to be used in a backup operation. The backup server retrieves the backup-related information from the backup catalog and, based on this information, instructs the appropriate storage node to load the backup media into the backup devices. Simultaneously, it instructs the backup clients to send their metadata to the backup server and to back up the data to the appropriate storage node. On receiving this request, the backup client sends tracking information to the backup server. The backup server writes this metadata on its backup catalog. The backup client sends the data to the storage node, and the storage node writes the data to the storage device. The storage node also sends tracking information to the backup server to keep it updated about the media being used in the backup process.

A restore process is manually initiated by the backup client. Upon receiving a restore request, the user opens the restore application to view the list of clients that have been backed up.

While selecting the client for which a restore request has been made, the user also needs to identify the client that will receive the restored data. Data can be restored on the same client or on another client, given the proper permissions. The user then selects the data to be restored. Note that because the entire information comes from the backup catalog, the restore application must also communicate with the backup server.

The backup server identifies the backup media required for the restore and notifies the storage node to load the backup media. Data is then read and sent to the client that has been identified to receive the restored data.



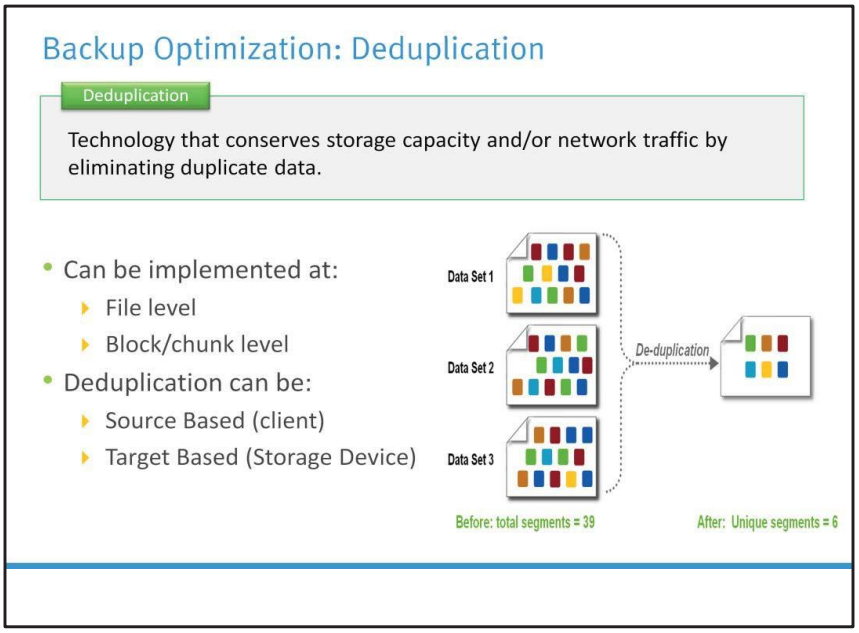
Tape drives, a low-cost option, are used extensively for backup. Tape drives are used to read/write data from/to a tape cartridge. Tape drives are referred to as sequential or linear access devices because the data is written or read sequentially. A physical tape library provides housing and power for a number of tape drives and tape cartridges.

Disks have now replaced tapes as the primary device for storing backup data, because of their performance advantages. Backup-to-disk systems offer ease of implementation, reduced cost, and improved quality of service. Apart from performance benefits in terms of data transfer rates, disks

also offer faster recovery when compared to tapes. Backing up to disk storage systems offers clear advantages due to their inherent random access and RAID- protection capabilities. In most backup environments, backup to disk is used as a staging area where the data is copied temporarily before transferring or staging it to tapes later. This enhances the backup performance. Some backup products allow for backup images to remain on the disk for a period of time even after they have been staged. This enables a much faster restore.

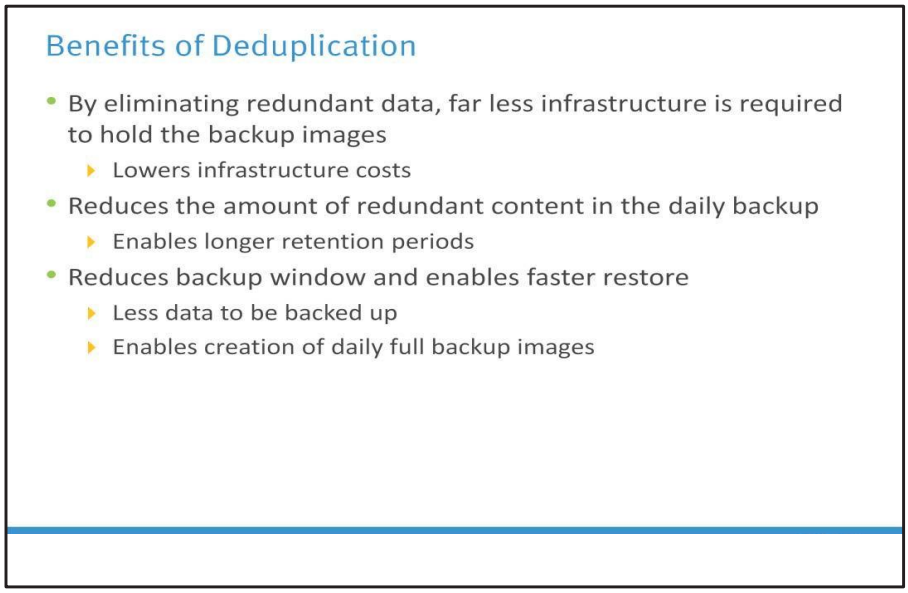
A virtual tape library (VTL) has the same components as that of a physical tape library, except that the majority of the components are presented as virtual resources. For a backup software, there is no difference between a physical tape library and a virtual tape library.

Virtual tape libraries use disks as backup media. Emulation software has a database with a list of virtual tapes, and each virtual tape is assigned a portion of a LUN on the disk. A virtual tape can span multiple LUNs, if required. File system awareness is not required while using backup to disk because virtual tape solutions use raw devices. Unlike a physical tape library, which involves mechanical delays, in a virtual tape library, response is almost instantaneous.

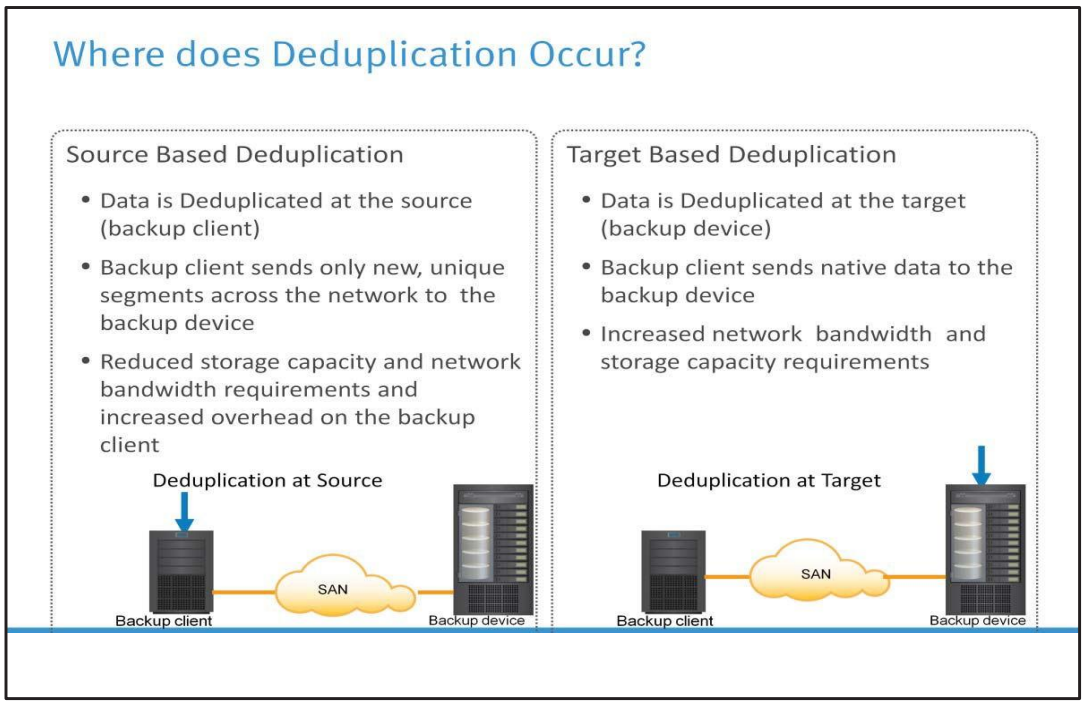


Deduplication refers to technology that searches for duplicate data (ex: blocks, segments) and discards duplicate data when located. When duplicate data is detected, it is not retained; instead, a "data pointer" is modified so that the storage system references an exact copy of that data already stored on disk. Furthermore, Deduplication alleviates the costs associated with keeping multiple copies of the same data. It also drastically reduces the storage requirements of a system. Some Deduplication solutions work at the file level, whereas some other Deduplication solutions work on a lower level, such as a block or variable chunk level.

Deduplication could occur close to where the data is created, which is often referred to as “Source Based Deduplication." It could occur close to where the data is stored, which is commonly called “Target Based Deduplication”.



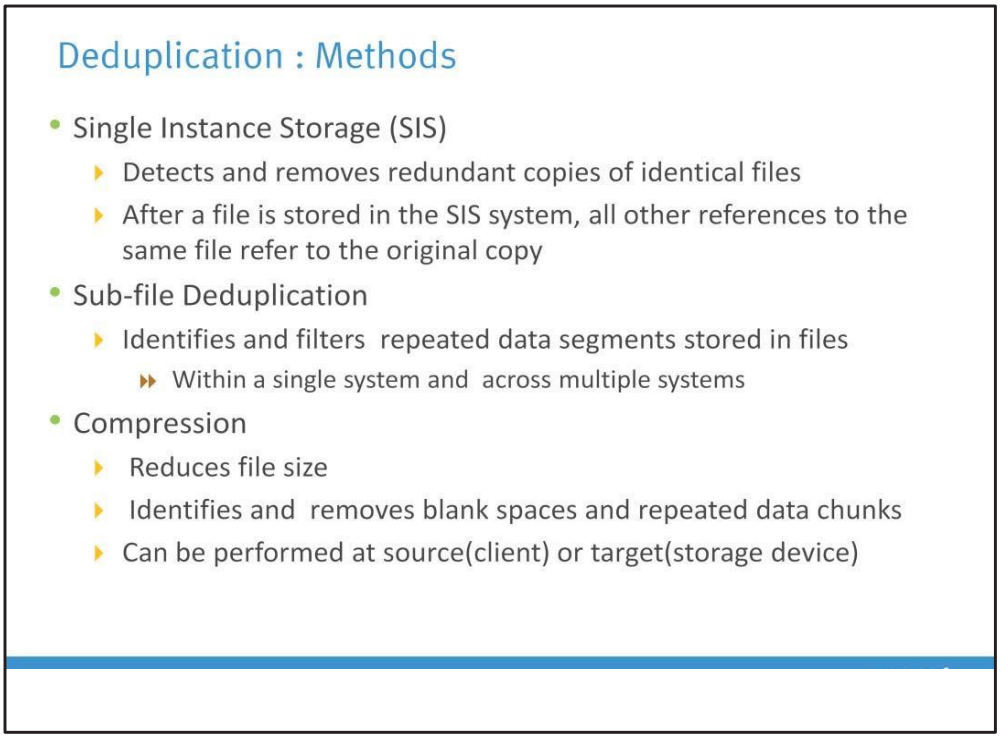
Deduplication directly results in reduced storage capacity requirements to hold backup images. Smaller capacity requirements means lower acquisition costs, reduced power and cooling costs, and reduced carbon footprint. As Deduplication reduces the amount of content in the daily backup, users can extend their retention policies. This can have a significant benefit to users who currently require longer retention, but are limited by the current processes and policies. Deduplication enables many organizations to create full backup images daily. Many of these organizations were forced to do weekly full backups and daily incremental due to backup window (predetermined time duration to complete the data backup) constraints. Deduplication reduces the storage capacity requirements, which therefore permits more aggressive backup policies with improved restore times.



Source based Deduplication eliminates redundant data at the source. This means that data Deduplication is performed at the start of the backup process—before the data is transmitted to the backup environment. Source based Deduplication can radically reduce the amount of backup data sent over networks during backup processes. This is important if there are bottlenecks in the backup

process related to networks, shared resources, or backup windows. Furthermore, there is also a substantial reduction in the capacity requirements to store the backup images. Source based Deduplication increases the overhead on the backup client, which impacts the backup performance.

Target based Deduplication is an alternative to source based Deduplication. Target based Deduplication happens at the backup device. Because Deduplication happens at the target, all the backup data need to be transferred over the network, which may consequently increase network bandwidth and capacity requirements.

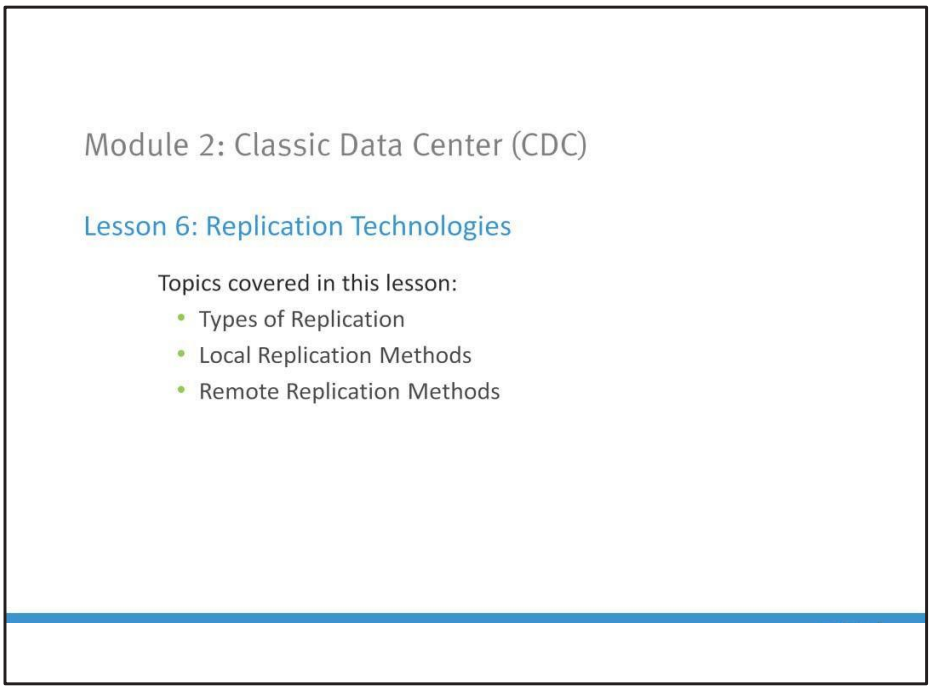


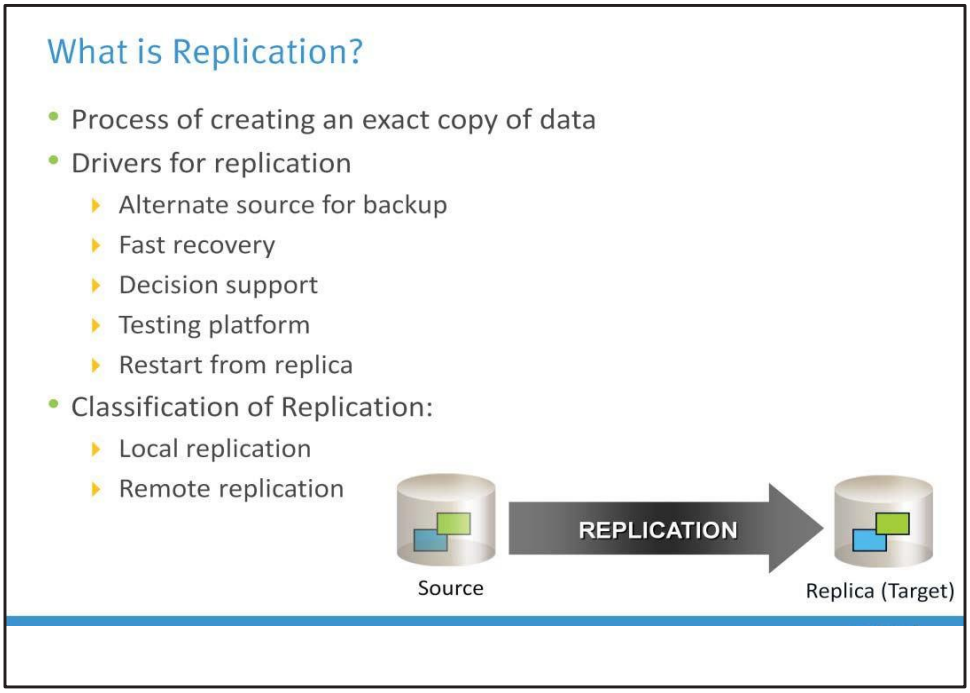
There are three methods for implementing Deduplication: Single Instance storage, Sub-file Deduplication, and compression.

Single Instance Storage (SIS) environments can detect and remove redundant copies of identical files. After a file has been stored in an SIS system, all other references to the same file will refer to the original, single copy. SIS systems compare the content of files to determine whether the incoming file is identical to an existing one in the storage system.

Sub-file Deduplication detects redundant data within and across files, as opposed to finding identical files in SIS implementations. Sub-file Deduplication identifies and filters repeated data segments stored in files within a single system and across multiple systems, over time. This ensures that each unique data segment is backed up only once across the enterprise. As a result, copied or edited files, shared applications, embedded attachments, and even daily changing databases generate only a small amount of the incremental backup data.

Compression reduces the size of individual files by locating patterns of blank spaces and repeated data chunks and then removing them. While compression is widely used in most backup solutions, it is not the most effective method of Deduplication because it does not compare data across files. Data compression performed at the source can utilize a significant amount of processing power of the compute system, but results in smaller files being sent across the network. In target-based compression, many backup devices, such as tape drives, natively perform compression.





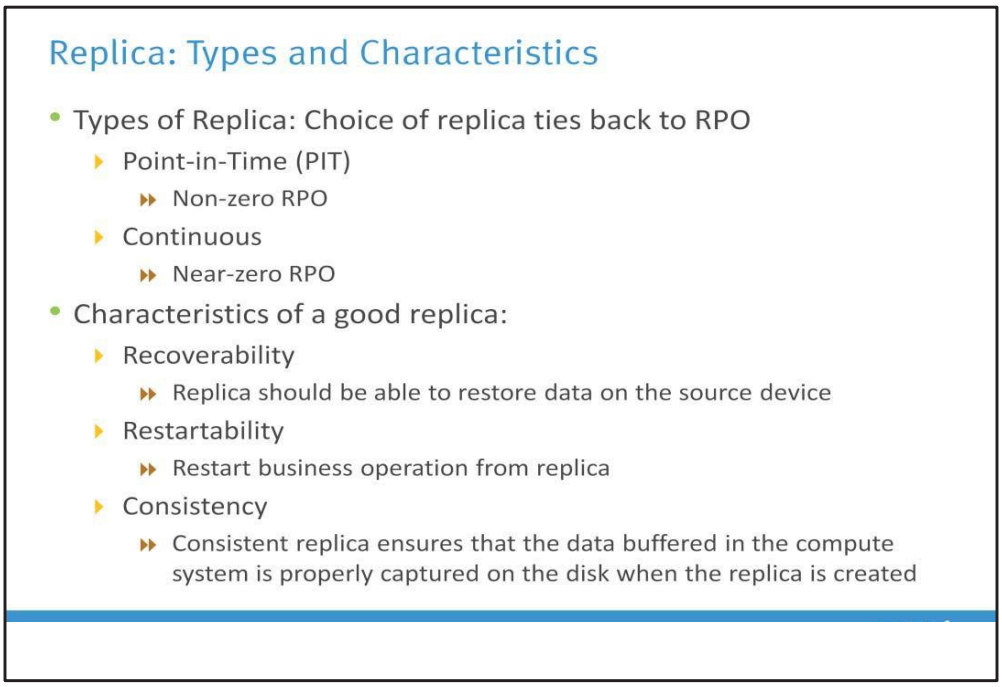
Replication is the process of creating an identical/exact copy of data. The exact copy of data which is created is called a replica. Creating one or more replicas of the production data is one of the ways to provide BC. These replicas can be used for recovery and restart operations in the event of data loss. The primary purpose of replication is to enable users to have the designated data at the right place, in a state appropriate to the recovery needs. This enables restarting business operations using the replicas. Replicas can be used to address a number of Business Continuity functions, such as:

- Providing an alternate source for backup to alleviate the impact on production
- Providing a source for fast recovery to facilitate faster RPO and RTO
- Enabling decision support activities, such as reporting. For example, a company may have

a requirement to generate periodic reports. Running the reports from the replicas greatly reduces the burden placed on the production volumes. Typically, reports are required periodically (Ex: once a day or once a week.)

- Developing and testing proposed changes to an application or an operating environment. For example, the application can be run on an alternate server using the replica volumes. Any proposed design changes can also be tested.
- Restarting an application from the replica in the event of a failure in the source volume

Replication is classified as: Local Replication and Remote Replication. These are discussed in detail later in the module.

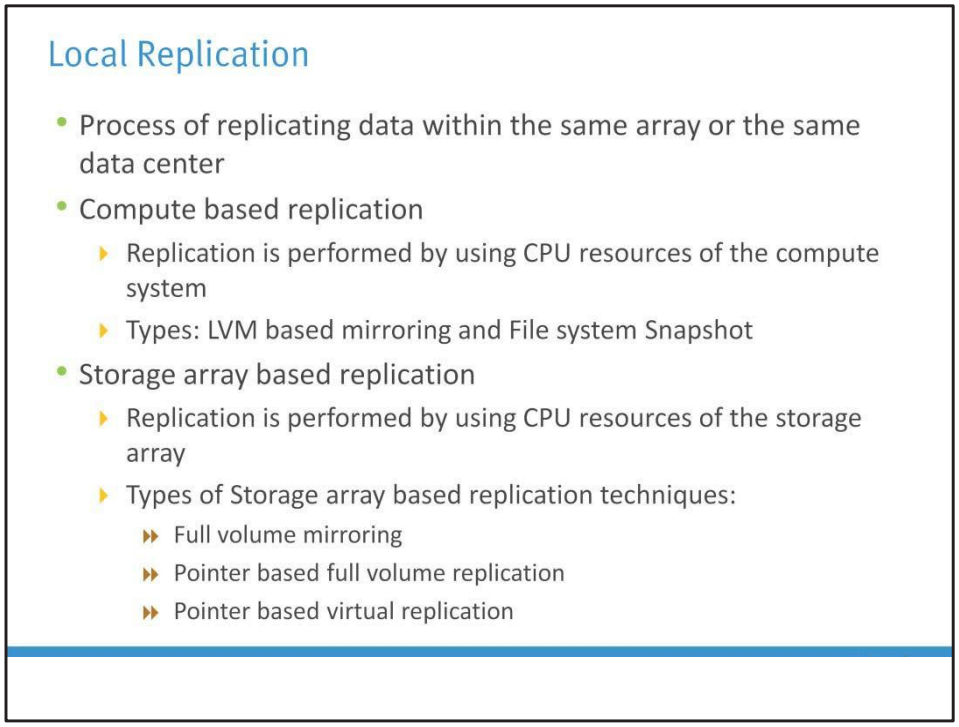


Key factors to consider with replicas:

- Replicas can be either Point-in-Time (PIT) or Continuous:
    - Point-in-Time (PIT) - the data on the replica is an identical image of the production at some specific timestamp.
    - For example, a replica of a file system is created at 4:00 PM on Monday. This replica would then be referred to as the Monday 4:00 PM Point-in-Time copy.
    - Note: The RPO will be a finite value with any PIT. The RPO will map to the time when the PIT was created to the time when any kind of failure on the production occurred. If there is a failure on the production at 8:00 PM and there is a 4:00 PM PIT available, the RPO would be 4 hours (8 – 4 = 4). To minimize RPO, periodic PITs should be taken.
    - Continuous replica - the data on the replica is synchronized with the production data at all times .
    - The objective of any continuous replication process is to reduce the RPO to zero.
- What makes a replica good:
    - Recoverability: The replication technology must allow for the restoration of data from the replicas to the production.
    - Restartability: It should be possible to restart business operation from the replica

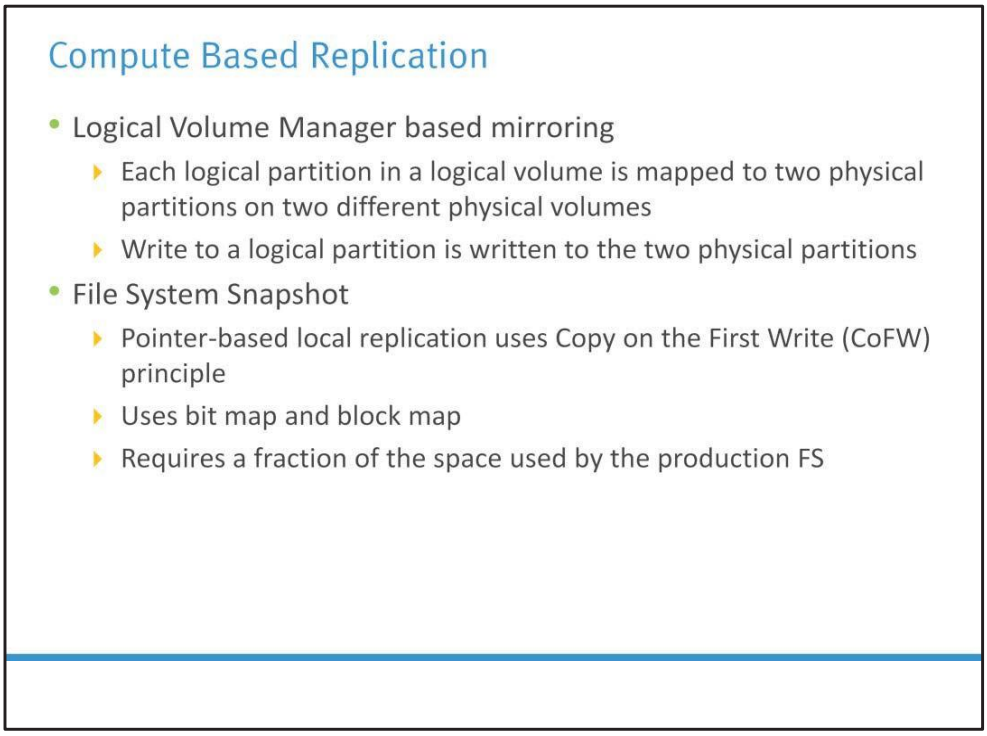
after failure of source.

- Consistency: A consistent replica ensures that the data buffered in the compute system is properly captured on the disk when the replica is created. Ensuring consistency is the primary requirement for all the replication technologies. In the case of file systems, consistency can be achieved either by un-mounting FS or by keeping FS online and flushing the compute system buffers before creating replica. Similarly, in the case of databases, either the database needs to be shutdown for creating consistent replica or the hot backup mode needs to be used for online databases.



Local Replication is the process of replicating data within the same array or the same data center. Local Replication technologies can be classified on the basis of the location where the replication is performed.

- Compute based – Replication is performed by using the CPU resources of the compute system via a software that is running on the compute system. Compute based local replication can be further categorized as LVM based mirroring and file system snapshot.
- Storage array based – Replication is performed on the storage array using the CPU resources of the array via the array's operating environment. In this case, the compute system is not burdened by the replication operations. The replica can be accessed by an alternate server for any business operations. In this replication, the required number of replica devices should be selected on the same array and then the data should be replicated between the source-replica pairs. Storage array based local replication can be further categorized as full volume mirroring, pointer based full volume replication, and pointer based virtual replication.



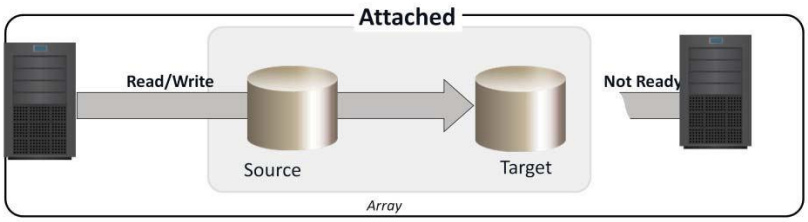
In an LVM-based local replication, the logical volume manager is responsible for creating and controlling the compute-level logical volume. An LVM has three components: physical volumes (physical disk), volume groups, and logical volumes. A volume group is created by grouping together one or more physical volumes. Logical volumes are created within a given volume group. In an LVM-based local replication, each logical partition in a logical volume is mapped to two physical partitions on two different physical volumes. An application write to a logical partition is written to the two physical partitions by the LVM device driver. This is also known as LVM mirroring. Mirrors can be split and the data contained therein can be independently accessed.

File system (FS) snapshot is a pointer-based local replication that requires a fraction of the space used by the production FS. This snapshot can be implemented by either FS itself or by LVM. It uses the Copy on First Write (CoFW) principle. In a CoFW mechanism, if a write I/O is issued to the production FS for the first time after the creation of snapshot, the I/O is held, and the original data of production FS corresponding to that location is moved to the snap FS (replica). Then, the new data is allowed to write on to the production FS. The bitmap and blockmap are updated accordingly. Any subsequent write to the same location will not initiate the CoFW activity.

When the snapshot is created, a bitmap and a block map are created in the metadata of the Snap FS. The bitmap is used to keep track of blocks that are changed on the production FS after creation of the snap. The block map is used to indicate the exact address from which the data is to be read when the data is accessed from the Snap FS. Immediately after creation of the snapshot, all reads from the snapshot will actually be served by reading the production FS. To read from the Snap FS, the bitmap is consulted. If the bit is 0, then the read is directed to the production FS. If the bit is 1, then the block address is obtained from the block map, and data is read from that address. Reads from the production FS are performed as usual.

Full Volume Mirroring

- Target is a full physical copy of the source device
- Target is attached to the source and data from the source is copied to the target
- Target is unavailable while it is attached
- Target device is as large as the source device



In full-volume mirroring, the target is attached to the source and established as a mirror of the source. The existing data on the source is copied to the target. New updates to the source are also updated on the target. After all the data is copied and both the source and the target contain identical data, the target can be considered a mirror of the source. While the target is attached to the source and the synchronization is taking place, the target remains unavailable to any other server. However, the production server can access the source.

Note: A compute system accessing data from one or more LUNs on the storage array is called a production compute system. These LUNs are known as source LUNs (devices/volumes), production LUNs, or simply the Source. A LUN (or LUNs) on which the data is replicated is called the Target LUN or simply the Target.

Pointer Based Full Volume Replication

- Provides a full copy of the source data on the target
- Target device is made accessible for business operations as soon as the replication session has started
- Point-in-Time (PIT) is determined by the time of session activation
- Two modes
  - Copy on First Access (deferred)
  - Full Copy mode
- Target device is at least as large as the source device

Similar to full-volume mirroring, pointer based full volume replication can provide full copies of the source data on the targets. Unlike full-volume mirroring, the target is made immediately available at the activation of the replication session. Hence, one need not wait for data synchronization and detachment of the target in order to access it. The time of activation defines the PIT copy of source. Pointer-based, full-volume replication can be activated in either the Copy on First Access (CoFA) mode or on the Full Copy mode.

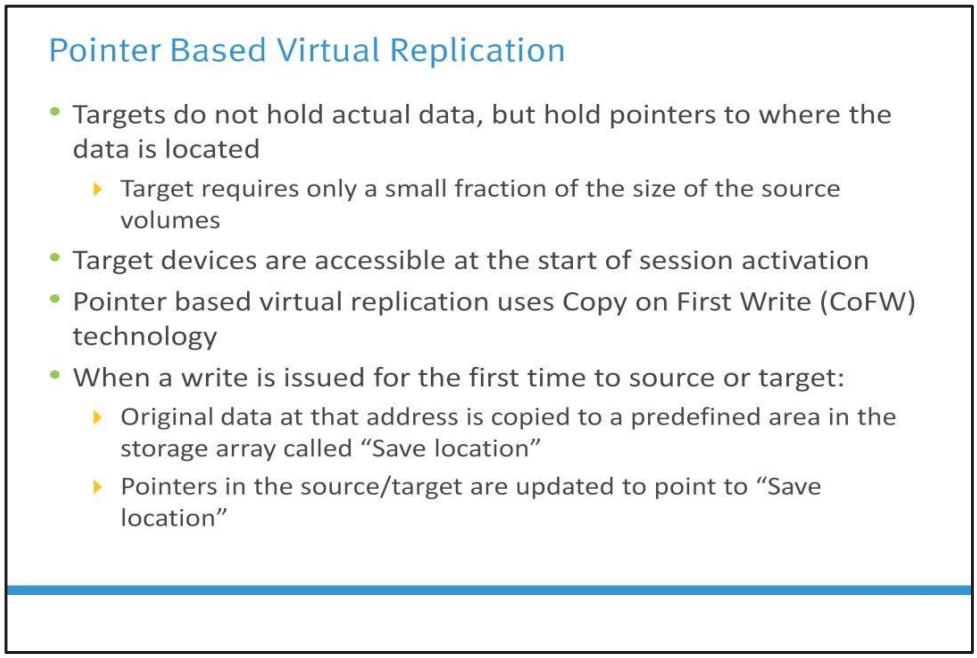
In either case, at the time of activation, a protection bitmap is created for all data on the source devices. Pointers are initialized to map the (currently) empty data blocks on the target to the corresponding original data blocks on the source. The granularity can range from 512 byte blocks to 64 KB blocks or higher. Data is then copied from the source to the target, based on the mode of activation. In a Full Copy mode, the target is made available immediately and all the data from the source is copied over to the target in the background.

- During this process, if a data block that has not yet been copied to the target is accessed, the replication process jumps ahead and moves the required data block first.
- When a full copy mode session is terminated (after full synchronization), the data on the target is still usable because it is a full copy of the original data. This makes the target a viable copy for recovery, restore, or other business continuity operations.

In the Copy on First Access mode (or the deferred mode), data is copied from the source to the target only when:

- A write is issued for the first time after the PIT to a specific address on the source.
- A read or write is issued for the first time after the PIT to a specific address on the target.

When a write is issued to the source for the first time after session activation, the original data at that address is copied to the target. After this operation, the new data is updated on the source. This ensures that the original data at the point-in-time of activation is preserved on the target.



In a pointer based virtual replication, at the time of session activation, the target contains pointers to the location of the data on the source. The target does not contain data, at any time. Hence, the target is known as a virtual replica. Similar to pointer based full volume replication, a protection bitmap is created for all the data on the source device, and the target is immediately accessible.

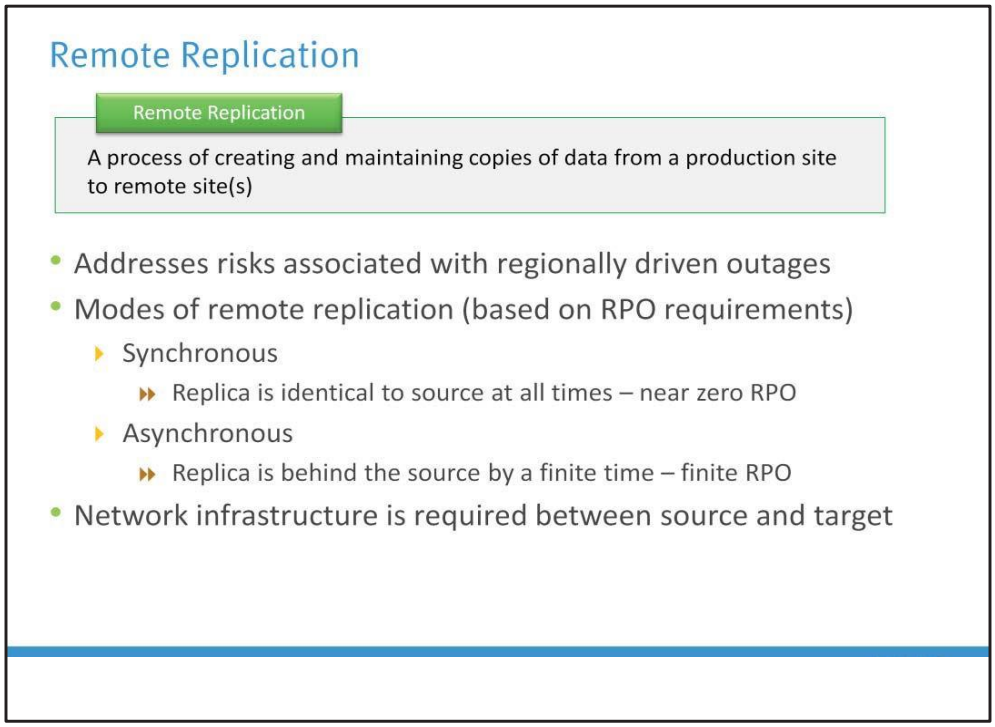
Granularity ranges from 512 byte blocks to 64 KB blocks or greater. The primary advantage of pointer based copies is the reduction in storage requirement for the replicas.

Pointer based virtual replication uses CoFW technology. When a write is issued to the source for the first time after session activation, the original data at that address is copied to a predefined area in the array. This area is generally termed the Save location. The pointer in the target is updated to point to the data address in the Save location. After this, the new write is updated on the source.

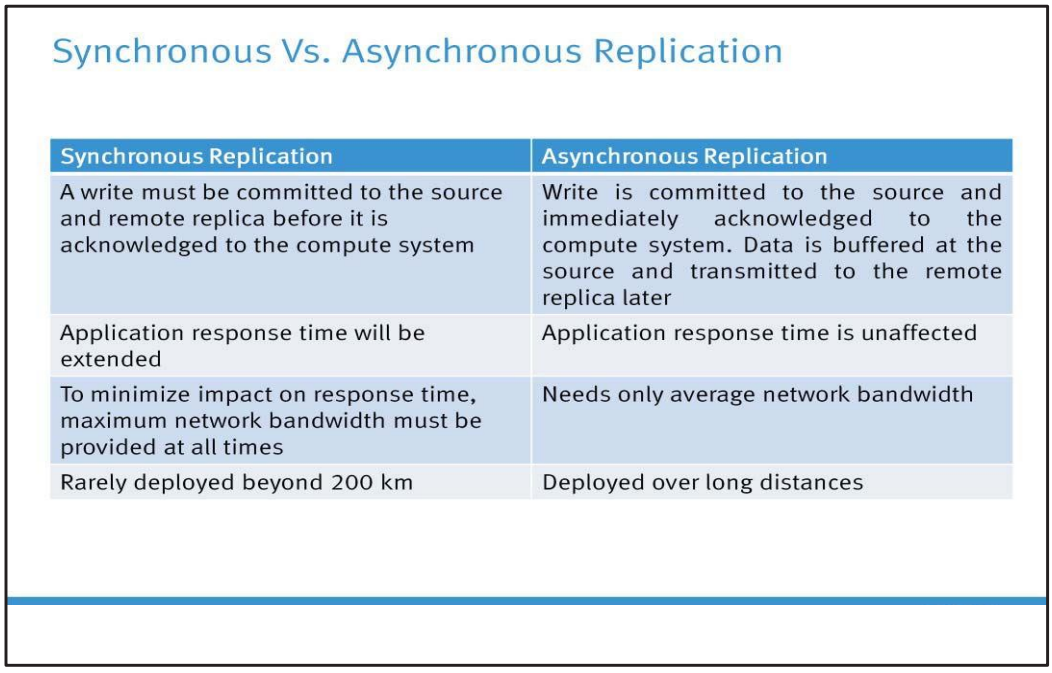
When a write is issued to the target for the first time after session activation, the original data is copied from the source to the Save location and similarly the pointer is updated to data in the Save location. Subsequent writes to the same data block on the source or the target do not trigger a copy operation.

When reads are issued to the target, unchanged data blocks since the session activation are read from the source. The original data blocks that have changed are read from the Save location.

Data on the target is a combined view of unchanged data on the source and data on the Save location. Unavailability of the source device invalidates the data on the target. As the target contains only pointers to data, the physical capacity required for the target is a fraction of the source device. The capacity required for the Save location depends on the amount of expected data change.

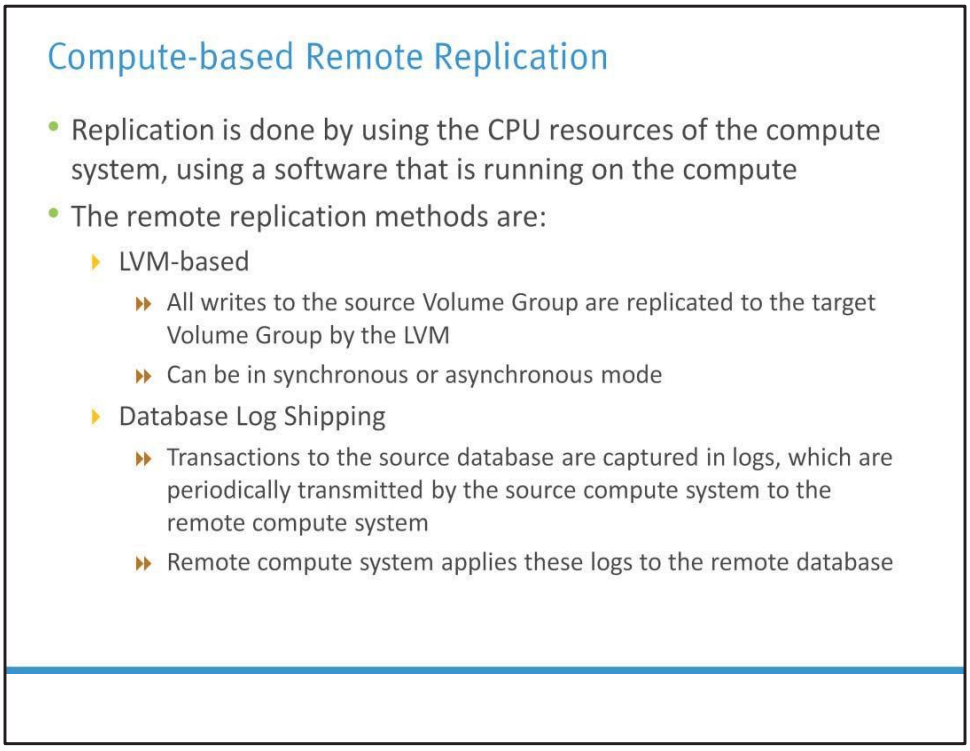


Remote replication is the process of creating replicas of production (local) data to remote sites (locations). Remote replicas help organizations mitigate the risks associated with regionally driven outages resulting from natural or human-made disasters. Similar to local replicas, they can also be used for other business operations. The infrastructure on which the data is stored at the primary site is called Source. The infrastructure on which the replica is stored at the remote site is referred to as Target. Data has to be transferred from the source site to a target site over some network. Two basic modes of remote replications are: Synchronous and Asynchronous replication.



In synchronous remote replication, writes must be committed to the source and the target, prior to acknowledging "write complete" to the compute system. Additional writes on the source cannot occur until each preceding write has been completed and acknowledged. This ensures that data is identical on the source and the replica at all times. Further, writes are transmitted to the remote site exactly in the order in which they are received at the source. Hence, write ordering is maintained. In the event of a failure of the source site, synchronous remote replication provides zero or near-zero RPO, as well as the lowest RTO. However, the application response time is increased with any synchronous remote replication. The degree of the impact on the response time depends on the distance between sites, available bandwidth, and the network connectivity infrastructure. The distances over which synchronous replication can be deployed depend on the application's ability to tolerate extension in response time. Typically, it is deployed for distances less than 200 KM (125 miles) between the two sites. To minimize the response time elongation, ensure that the Max bandwidth is provided by the network at all times.

In asynchronous remote replication, a write is committed to the source and immediately acknowledged to the compute system. Data is buffered at the source and transmitted to the remote site later. Data at the remote site will be behind the source by at least the size of the buffer. Hence, asynchronous remote replication provides a finite (nonzero) RPO. RPO depends on the size of the buffer, available network bandwidth, and the write workload to the source. There is no impact on the application response time because the writes are acknowledged immediately by the source. This enables deployment of asynchronous replication over extended distances. Asynchronous remote replication can be deployed over distances ranging from several hundred to several thousand kilometers between two sites. The available network bandwidth should be at least equal to the average write workload. Data is buffered during times when the bandwidth is not enough; thus, sufficient buffers should be designed into the solution.



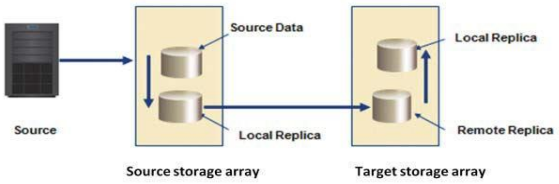
Compute-based remote replication implies that all the replication is done by using the CPU resources of the compute system using a software that is running on the compute system. The following are the compute-based remote replication methods:

LVM-based replication is performed and managed at the volume group level. Writes to the source volumes are transmitted to the remote compute system by the LVM. The LVM on the remote compute system receives the writes and commits them to the remote volume group. Prior to the start of replication, identical volume groups, logical volumes, and file systems are created at the source and target sites. LVM-based remote replication supports both synchronous and asynchronous modes of data transfer. In the asynchronous mode, writes are queued in a log file at the source and sent to the remote compute system in the order in which they were received.

Database replication via log shipping is a compute-based replication technology supported by most databases. Transactions to the source database are captured in logs, which are periodically transmitted by the source compute system to the remote compute system. The remote compute system receives the logs and applies them to the remote database. Prior to starting the production work and replicating the log files, all the relevant components of the source database are replicated to the remote site. This is done while the source database is shut down. After this step, production work is started on the source database. The remote database is started in a standby mode. Typically, in the standby mode, the database is not available for transactions. All DBMSs switch the log files at preconfigured time intervals, or when a log file is full.

Storage Array Based Remote Replication

- Performed by array operating environment
  - Synchronous Replication
  - Asynchronous Replication
  - Disk buffered Replication (shown in the figure)
    - Combination of local and remote replications
    - RPO usually in the order of hours
    - Low bandwidth requirements
    - Extended distance solution



In a storage array based remote replication, replication is performed by the array operating environment. Storage array based remote replication technologies support three modes of operations: Synchronous, asynchronous, and disk buffered.

Disk-buffered replication is a combination of local and remote replication technologies. A consistent PIT local replica of the source device is first created. Then, the data on the local replica in the source array is transmitted to its remote replica in the target array. Optionally, a local PIT replica of the remote device on the target array can be created. The frequency of this cycle of operations depends on the available link bandwidth and the data change rate on the source device.

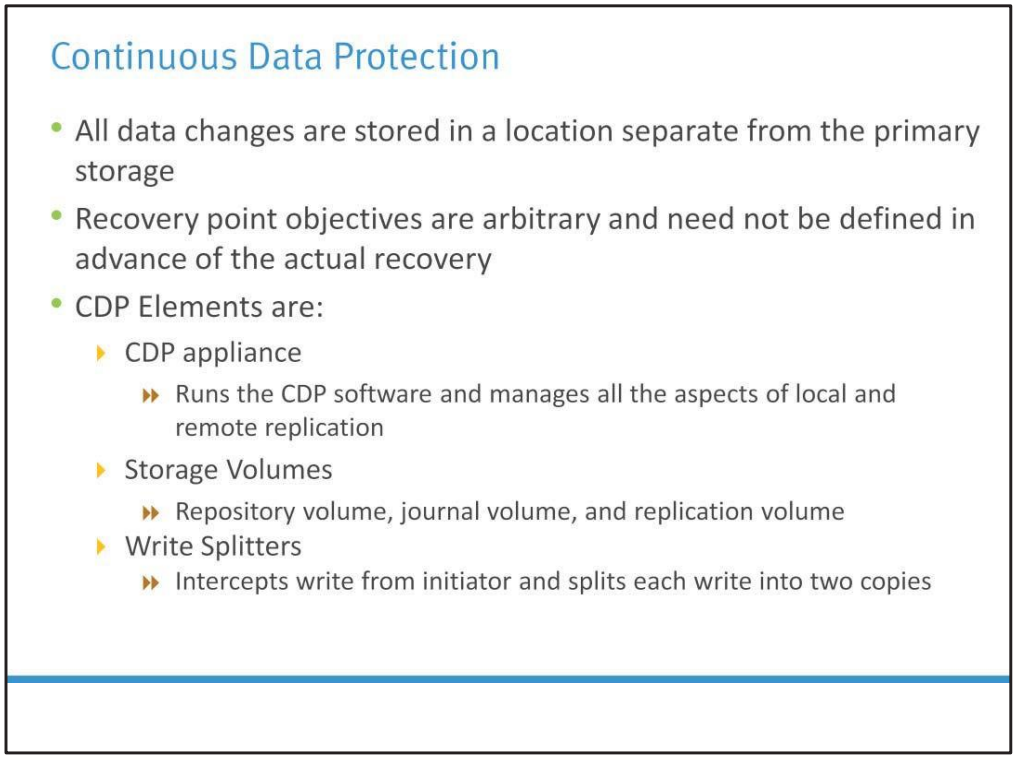
Advanced Replication Technologies

- Three site replication
  - Eliminates disadvantages of two site replication
  - Replicates data to two remote sites
- SAN based replication
  - Allows replication between heterogeneous vendor storage arrays over SAN/WAN
- Continuous Data Protection (CDP)
  - Changes to data are continuously captured or tracked

Three-site replication is used for the mitigation of risks identified in two-site replication. In a three-site replication, data from the source site is replicated to two remote sites. Replication may be synchronous to one of the two sites, and provides a zero-RPO solution. It may be asynchronous or disk buffered to the other remote site, and provides a finite RPO.

SAN-based remote replication allows the replication of data between heterogeneous vendor storage arrays. Data is moved from one array to the other over SAN/WAN. The technology is application and server operating system independent, because the replication operations are performed by one of the storage arrays. There is no impact on the production servers or the LAN because replication is done by the array and the data is moved over the SAN.

Traditional data protection technologies do not meet the needs of all applications in a CDC. Mission-critical applications require instant and unlimited data recovery point options.
Continuous data protection captures all writes and maintains consistent point in time images.



With CDP, recovery from data corruption poses no problem, because it allows going back to a point-in-time image prior to the data corruption incident. CDP provides faster recovery and unlimited recovery point at both the local and remote sites.

CDP systems may be block, file, or application-based and can provide fine granularities of restorable objects. All CDP solutions incorporate three key attributes, such as, data changes are continuously captured, all data changes are stored in a separate location from the primary storage, and RPO are arbitrary and not required to be defined in advance.

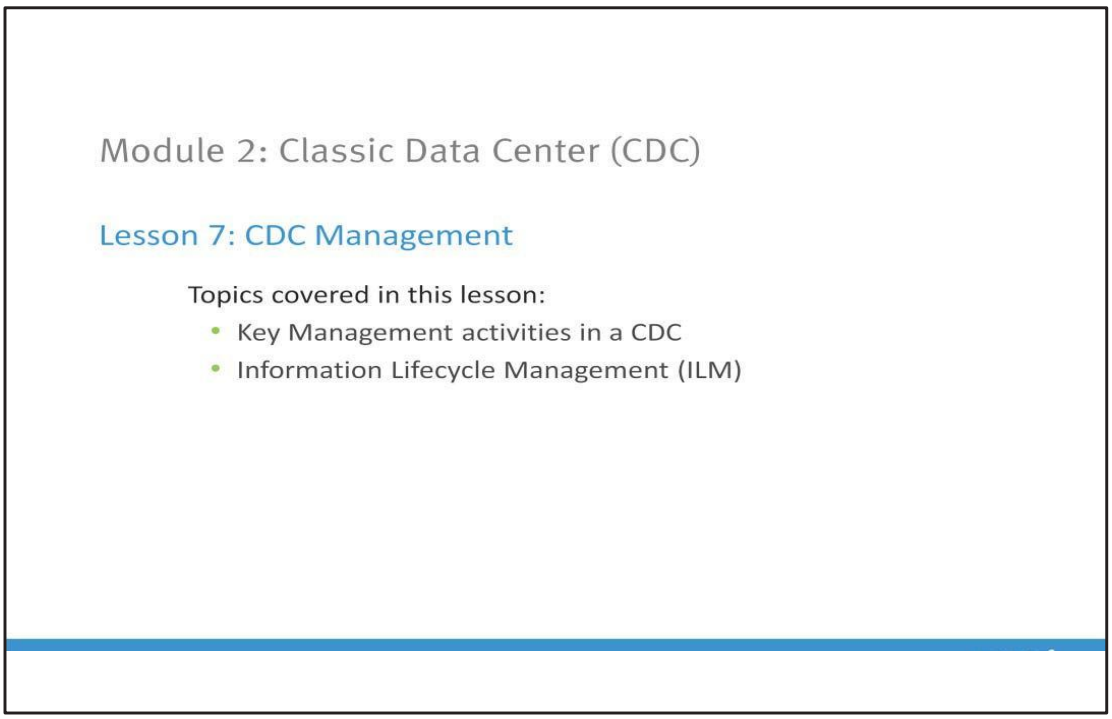
CDP elements include the CDP Appliance, Storage Volumes, and Splitters. CDP appliance is the intelligent custom built-platform that runs the CDP software and manages all the aspects of local and remote data replication. During replication, the CDP appliance at the source site makes intelligent decisions regarding when and what data to transfer to the target site.

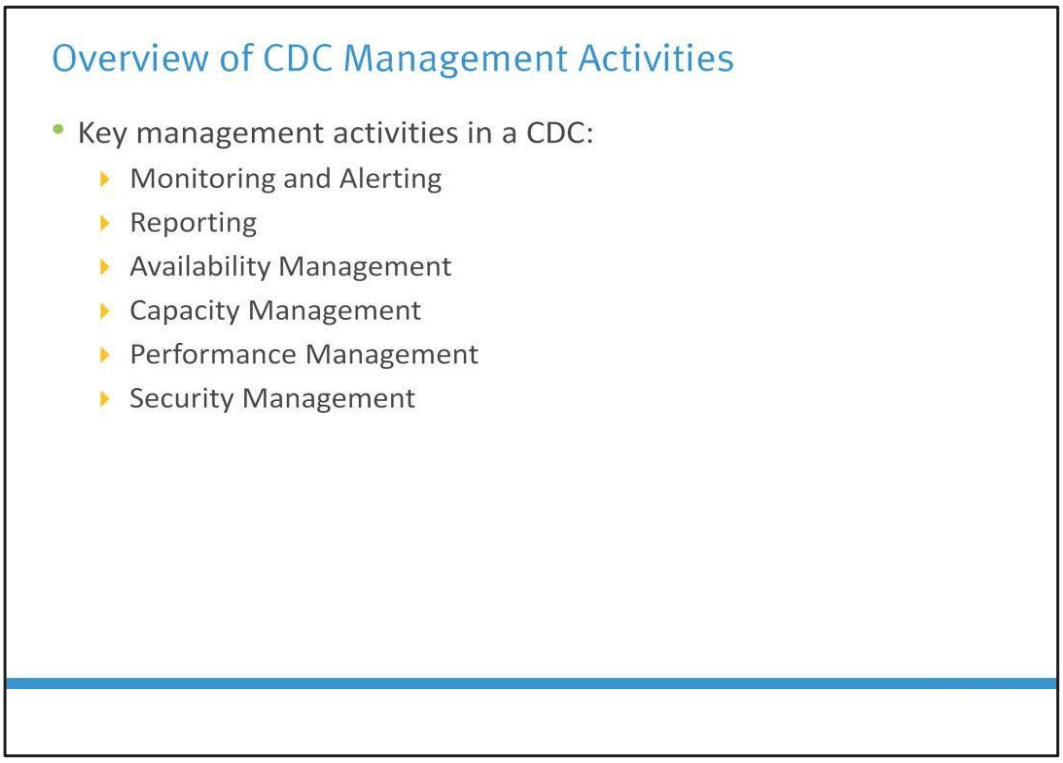
The three types of storage volumes used in CDP are repository volume, journal volume, and replication volume. The Repository volume must be a dedicated volume on the SAN-attached storage at each site. It stores configuration information about the CDP appliance.

Journal Volumes store all data changes on the primary storage. The journal contains the metadata and data that will allow rollbacks to various recovery points. The amount of space that is configured for the journal will determine how backward the recovery points can go.

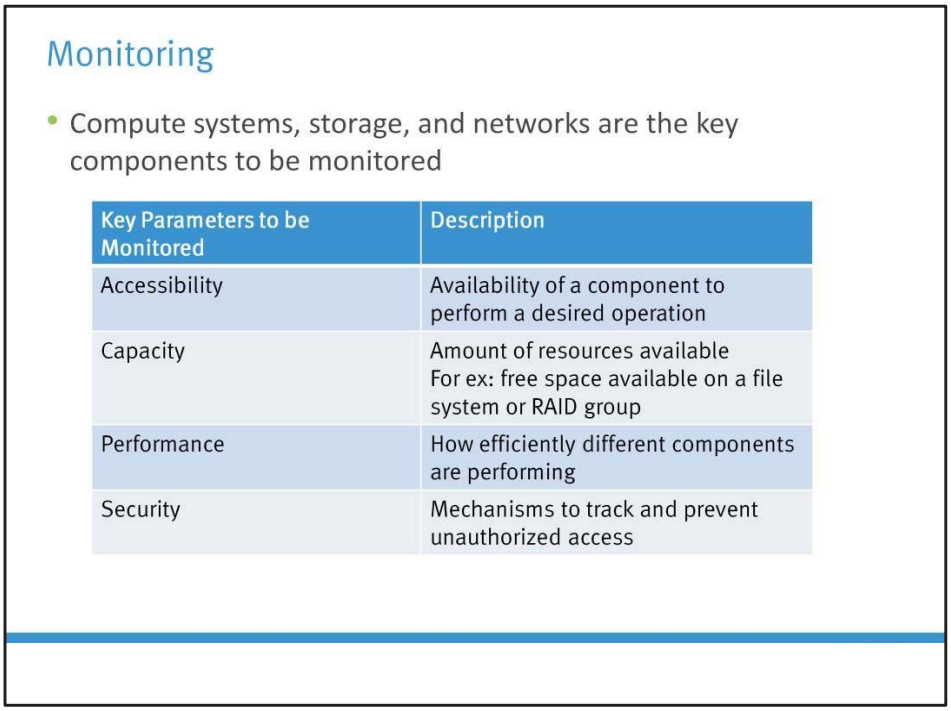
Replication Volumes refer to the data volumes to be replicated. Write splitter intercepts writes from the initiator and splits each write into two copies; one copy is sent to CDP appliance for replication and the other, to the designated production volume. Write splitter may exist in the compute

system, fabric, or storage array.





Managing a CDC involves many tasks. The key management activities in a CDC are monitoring and alerting, reporting, availability management, capacity management, performance management, and security management. These are explained later in the module.



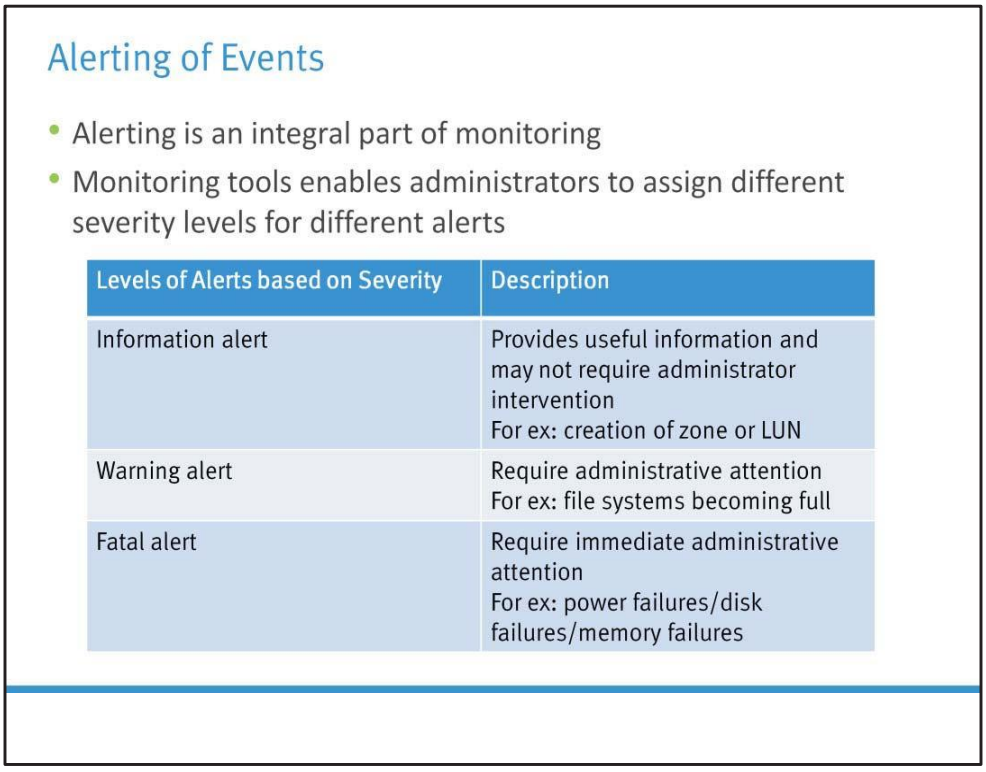
Monitoring helps to analyze the status and utilization of various storage infrastructure components. Compute systems, networks, and storage are the key components that should be monitored for accessibility, capacity, performance, and security. Accessibility refers to the availability of a component to perform a desired operation. Monitoring hardware components or software components (ex: a database instance) for accessibility involves checking their availability status by listening to pre-determined alerts from devices. For example, a port may go down resulting in a chain of availability alerts.

Capacity refers to the amount of storage infrastructure resources available. Examples of capacity monitoring include examining the free space available on a file system or a RAID group, the mailbox quota allocated to users, or the numbers of ports available on a switch.

Inadequate capacity could lead to degraded performance or even application/service availability. Capacity monitoring ensures uninterrupted data availability and scalability by averting outages before they occur. For example, if a report indicates that 90 percent of the ports are utilized in a particular SAN fabric, a new switch should be added if more arrays and servers need to be installed on the same fabric. Capacity monitoring is preventive and predictive, and usually leveraged with advanced analytical tools for trend analysis. These trends help to understand emerging challenges and can provide an estimation of time needed to meet them.

Performance monitoring evaluates how efficiently different storage infrastructure components are performing and helps to identify bottlenecks. Performance monitoring usually measures and analyzes behavior in terms of response time or in terms of the ability to perform at a certain predefined level. It also deals with utilization of resources, which affects the way resources behave and respond. Performance measurement is a complex task that involves assessing various components on several interrelated parameters. The number of I/Os to disks, application response time, network utilization, and server CPU utilization are examples of performance monitoring.

Monitoring CDC resources for security helps to track and prevent unauthorized access and login failures, whether accidental or malicious. Security monitoring also helps to tracks unauthorized configuration changes of storage infrastructure elements. For example, security monitoring tracks and reports the initial zoning configuration performed and all subsequent changes. Physical security of a storage infrastructure is also continuously monitored using badge readers, biometric scans, or video cameras.

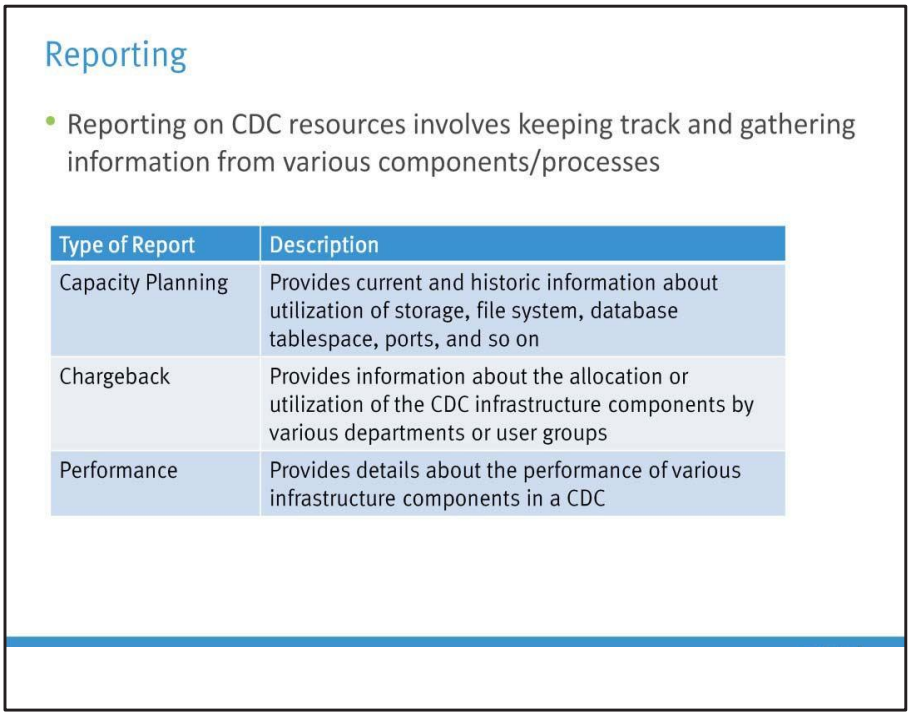


Alerting of events is an integral part of monitoring. There are certain conditions observed by monitoring, such as failure of power, disks, memory, or switches, which may impact the availability of services that require immediate administrative attention. Other conditions, such as a file system reaching a capacity threshold or a soft media error, are considered warning signs, and may also require administrative attention.

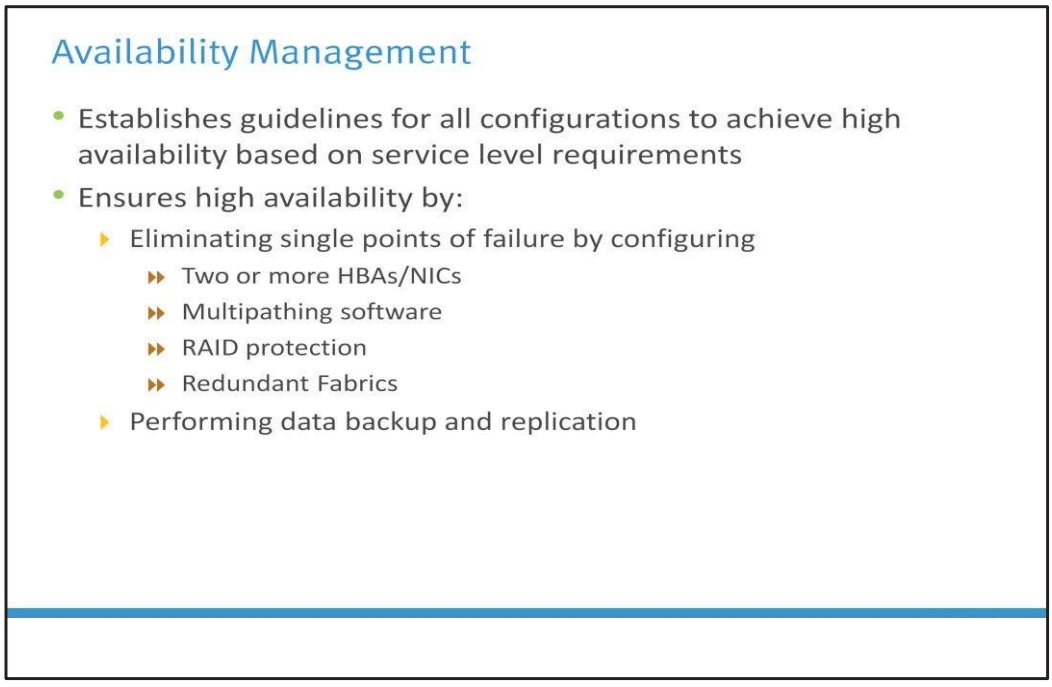
Monitoring tools enable administrators to assign different severity levels for different conditions in the storage infrastructure. Whenever a condition with a particular severity level occurs, an alert is sent to the administrator to initiate a corrective action. Alert classifications can range from information alerts to fatal alerts.

- Information alerts provide useful information that does not require any intervention by the administrator. Creation of zone or LUN is an example of an information alert.
- Warning alerts require administrative attention so that the alerted condition is contained and does not affect accessibility. For example, when an alert indicates a soft media error on a disk that is approaching a predefined threshold value, the administrator can decide whether the disk needs to be replaced.
- Fatal alerts require immediate attention because the condition could affect overall performance or availability. For example, if a disk fails, the administrator must ensure that it is replaced quickly.

Alerts can be assigned a severity level based on the impact of the alerted condition. Continuous monitoring, in conjunction with automated alerting, enables administrators to respond to failures quickly and proactively. Alerting provides information to prioritize the administrator's response to events.



It is difficult for businesses to keep track of the resources they have in their CDCs, for example, the number of storage arrays, the array vendors, mode of usage of the storage arrays, and the applications. Reporting on CDC resources involves keeping track and gathering information from various components/processes. This information is compiled to generate reports for capacity planning, chargeback, performance, and so on. Capacity planning reports also contain current and historic information about storage utilization, file system, database tablespace, and ports. Chargeback reports contain information about the allocation or utilization of CDC infrastructure components by various departments or user groups. Performance reports provide details about the performance of various infrastructure components in a CDC.

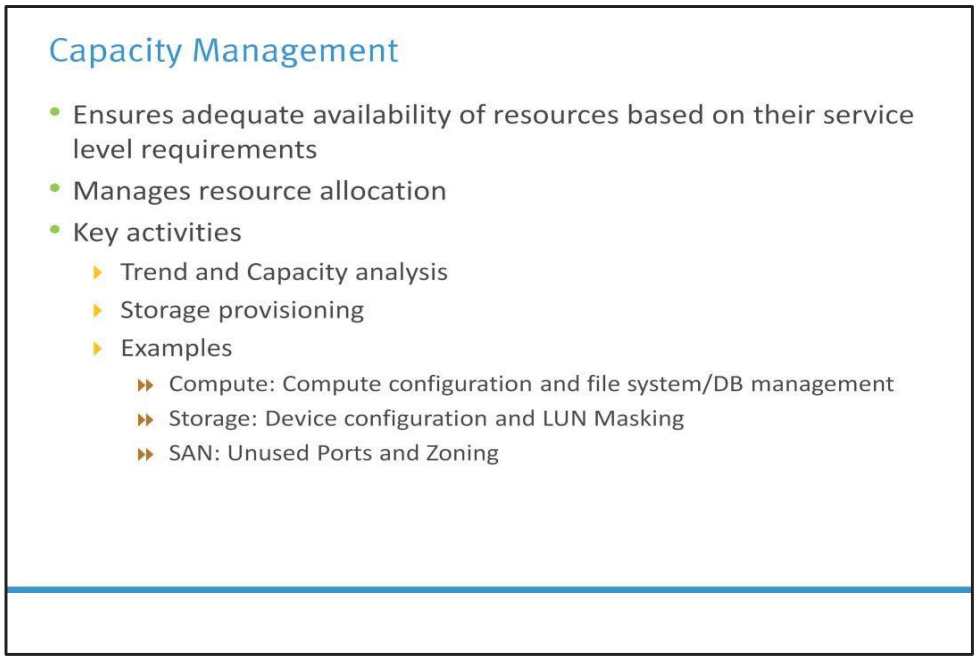


The critical task in availability management is establishing a proper guideline for all configurations to ensure availability based on service levels. For example, when a compute is

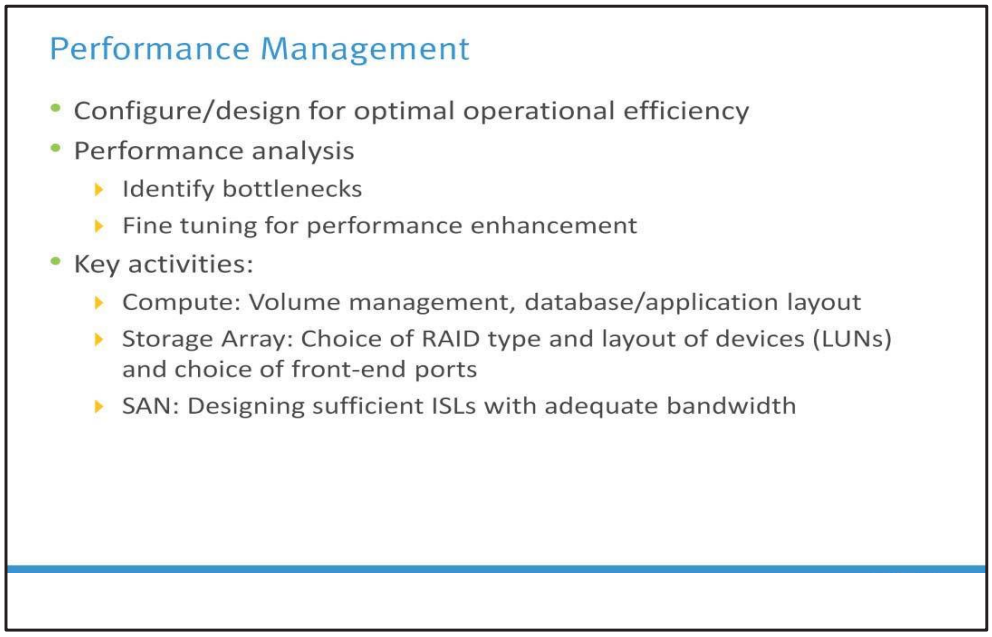
deployed to support a critical business function, the highest availability standard is usually required. This is generally accomplished by:

- Deploying two or more HBAs/NICs, multipathing software with path failover capability, and server clustering.
- Connecting the server to the storage array using at least two independent fabrics and switches that have built-in redundancy.
- Availing storage devices with RAID protection to the server using at least two front- end ports.

In addition, these storage arrays should have built-in redundancy for various components, should support backup and replication (both local and remote).

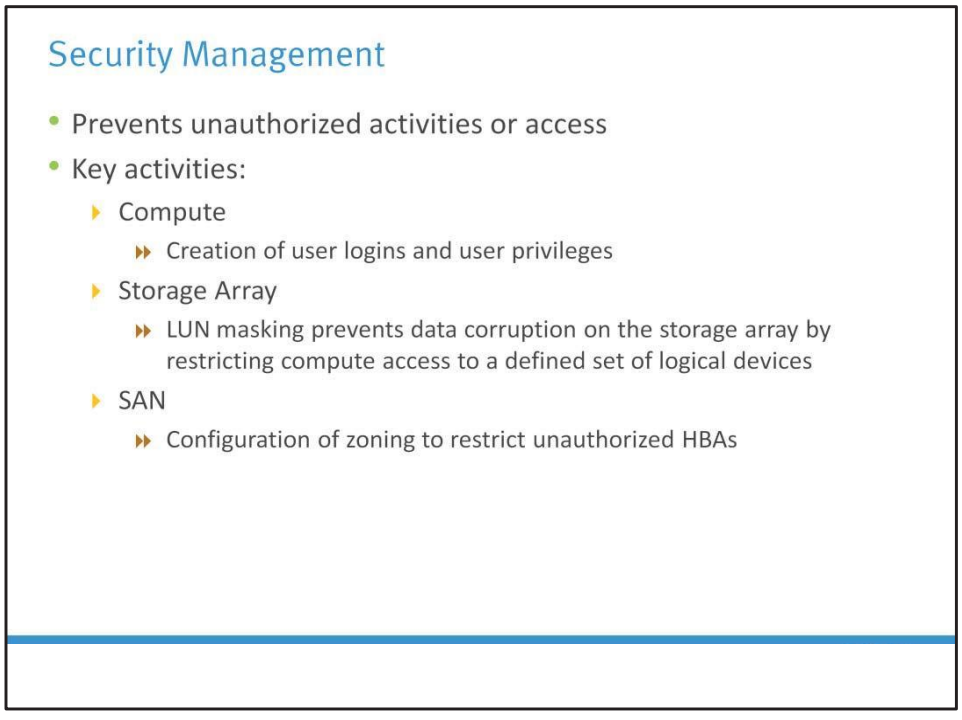


The goal of capacity management is to ensure adequate availability of resources for all services based on their service-level requirements. Capacity management provides capacity analysis and compares the allocated storage to the forecasted storage on a regular basis. It also provides a trend analysis of the actual utilization of allocated storage. Additionally, it also provides information on the rate of consumption, which must be rationalized against storage acquisition and deployment timetables. Storage provisioning is an example of capacity management. It involves activities such as device configuration and LUN masking on the storage array and zoning configuration on the SAN and HBA components. Capacity management also takes into account the future needs of resources, and setting up of monitors and analytics to gather such information.

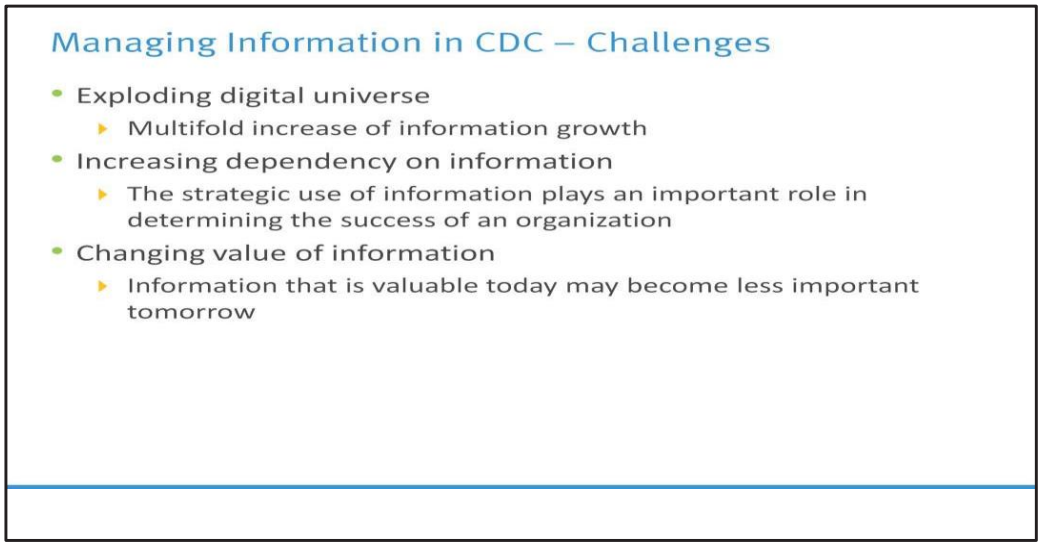


Performance management ensures the optimal operational efficiency of all components. Performance analysis is an important activity that helps to identify the performance of storage infrastructure components. This analysis provides information on whether a component is meeting the expected performance levels. Several performance management activities are initiated for the deployment of an application or server in the existing storage infrastructure. Every component must be validated for adequate performance capabilities, as defined by the service levels. For example, to optimize the expected performance levels, the activities on the server, such as the volume configuration, designing the database, application layout configuration of multiple HBAs, and intelligent multipathing software, must be fine- tuned. The performance management tasks on a SAN include designing sufficient ISLs in a multi-switch fabric with adequate bandwidth to support the required performance levels.

While considering the end-to-end performance, the storage array configuration tasks include selecting the appropriate RAID type and LUN layout, front-end and back-end ports, and LUN accessibility (LUN masking).



Security management prevents unauthorized access and configuration of storage infrastructure components. For example, while deploying an application or a server, security management tasks include managing the user accounts and access policies that authorize users to perform role-based activities. The security management tasks in a SAN environment include configuration of zoning to restrict an HBA's unauthorized access to the specific storage array ports. LUN masking prevents data corruption on the storage array by restricting compute access to a defined set of logical devices.

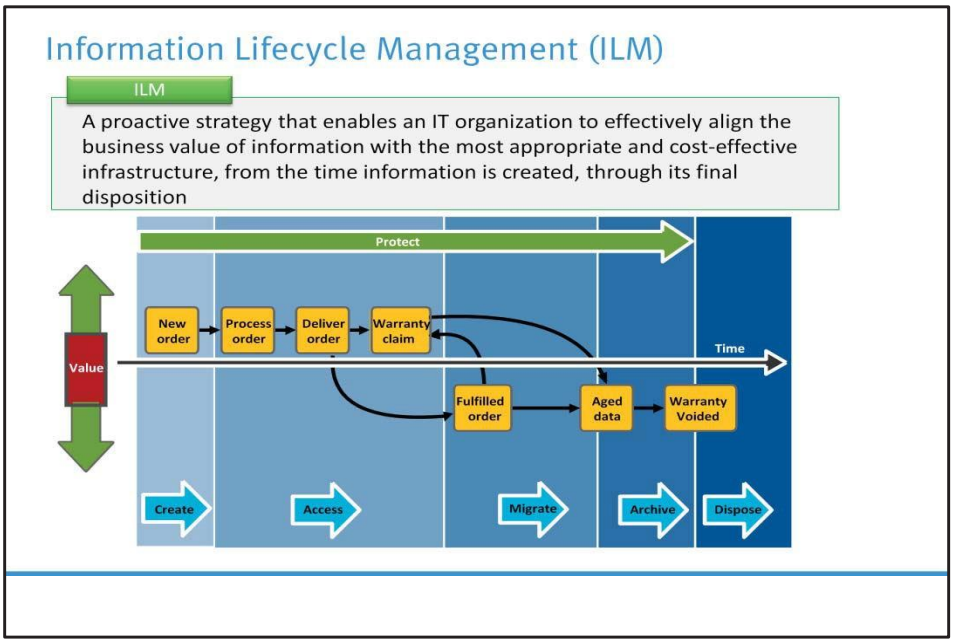


In order to frame an effective information management policy, organisations need to consider the following key challenges of information management:

Exploding digital universe: The rate of information growth is increasing exponentially. Duplication of data to ensure high availability and its repurposing have also contributed to the multifold increase of information growth.

Increasing dependency on information: The strategic use of information plays an important role in determining the success of an organization and provides competitive advantages in the marketplace.

Changing value of information: Information that is valuable today may become less important tomorrow. The value of information often changes over time. Framing a policy to meet these challenges involves understanding the value of information over its lifecycle



An information lifecycle entails "change in the value of information" over time.

When data is first created, it often has the highest value and is used frequently. As data ages, it is accessed less frequently and is of less value to the organization. Understanding the information lifecycle helps to deploy appropriate storage infrastructure, according to the changing value of information. For example, in a sales order application, the value of the information changes from the time the order is placed until the time that the warranty becomes void. The value of the information is highest when a company receives a new sales order and processes it to deliver the product. After order fulfillment, the customer or order data need not be available for real-time access. The company can transfer this data to a less expensive secondary storage with lower accessibility and availability requirements, unless or until a warranty claim or another event triggers its need. After the warranty becomes void, the company can archive or dispose off the data to create space for other high-value information.

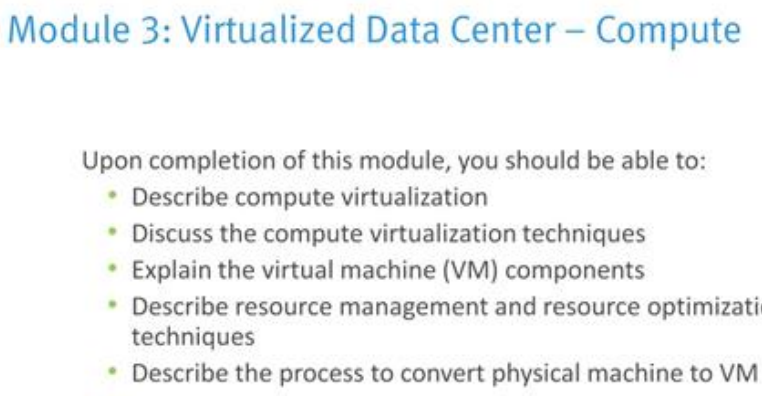
Today's business requires data to be protected and available 24 × 7. Data centers can accomplish this with the optimal and appropriate use of storage infrastructure. An effective information management policy is required to support this infrastructure and leverage its benefits.

Information lifecycle management (ILM) is a proactive strategy that enables an IT organization to effectively manage data throughout its lifecycle, based on predefined business policies. This allows the IT organization to optimize the storage infrastructure for maximum return on investment. An ILM strategy should include the following characteristics:

- Business-centric: It should be integrated with the key processes, applications, and initiatives of the business to meet both the current and future growth in information.
- Centrally managed: All the information assets of an organization should be under the purview of the ILM strategy.
- Policy-based: The implementation of ILM should not be restricted to a few departments. ILM should be implemented as a policy and should encompass all business applications, processes, and resources.
- Heterogeneous: An ILM strategy should take into account all types of storage platforms and operating systems.
- Optimized: As the value of information varies, an ILM strategy should consider different storage requirements and allocate storage resources based on the information's value to the organization
- Tiered Storage: Tiered storage is an approach to define different storage levels so as to reduce the total storage cost. Each tier has different levels of protection, performance, data access frequency, and other considerations. Information is stored and moved among different tiers based on its value over time. For example, mission-critical, most accessed information may be stored on Tier 1 storage, which consists of high performance media and has the highest level of protection. Data accessed moderately and other important data are stored on Tier 2 storage, which may be a less expensive media that offers moderate performance and protection. Rarely accessed or event-specific information may be stored on lower tiers of storage.

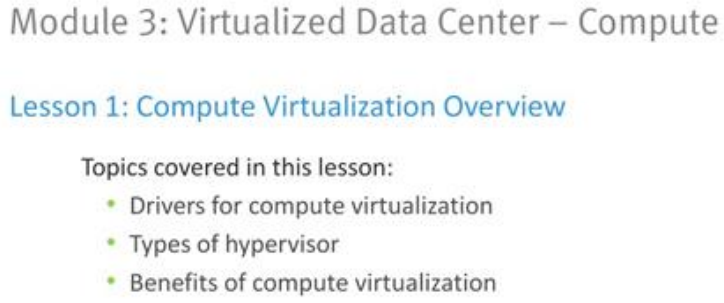
UNIT-3

Virtualized Data Center



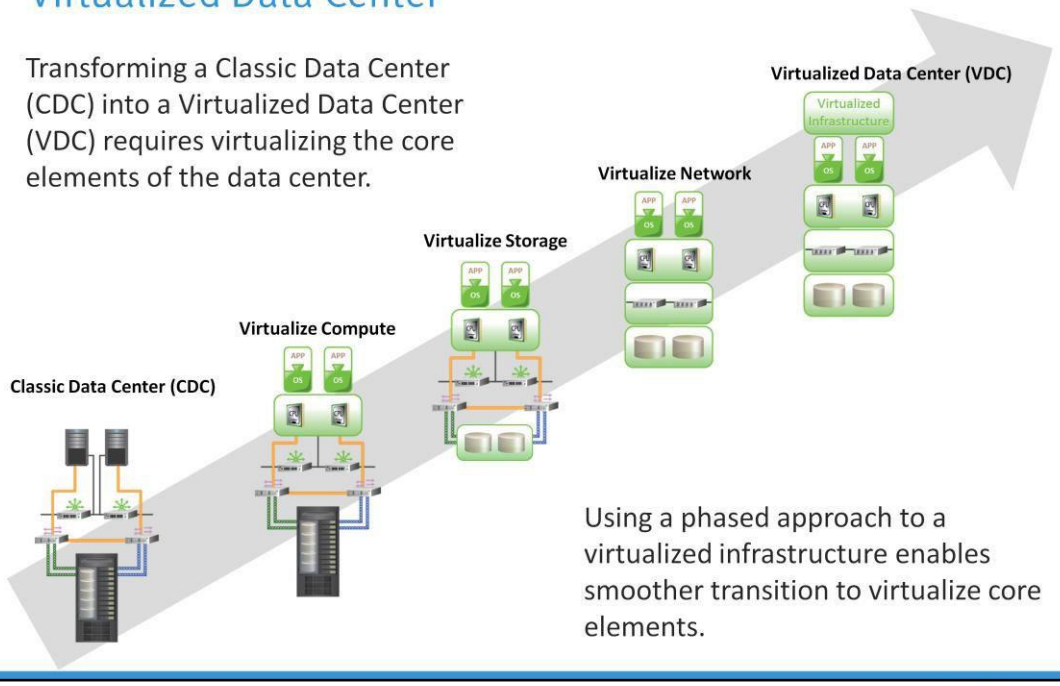
This module focuses on the compute aspect of the Virtualized Data Center (VDC). It explains the fundamental concepts of compute virtualization and describes compute virtualization techniques. This module also details virtual machine (VM) components and management of compute resources. Finally, it describes the process to convert physical machine to VM.

Virtualization is the first step towards building a cloud infrastructure. Transforming a Classic Data Center (CDC) into a Virtualized Data Center (VDC) requires virtualizing the core elements of the data center. A phased approach to virtualize an infrastructure enables a smooth transition from Classic Data Center to Virtualized Data Center.



Virtualized Data Center

Transforming a Classic Data Center (CDC) into a Virtualized Data Center (VDC) requires virtualizing the core elements of the data center.



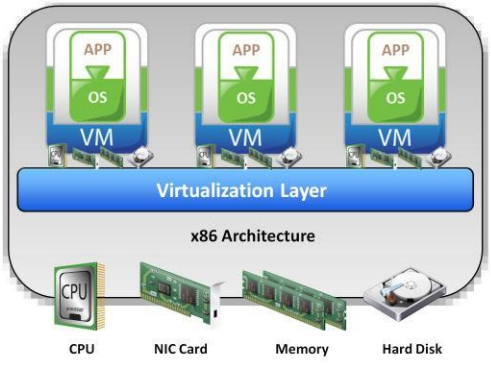
Using a phased approach to a virtualized infrastructure enables smoother transition to virtualize core elements.

Compute Virtualization

Compute Virtualization

It is a technique of masking or abstracting the physical compute hardware and enabling multiple operating systems (OSs) to run concurrently on a single or clustered physical machine(s).

- Enables creation of multiple virtual machines (VMs), each running an OS and application
  - VM is a logical entity that looks and behaves like physical machine
- Virtualization layer resides between hardware and VMs
  - Also known as hypervisor
- VMs are provided with standardized hardware resources



Compute virtualization is a technique of masking or abstracting the physical hardware

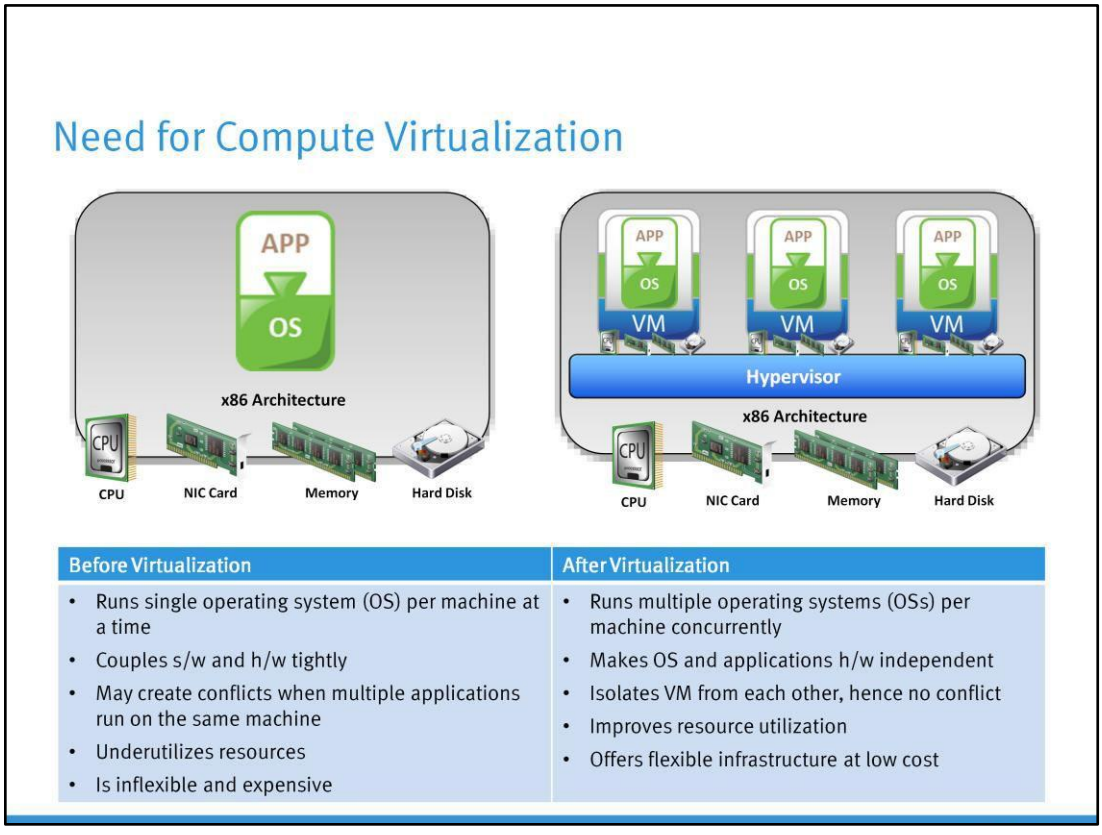
from the operating system and enabling multiple operating systems to run concurrently on a single or clustered physical machine(s). This technique encapsulates an operating system and an application into a portable virtual machine (VM).

A virtual machine is a logical entity that looks and behaves like a physical machine. Each operating system runs on its own virtual machines.

In compute virtualization, a virtualization layer resides between the hardware and virtual machine (on which an operating system is running). The virtualization layer is also known as hypervisor. The hypervisor provides standardized hardware resources (for example: CPU, Memory, Network, etc.) to all the virtual machines.

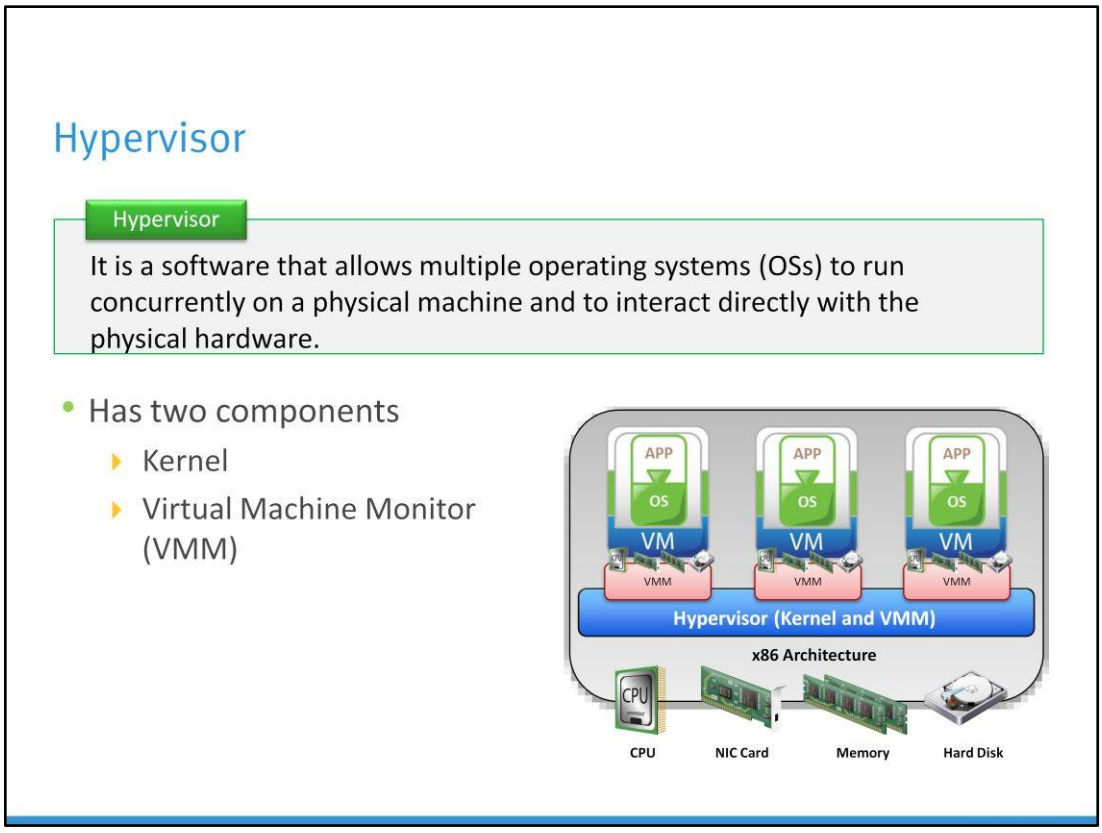
Note:

The terms physical machine, host machine, compute, and server are interchangeably used throughout the course. The terms virtual machine, guest machine, virtual compute, and virtual server are interchangeably used.



Traditionally, one operating system (OS) per compute system (physical machine) is deployed because the operating system and the hardware are tightly coupled and cannot be separated. Only one application is deployed per compute system to minimize the potential resource conflict. This causes organizations to purchase new physical machines for every application they deploy, resulting in expensive and inflexible infrastructure. Further, these compute systems remain underutilized - it is very common to find compute systems running at 15 – 20% utilization. This compounding over many machines within a Classic Data Center (CDC) leads to poor utilization of physical machines.

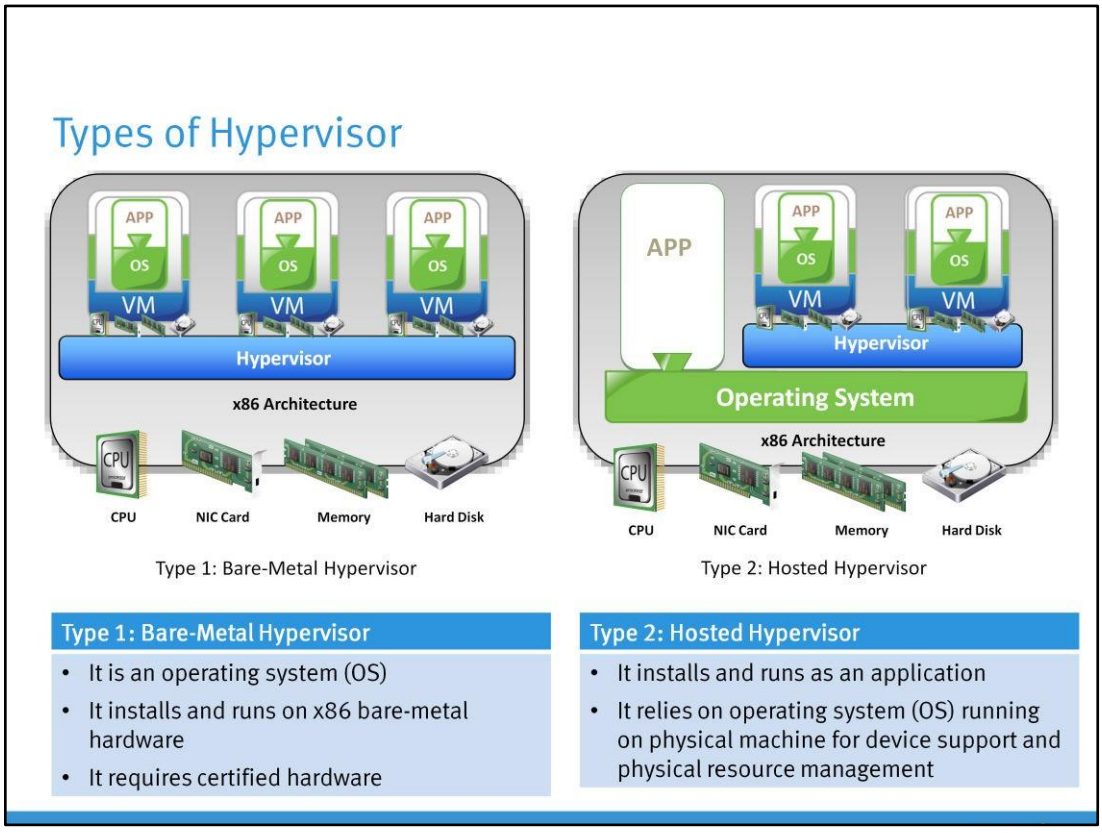
Compute virtualization enables to overcome these challenges by allowing multiple operating systems and applications to run on a single physical machine. This technique significantly reduces acquisition cost and improves utilization.



Hypervisor is a compute virtualization software that enables multiple operating systems (OSs) to run on a physical machine concurrently. The hypervisor interacts directly with the physical resources of the x86 based compute system. The hypervisor is a key component of data center consolidation efforts. By nature, it allows multiple operating systems and applications to reside on the same physical machine.

Hypervisor has two key components: kernel and Virtual Machine Monitor (VMM).

1. A hypervisor kernel provides the same functionality as other operating systems, such as process creation, file system management, and process scheduling. It is designed to specifically support multiple virtual machines and to provide core functionalities, such as resource scheduling, I/O stacks, etc.
2. The Virtual Machine Monitor is responsible for actually executing commands on the CPUs and performing Binary Translation (BT). A Virtual Machine Monitor abstracts hardware to appear as a physical machine with its own CPU, memory, and I/O devices. Each virtual machine is assigned a Virtual Machine Monitor that has a share of the CPU, memory, and I/O devices to successfully run the virtual machine. When a virtual machine starts running, the control is transferred to the Virtual Machine Monitor, which subsequently begins executing instructions from the virtual machine.

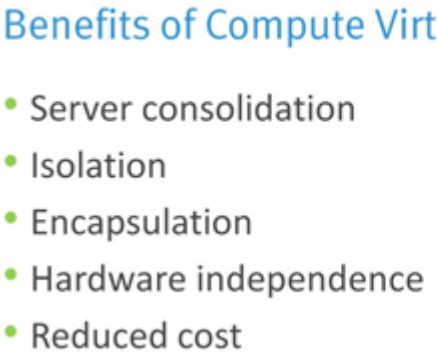


Hypervisors are categorized into two types: hosted hypervisor and bare-metal hypervisor.

Type 1 (Bare-metal hypervisor): In this type, the hypervisor is directly installed on the x86 based hardware. Bare-metal hypervisor has direct access to the hardware resources. Hence, it is more efficient than a hosted hypervisor.

Type 2 (Hosted hypervisor): In this type, the hypervisor is installed and run as an application on top of an operating system. Since it is running on an operating system, it supports the broadest range of hardware configurations.

A hypervisor is the primary component of virtualization that enables compute system partitioning (i.e. partitioning of CPU and memory). We will focus on type 1 hypervisors because it is most predominantly used within Virtualized Data Center (VDC).

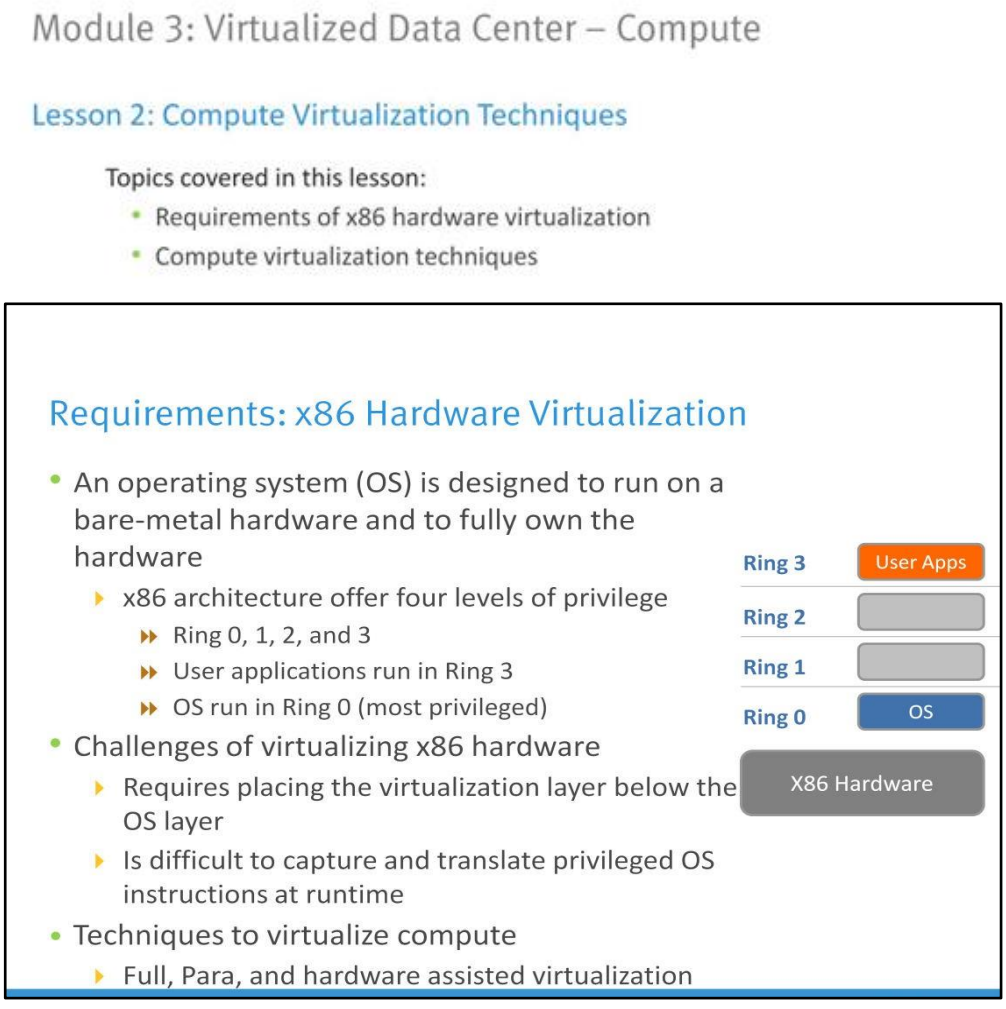


Compute virtualization offers the following benefits:

- **Server Consolidation:** Compute virtualization enables running multiple virtual machines

on a physical server. This reduces the requirement for physical servers.

- **Isolation**: While virtual machines can share the physical resources of a physical machine, they remain completely isolated from each other as if they were separate physical machines. If, for example, there are four virtual machines on a single physical machine and one of the virtual machines crashes, the other three virtual machines remain unaffected.
- **Encapsulation**: A virtual machine is a package that contains a complete set of virtual hardware resources, an operating system, and applications. Encapsulation makes virtual machines portable and easy to manage. For example, a virtual machine can be moved and copied from one location to another just like a file.
- **Hardware Independence**: A virtual machine is configured with virtual components such as CPU, memory, network card, and SCSI controller that are completely independent of the underlying physical hardware. This gives the freedom to move a virtual machine from one x86 machine to another without making any change to the device drivers, operating system, or applications.
- **Reduced Cost:** Compute virtualization reduces the following direct costs:
    - Space (leased or owned) for physical machines, power and cooling, Hardware (including
      switches and Fibre Channel HBA), and annual maintenance



x86 based operating systems (OS) are designed to run directly on the bare-metal

hardware. So, they naturally assume that they fully 'own' the compute hardware. As shown in the figure on this slide, the x86 CPU architecture offers four levels of privilege known as Ring 0, 1, 2, and 3 to operating systems and applications to manage access to the compute hardware. While the user-level applications typically run in Ring 3, the operating system needs to have direct access to the hardware and must execute its privileged instructions in Ring 0. Privileged instruction is a class of instructions that usually includes interrupt handling, timer control, and input/output instructions.

These instructions can be executed only when the compute is in a special privileged mode, generally available to an operating system, but not to user programs.

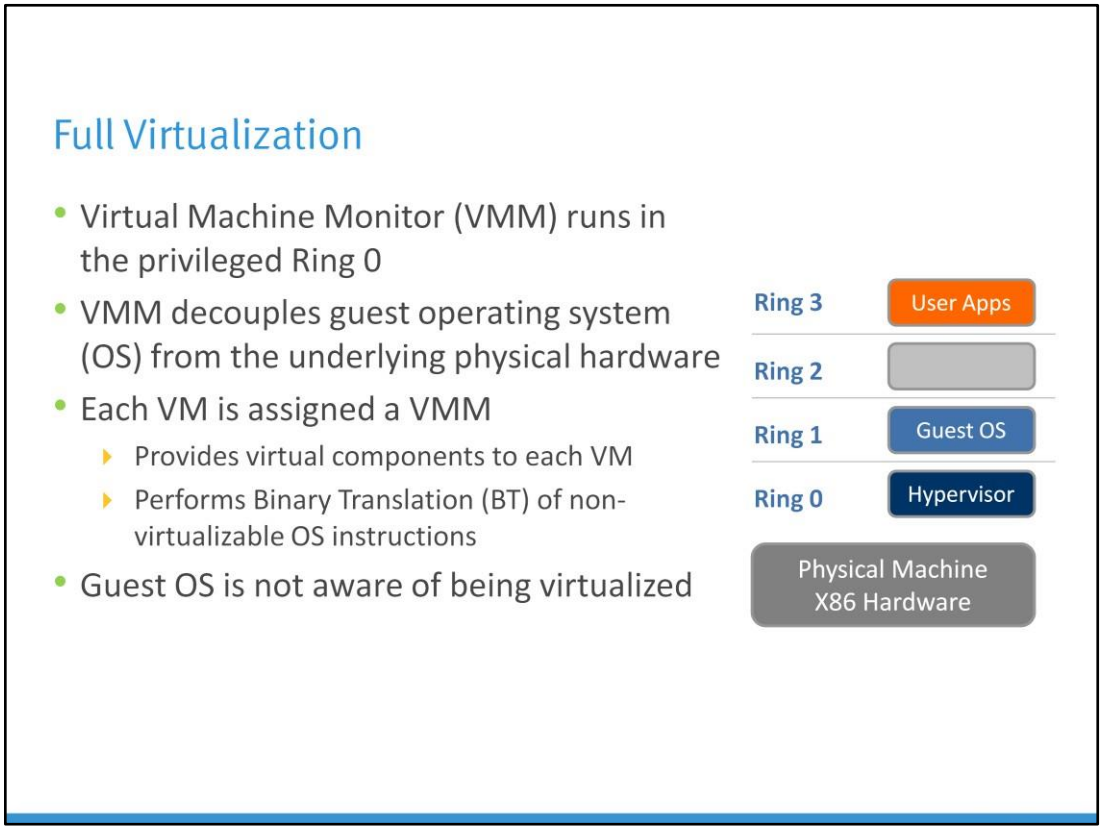
Virtualizing the x86 architecture requires placing a virtualization layer below the operating system (which expects to be in the most privileged Ring 0) to create and manage the virtual machines that deliver shared resources.

Further complicating the situation, some privileged operating system instructions cannot effectively be virtualized because they have different semantics when they are not executed in Ring

1. The difficulty in capturing and translating these privileged instruction requests at runtime was the challenge that originally made x86 architecture virtualization look impossible.

The three techniques that now exist for handling privileged instructions to virtualize the CPU on x86 architecture are as follows:

2. Full virtualization using Binary Translation (BT)
3. Operating systems-assisted virtualization or Paravirtualization
4. Hardware assisted virtualization

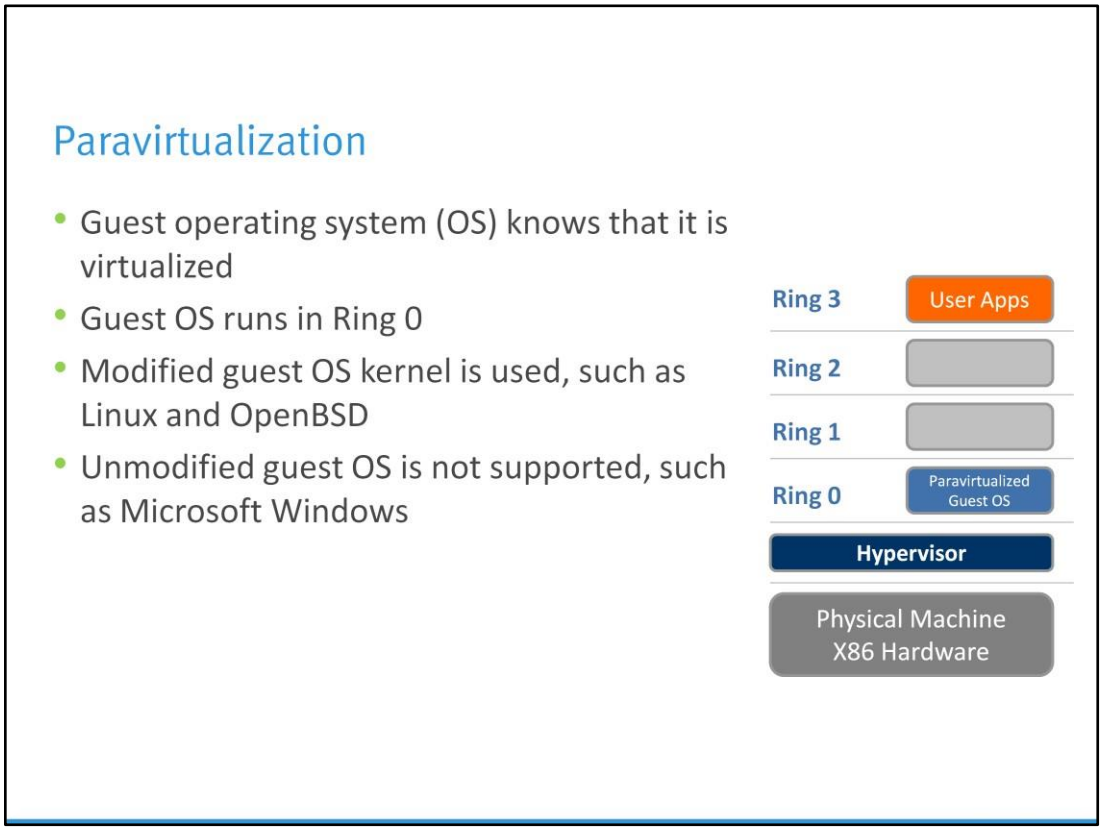


In a full virtualization, Binary Translation (BT) of operating system (OS) instructions is essential. Binary Translation means replacing the guest operating system (an operating system

running on a virtual machine) instructions that cannot be virtualized, with new instructions that have the same effect on the virtual hardware. Application requests work as they would otherwise on a physical machine. Each virtual machine is assigned a Virtual Machine Monitor (VMM), which performs Binary Translation and provides each virtual machine all the services similar to a physical compute, including a virtual BIOS and virtual devices.

Binary Translation provides 'Full Virtualization' because the hypervisor completely decouples the guest operating system from the underlying hardware. The guest operating system is not aware that it is being virtualized and requires no modification.

VMware ESX/ESXi and Microsoft Hyper-V are product examples that implement the full virtualization technique.



In paravirtualization, guest operating systems (OSs) are aware of being virtualized. In this approach, the guest operating system kernel is modified to eliminate the need for Binary Translation. While it is possible to modify open source operating systems, such as Linux and OpenBSD, it is not possible to modify "closed" source operating systems such as Microsoft Windows. Paravirtualization is possible in open source operating systems. A full virtualization approach should be adopted for unmodified guest operating systems such as Microsoft Windows.

Xen and KVM are product examples of paravirtualization.

Hardware Assisted Virtualization

- Achieved by using hypervisor-aware CPU to handle privileged instructions
  - Reduces virtualization overhead caused due to full and paravirtualization
  - CPU and Memory virtualization support is provided in hardware
- Enabled by AMD-V and Intel VT technologies in the x86 processor architecture

Hardware assisted virtualization is accomplished by making hypervisor-aware CPU to handle privileged instructions. Currently, both Intel and AMD offer x86 CPUs with extensions. Though virtualizing the x86 instruction set decreases the hypervisor overhead, it does increase the CPU overhead. Generally, this is seen as negligible, because the machine is usually bound by memory than processing power.

The hardware virtualization support enabled by AMD-V and Intel VT technologies introduces virtualization in the x86 processor architecture. Currently, hardware assisted virtualization supports both CPU and memory virtualization.

Module 3: Virtualized Data Center – Compute

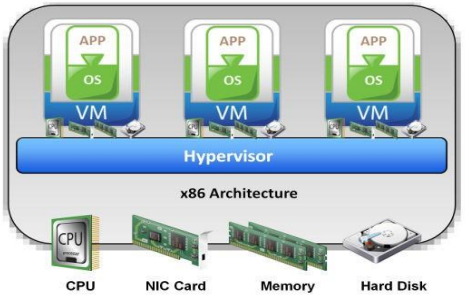
Lesson 3: Virtual Machine

Topics covered in this lesson:

- Virtual machine (VM) files
- File systems that manage Virtual machine files
- Virtual machine hardware
- Virtual machine console

Virtual Machine

- From a user's perspective, a logical compute system
  - Runs an operating system (OS) and application like a physical machine
  - Contains virtual components such as CPU, RAM, disk, and NIC
- From a hypervisor's perspective
  - Virtual machine (VM) is a discrete set of files such as configuration file, virtual disk files, virtual BIOS file, VM swap file, and log file



From a user's perspective, a virtual machine (VM) is a logical compute system just like a physical machine that runs an operating system (OS) and application. An operating system that runs within a virtual machine is called a guest operating system. At a time, only one supported guest operating system can run on a single virtual machine. Each virtual machine is independent and can run its own application.

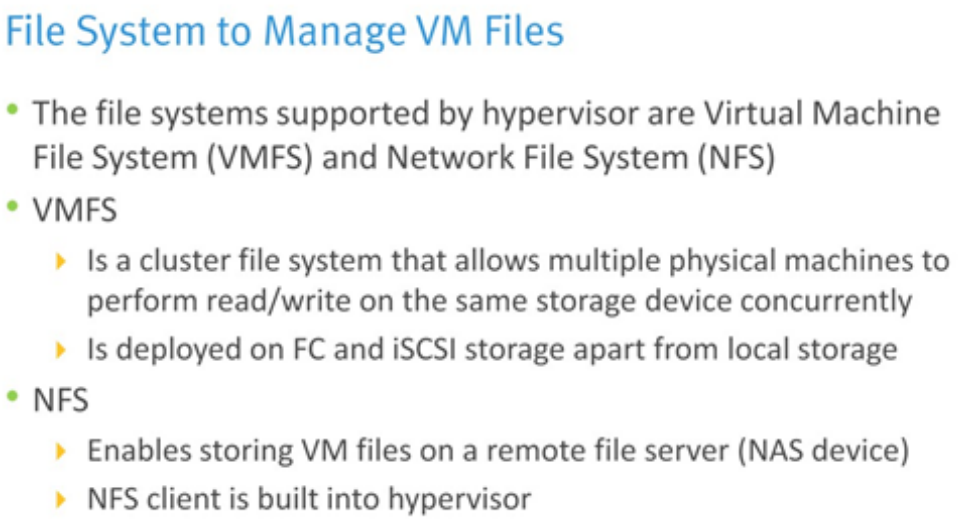
From a hypervisor perspective, a virtual machine is a discrete set of files. The set includes a configuration file, virtual disk files, virtual BIOS file, virtual machine swap file, and a log file.

Virtual Machine Files

File name	Description
Virtual BIOS File	• Stores the state of the virtual machine's (VM's) BIOS
Virtual Swap File	• Is a VM's paging file which backs up the VM RAM contents • The file exists only when VM is running
Virtual Disk File	• Stores the contents of the VM's disk drive • Appears like a physical disk drive to VM • VM can have multiple disk drives
Log File	• Keeps a log of VM activity • Is useful for troubleshooting
Virtual Configuration File	• Stores the configuration information chosen during VM creation • Includes information such as number of CPUs, memory, number and type of network adaptors, and disk types

The table displayed on this slide lists the files that make up a virtual machine (VM).

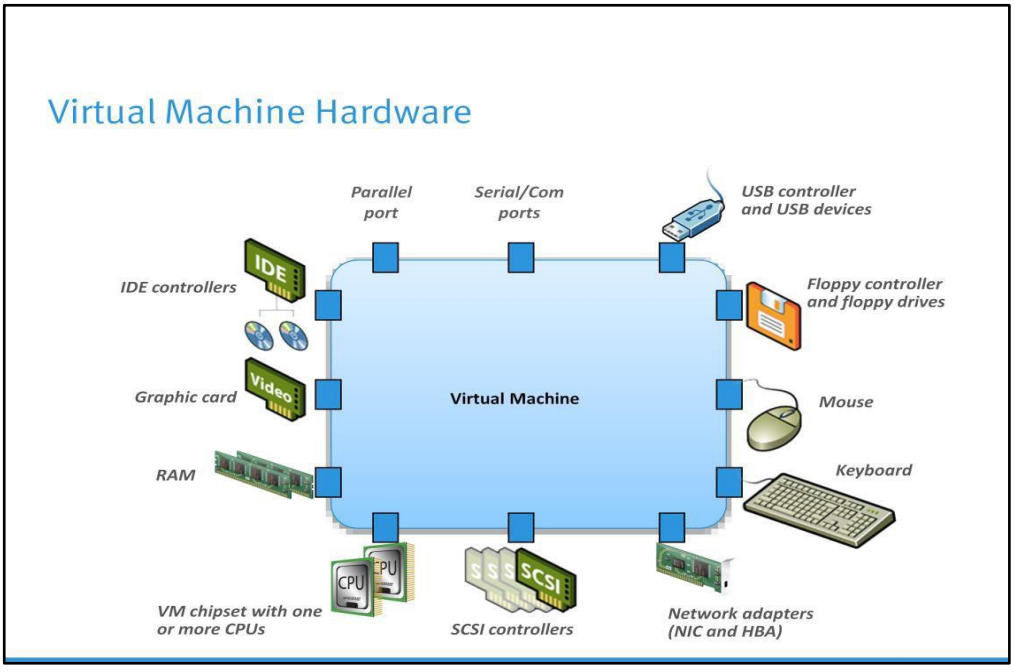
- **Virtual BIOS file**: It stores the state of virtual machine BIOS.
- **Virtual Machine swap file**: It is the paging file of a virtual machine, which backs up the virtual machine RAM contents. This file exists only when the virtual machine is running.
- **Virtual disk file**: It stores the contents in the disk drive of the virtual machine. A virtual disk file appears as a physical disk drive to the virtual machine. A virtual machine can have multiple virtual disk files. Each virtual disk file appears like a separate disk drive.
- **Log file**: This file keeps a log of virtual machine activities. This file may be useful in troubleshooting if a problem is encountered.
- **Configuration file**: It stores the configuration information chosen during creating virtual machines. This includes information such as: virtual machine name, inventory location, guest operating system, virtual disk parameters, number of CPUs and memory sizes, number of adaptors and associated MAC addresses, the networks to which the network adapters connect, SCSI controller type, and the disk type.



The file system supported by hypervisor are Virtual Machine File System (VMFS) and Network File System (NFS).

The Virtual Machine File System is a clustered file system optimized to store virtual machine files. Virtual Machine File System can be deployed on Fibre Channel and iSCSI storage, apart from the local storage. The virtual disks are stored as files on a VMFS.

Network File System enables storing virtual machine files on remote file servers (NAS device) accessed over an IP network. The Network File System client built into the hypervisor uses the Network File System protocol to communicate with the NAS device.



A virtual machine uses a virtual hardware. Each guest operating system sees the hardware devices as if they were physical and owned by them. All virtual machines have standardized hardware.

Standardized hardware makes virtual machine portable across physical machines.

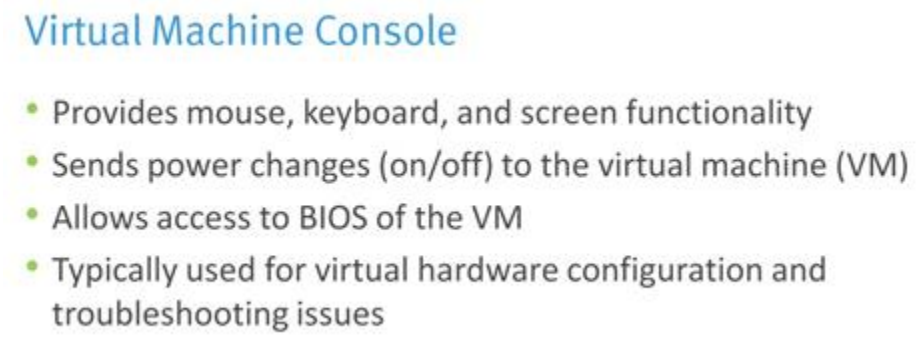
Virtual machine can be configured with a virtual CPU, memory, and other virtual hardware devices such as virtual hard disk, virtual Ethernet cards, virtual CD/DVD drives, virtual floppy drives, USB controllers, and SCSI controllers. These components can be added while creating a new virtual machine or when required. In the current implementation of hypervisors, not all devices are available to add and configure; for example video devices cannot be added, but the available video device can be configured.

VM Hardware Components

Virtual Hardware	Description
vCPU	• Virtual machine (VM) can be configured with one or more virtual CPUs • Number of CPUs allocated to a VM can be changed
vRAM	• Amount of memory presented to the guest operating system (OS) • Memory size can be changed based on requirement
Virtual Disk	• Stores VM's OS and application data • A VM should have at least one virtual disk
vNIC	• Enables a VM to connect to other physical and virtual machines
Virtual DVD/CD-ROM Drive	• It maps a VM's DVD/CD-ROM drive to either a physical drive or an .iso file
Virtual Floppy Drive	• It maps a VM's floppy drive to either a physical drive or an .flp file
Virtual SCSI Controller	• VM uses virtual SCSI controller to access virtual disk
Virtual USB Controller	• Maps VM's USB controller to the physical USB controller

A virtual machine (VM) can be configured with the following virtual components:

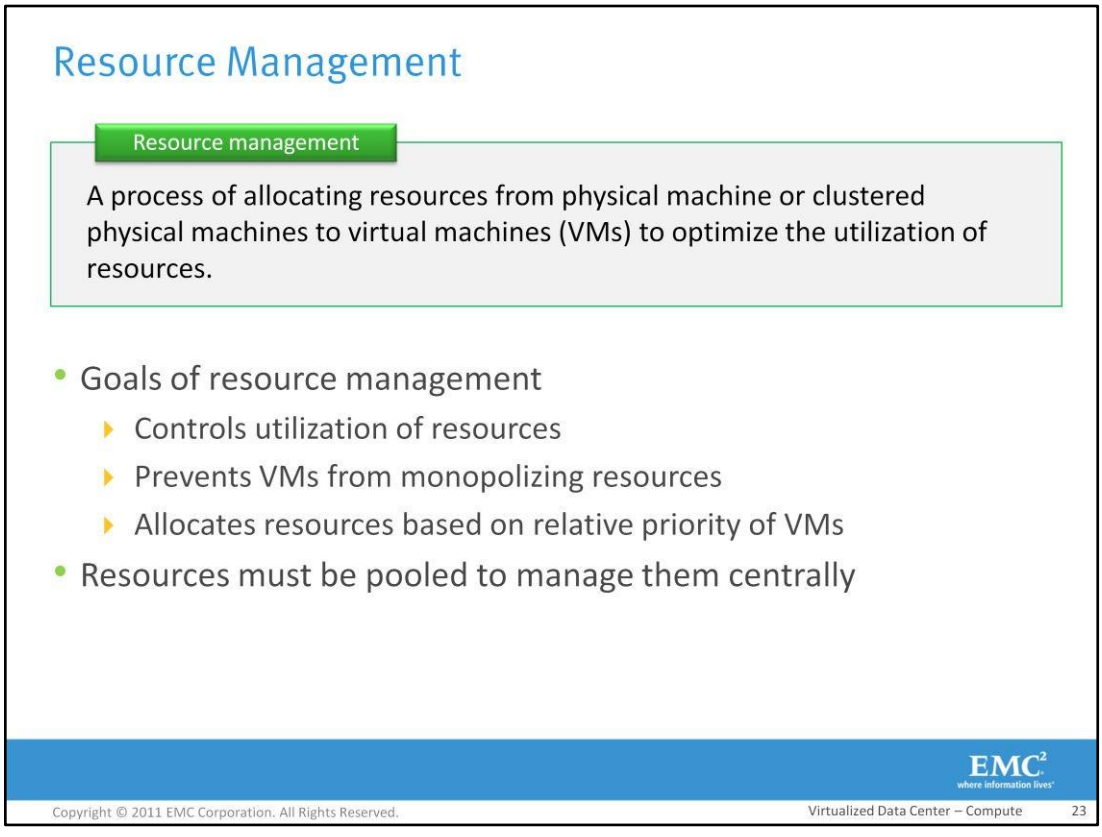
- **Virtual Central processing Unit (vCPU)**: A virtual machine can be configured with one or more vCPU when it is created. The number of vCPUs can be increased or decreased based on requirements.
- **Virtual Random Access Memory (vRAM)**: vRAM is the amount of memory allocated to a virtual machine. It is visible to a guest operating system. This memory size can be changed based on requirements.
- **Virtual Disk**: A virtual disk stores the virtual machine's operating system, program files, application data, and other data associated with the virtual machine. A virtual machine should have at least one virtual disk.
- **Virtual Network Adaptor (vNIC)**: It provides connectivity between virtual machines on the same compute system, between virtual machines on different compute systems, and between virtual and physical machines. vNIC functions exactly like a physical NIC.
- **Virtual DVD/CD-ROM and floppy drives**: These devices enable to map the virtual machines drive to either the physical drive or to the image file (such as .iso for CD/DVD and .flp for floppy) on the storage.
- **Virtual SCSI Controller**: A virtual machine uses virtual SCSI controller to access virtual disks.
- **Virtual USB Controllers**: Enables a virtual machine to connect to the physical USB controller and to access the USB device connected.



A virtual machine (VM) console provides the mouse, keyboard, and screen functionalities. To install an operating system (OS), a virtual machine console is used. The virtual machine console allows access to the BIOS of the virtual machine. It offers the ability to power the virtual machine on/off and to reset it.

The virtual machine console is normally not used to connect to the virtual machine for daily tasks. It is used for tasks such as virtual hardware configuration and troubleshooting issues.

This lesson covers the process of pooling and managing resources, explains how the resources are controlled using share, limit, and reservation. It also describes the CPU and memory optimization techniques.



Resource management is the allocation of resources from a physical machine or clustered physical machines to virtual machines (VMs). It helps optimize the utilization of resources because the demand for resource varies over time. It also allows to dynamically reallocate resources so that the available capacity can be used more efficiently.

In addition to controlling resource utilization, resource management helps prevent virtual machines from monopolizing resources and guarantees predictable service levels. With resource management, resources are allocated based on the relative priority of virtual machines. For example, consider two virtual machines with different priority levels, one at a high level and the other at a low level. In this case, more resources are allocated to the virtual machine with high priority.

In order to allocate resources, they must be pooled and managed centrally.

Note: Resource management includes management of CPU, memory, network, and storage. This module focuses on the management of CPU and memory resources.

Resource Pool

Resource pool

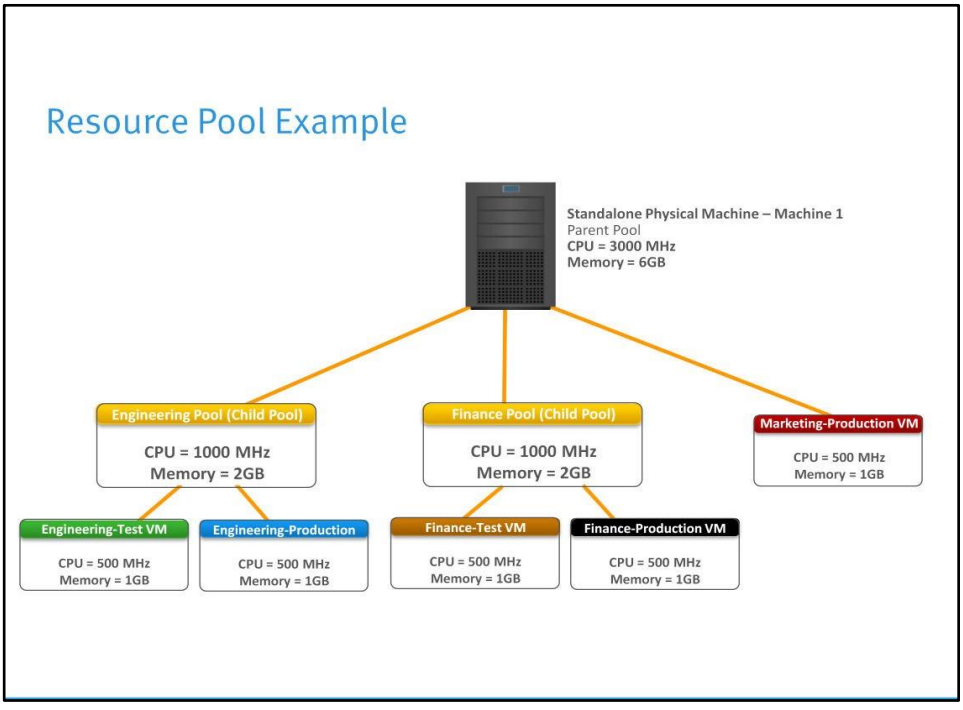
It is a logical abstraction of aggregated physical resources that are managed centrally.

- Created from a physical machine or cluster
- Administrators may create child resource pool or virtual machine (VM) from the parent resource pool
- Reservation, limit, and share are used to control the resources consumed by resource pools or VMs

A resource pool is a logical abstraction of aggregated physical resources that are managed centrally. Each physical machine and cluster has a parent resource pool that groups the resources of that physical machine or cluster. Administrators may create child resource pools from the parent resource pool. Each child resource pool owns some of the parent's resources. A parent resource pool can contain child resource pools, virtual machines, or both.

For each resource pool and virtual machine, reservation, limit, and share can be specified. Reservation, limit, and share are used to control the resources consumed by a child resource pool or virtual machines.

Resource Pool Example

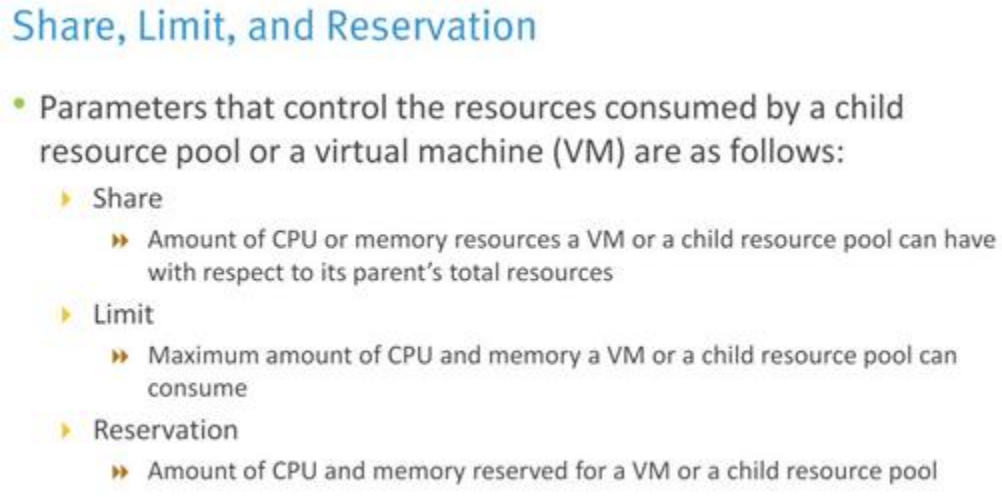


This slide illustrates an example where resources of a parent resource pool are distributed among child resource pools and virtual machines. The resources of the child resource pools are further distributed among virtual machines (VMs).

The parent resource pool includes the sum of all CPUs power (in megahertz) and the sum of all the capacity of installed RAM (in megabytes) available in the compute environment (physical machine or cluster).

On this slide, the parent resource is a physical machine named 'Machine 1'. It has 3000 MHz of CPU and 6GB of RAM, available for use by child resource pools and/or virtual machines.

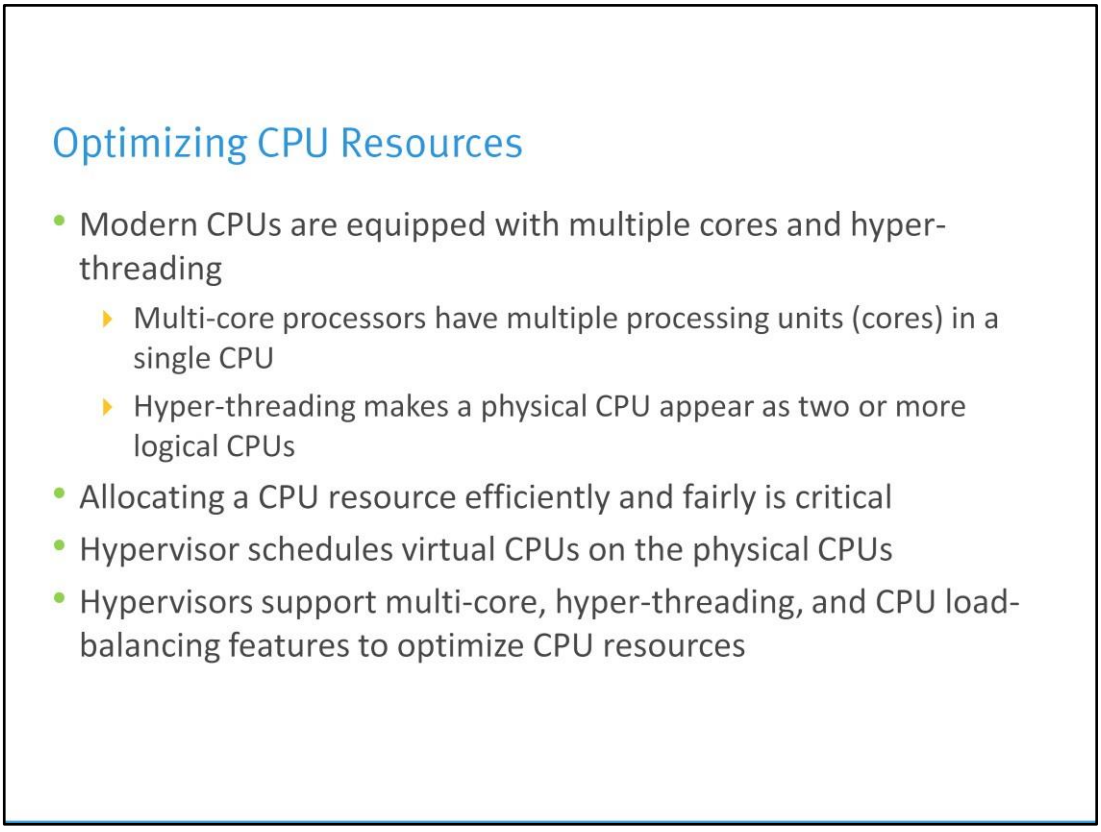
A child resource pool uses allocated resources from the parent resource pool for the virtual machines. The child resource pool cannot exceed the capacity of the parent resource pool. Creating a child pool reserves resources from the parent pool, irrespective of whether or not the virtual machine in the child pool is powered on.



Reservation, limit, and share are mechanisms to control the resources consumed by a child resource pool or virtual machine (VM).

1. Share
    - **Child Resource Pool**: When resources are scarce and when resource contention occurs, the share value defines the relative priority of child resource pools in a parent pool.
    - **Virtual machine**: Similar to a child resource pool, share specifies the relative priority of a virtual machine. If a virtual machine has twice as many CPU/memory share as another Virtual Machine, it is entitled to consume twice as much CPU/memory when the virtual machines are competing for resources.
2. Limit
    - **Child Resource Pool**: It defines the maximum amount of CPU (MHz) and memory (MB) that a child resource pool is allowed to consume.
    - **Virtual machine**: Similar to a resource pool, limit defines the maximum amount of CPU (MHz) and memory (MB) that a virtual machine is allowed to consume. The maximum amount of memory and CPU a virtual machine may consume is configured when it is created, and can only be modified when the virtual machine is powered off.

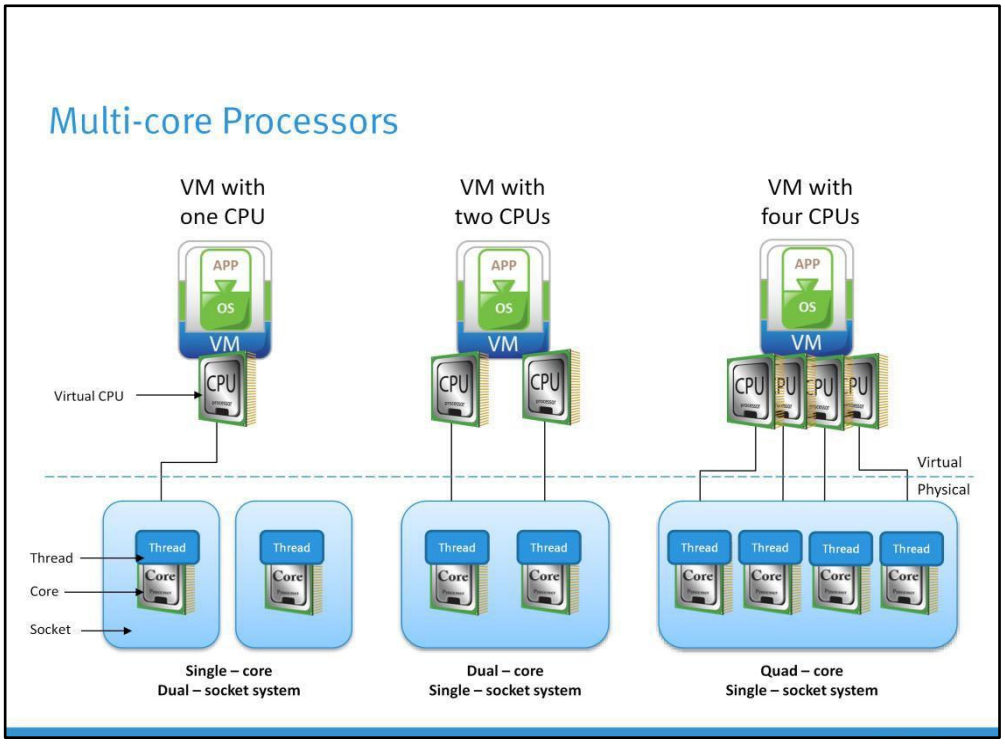
3. Reservation
    - **Child Resource Pool**: It defines the amount of CPU (MHz) and memory (MB) that is reserved for a child resource pool.
    - **Virtual machine**: Similar to child resource pool, virtual machine defines the amount of CPU (MHz) and memory (MB) reserved for a virtual machine. If the virtual machine does not use the total amount of its CPU and memory reserved, the unused portion will be available for use by other virtual machines until the virtual machine needs it. A virtual machine will not power on if the amount of resources defined in reservation is not available in the pool.



Modern CPUs are equipped with multiple cores per CPU and hyper-threading features. A multi-core CPU is an integrated circuit to which two or more processing units (cores) have been attached for enhanced performance and more efficient, simultaneous processing of multiple processes. Hyper- threading makes a physical CPU appear as two or more logical CPUs.

Today's data centers deploy servers with multi core and hyper-threading features in their environment. The role of the hypervisor scheduler is to assign a physical CPU resource to the virtual CPU in a way that meets system objectives, such as responsiveness, throughput, and utilization. A conventional Operating System schedules a process or thread on a CPU, while a hypervisor schedules virtual CPUs of virtual machines on the physical machines.

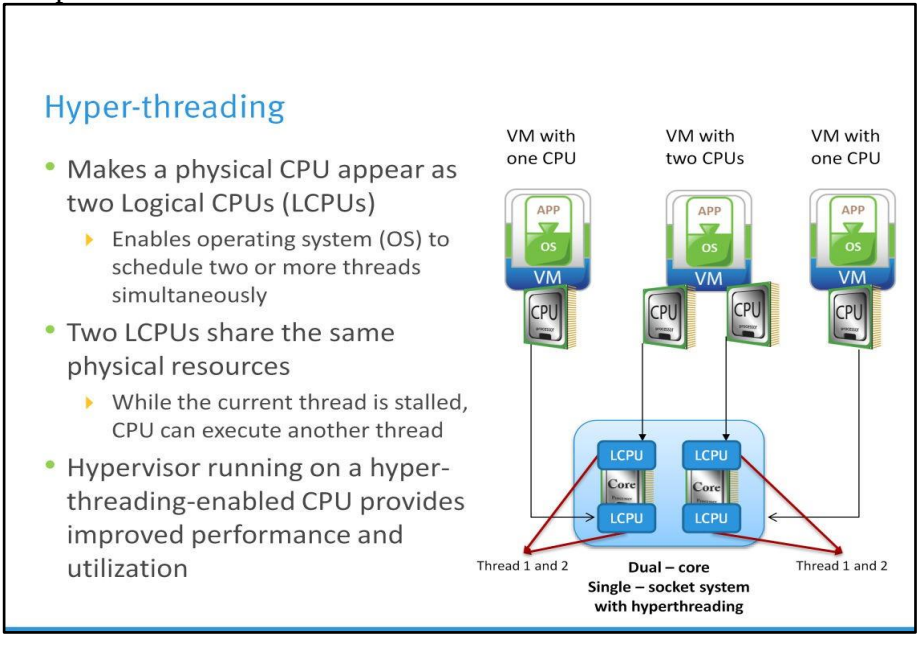
A hypervisor supports and optimizes the CPU resources using modern CPU features such as multi- core and hyper-threading. They also support CPU load balancing.



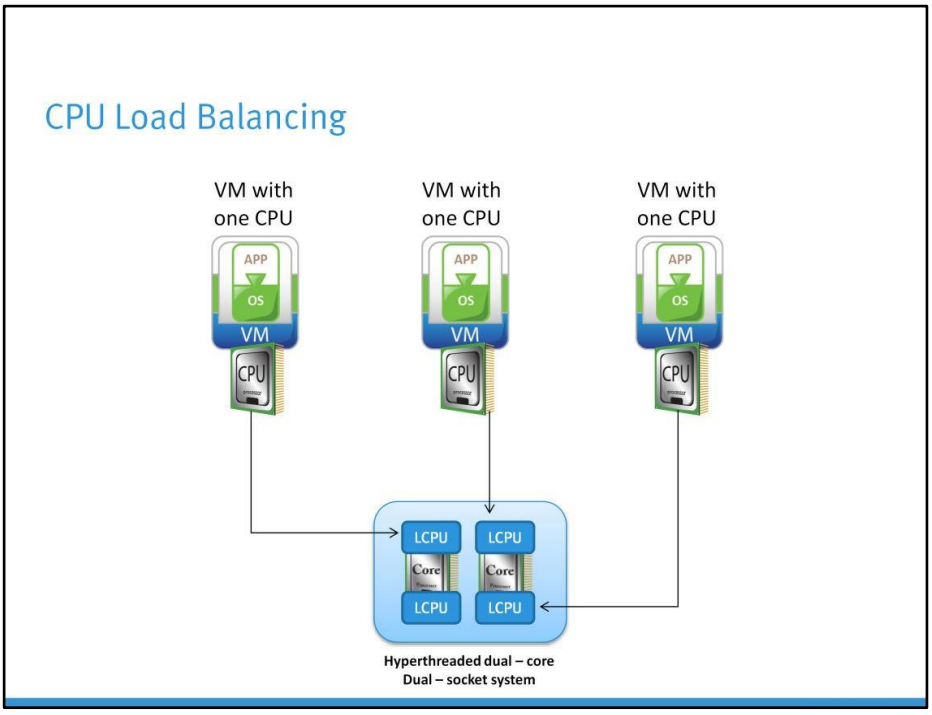
Multi-core CPUs provide many advantages to a hypervisor that performs multitasking of virtual machines. A dual-core CPU, for example, can provide almost double the performance of a single- core CPU by allowing two virtual CPUs to execute at the same time.

Intel and AMD have developed CPUs that combine two or more cores into a single integrated circuit, called a socket.

A hypervisor running on a physical machine can have single core, dual core, or quad core CPUs. Virtual machines can be configured with one or more virtual CPUs. When a virtual machine is scheduled, its virtual CPUs are scheduled to run on a physical CPU by the hypervisor. To maximize the overall utilization and performance, a hypervisor scheduler optimizes the placement of virtual CPUs onto different sockets.



Hyper-threading makes a physical CPU appear as two or more Logical CPUs (LCPUs), allowing the operating system (OS) to schedule two threads or processes simultaneously. The two threads cannot be executed at the same time because the two logical CPUs share a single set of physical resources. When physical resources are not in use by the current thread in a physical CPU, especially when the processor is stalled (due to a cache miss or data dependency), a hyper- threading-enabled CPU can use those execution resources to execute another scheduled task. When a hypervisor runs on a hyper-threading enabled CPU, it provides improved performance and utilization.



When a hypervisor is running on multi-processor and hyper-threading-enabled compute systems, balancing the load across CPUs is critical to the performance. In this environment, load balancing is achieved by migrating a thread from one logical CPU (over utilized) to another (under utilized) to keep the load balanced. The hypervisor intelligently manages the CPU load by spreading it smoothly across the CPU cores in the compute system. At regular intervals, the hypervisor looks to migrate the CPU of a virtual machine (virtual CPU) from one logical CPU to another to keep the load balanced.

If the logical CPU has no work assigned, it is put into a halted state. This action frees its physical resources and allows the virtual machine running on the logical CPU of the same core to use all the physical resources of that core.

Optimizing Memory Resource

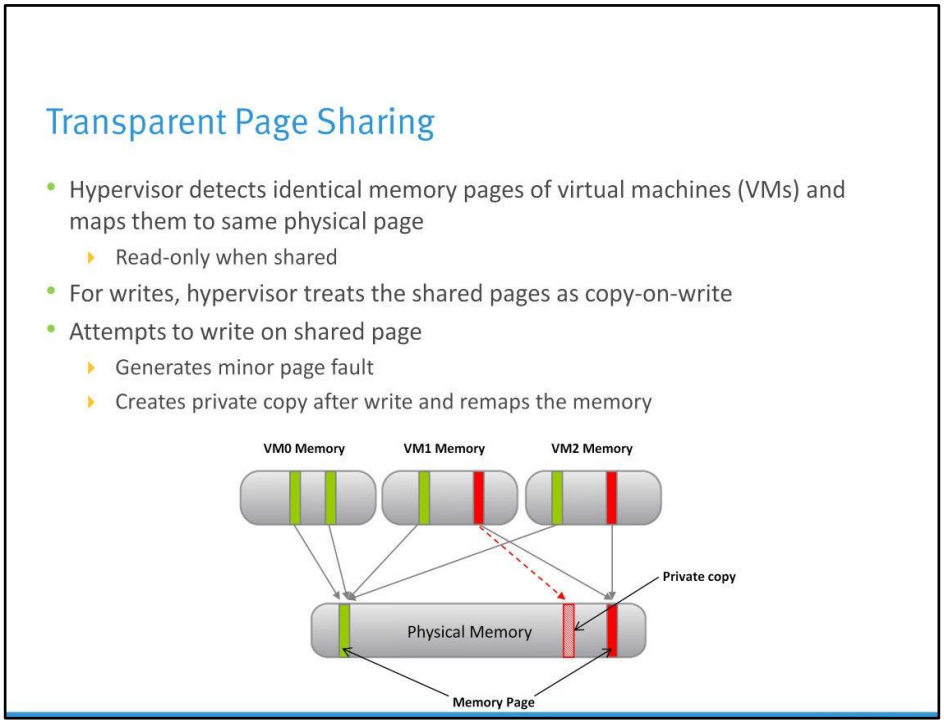
- Hypervisor manages a machine's physical memory
  - Part of this memory is used by the hypervisor
  - Rest is available for virtual machines (VMs)
- VMs can be configured with more memory than physically available, called 'memory overcommitment'
  - Memory optimization is done to allow overcommitment
- Memory management techniques are Transparent page sharing, memory ballooning, and memory swapping

The hypervisor manages the physical memory of compute systems. A part of memory is consumed by the hypervisor and rest of the memory is available for virtual machines (VMs).

Hypervisor allows configuring more memory to the virtual machines than what is physically available. This is known as over commitment of memory. Memory over commitment is allowed because, typically, some virtual machines are lightly loaded, compared to others. Their memory may be used infrequently, and much of the time, their memory remains idle. Memory over commitment allows the hypervisor to use memory reclamation techniques to take the inactive or unused memory away from the idle virtual machines and give it to other virtual machines that will actively use it.

For example, consider a physical machine with 4GB physical memory running three virtual machines with 2GB VM memory each. Without memory over commitment, only one virtual machine can be run because the hypervisor cannot reserve the physical memory for more than one virtual machine. This is because a hypervisor will consume some memory space and each virtual machine has a memory overhead.

In order to effectively support memory over commitment, the hypervisor must provide efficient physical memory reclamation techniques. A hypervisor supports three techniques to reclaim memory: transparent page sharing, ballooning, and memory swapping.



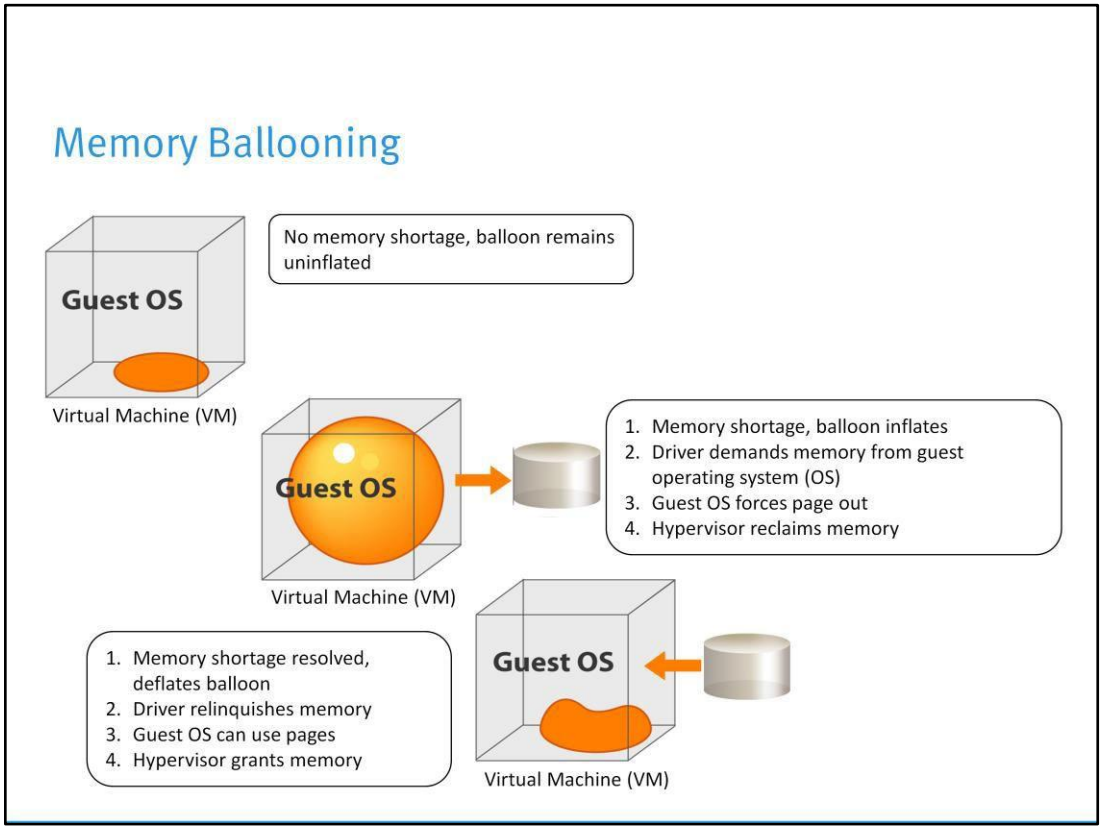
When multiple virtual machines (VMs) are running, some of them may have identical sets of memory content. This presents opportunities to share memory across virtual machines (as well as to share within a single virtual machine). For example, several virtual machines may run the same guest operating system, have the same applications, or contain the same user data. With page sharing, the hypervisor can reclaim the redundant copies and keep only one copy, which is shared by multiple virtual machines in the physical memory. As a result, the total amount of physical memory consumed by virtual machines is reduced and memory over commitment becomes possible.

A hypervisor identifies redundant page copies by their contents. This means that pages

with identical content can be shared regardless of when, where, and how those contents are generated. The hypervisor scans the content of VM memory for sharing opportunities.

After the candidate virtual machine page content is confirmed to match the content of an existing physical page, the virtual machine memory pointer is updated to point to the shared location, and the redundant physical memory copy is reclaimed. This mapping process is invisible to the virtual machine.

When a write occurs on the shared page, the standard Copy-on-Write (CoW) technique is used to handle these writes. Any attempt to write to the shared pages will generate a minor page fault. After a page fault occurs, the hypervisor will transparently create a private copy of the page for that virtual machine. It remaps the pointer to this private copy of the virtual machines. In this way, virtual machines can safely modify the shared pages without disrupting other virtual machines sharing that memory.

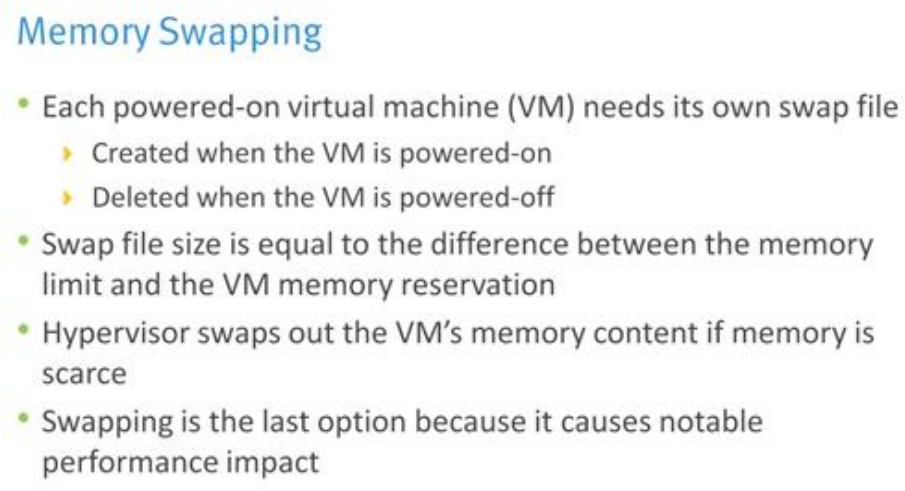


Ballooning is a completely different memory reclamation technique, compared to page sharing. When a virtual machine (VM) must yield memory, the best thing is to let the guest operating system of the VM select the memory pages to give up. The virtual machine knows which pages have been least recently used and can be freed up. Ballooning technique makes the guest operating system free some of the virtual machine memory.

The balloon driver is installed in the guest operating system as a pseudo-device driver and communicates with the hypervisor. The guest operating system is aware of the fact that the balloon driver is installed but not aware of its purpose. The driver's function is to demand memory from the guest operating system and later to relinquish it under the control of the hypervisor. The guest operating system is not aware of the communication taking place between the balloon driver and the hypervisor.

When a compute is not under memory pressure, no virtual machine's balloon is inflated. But when memory becomes scarce, the hypervisor chooses a virtual machine and inflates its balloon. It instructs the balloon driver in the virtual machine to demand memory from the guest operating system. The guest operating system complies by yielding memory, according to its own algorithms. The hypervisor can assign the relinquished pages to other virtual machines that require more memory.

The hypervisor can safely reclaim this physical memory because neither the balloon driver nor the guest operating system relies on the contents of these pages. This means that no processes in the virtual machine will intentionally access those pages to read or write. If any of these pages are re- accessed by the virtual machine for some reason, the hypervisor will treat it as normal virtual machine memory allocation and allocate a new physical page for that virtual machine.



As a last effort to manage excessively overcommitted physical memory, the hypervisor will swap the virtual machine's (VM's) memory content to their swap files. When virtual machines are powered on, the hypervisor creates and assigns one swap file to each virtual machine. This swap file stores the virtual machine's memory contents. If a physical machine cannot get enough memory through page sharing and memory ballooning, the hypervisor forcibly reclaims memory from virtual machines by memory swapping. The hypervisor copies the VM page contents to their corresponding swap files before assigning the pages to the virtual machines that need memory.

The swap file size is determined by the difference between the virtual machine's configured memory (or its memory limit) and its reservation.

Whenever a swap file is actively used, performance is severely penalized. When the virtual machine is powered off, the swap file of that virtual machine is deleted. When the virtual machine is powered back on, the swap file for that virtual machine is re-created.

Virtual Machine Affinity

- VM to VM affinity:
  - Selected VMs should run on same hypervisor
    - To improve performance ,if VMs are communicating with each other heavily
  - Anti-affinity ensures that selected VMs are not together on a hypervisor (ex: for availability reasons)
- VM to physical server affinity:
  - Specify whether selected VM can be placed only on a particular hypervisor (ex: for licensing issues)
  - Anti-affinity is allowing VM to move on different hypervisors in a cluster (ex: for high availability or performance requirements)

Understanding virtual machine affinity is very important, especially when considering VM migrations within a virtualized data center or to the Cloud. VM affinity can be classified as affinity between individual VMs, and affinity between a VM and a hypervisor in a clustered server environment.

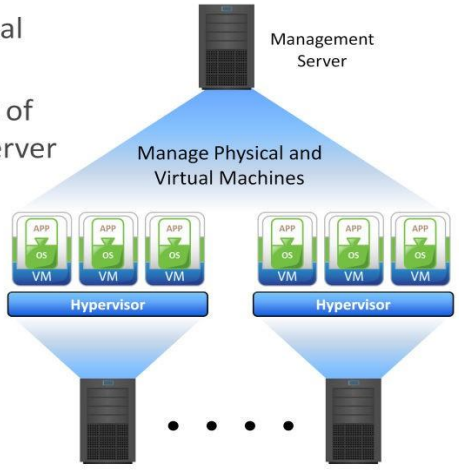
Affinity between virtual machines specifies that either the selected virtual machines be placed on the same host (affinity) or on different hosts (anti-affinity). Keeping virtual machines together may be beneficial in terms of performance, if VMs are communicating heavily with one another.

Conversely, anti-affinity requires selected VMs to be placed on different hypervisors, probably for availability or load balancing reasons.

VM to hypervisor (or physical server) affinity relationship specifies whether a virtual machine can be placed only on a particular physical server or it is allowed to migrate on a different server, if required.

Resource Management Tool

- Provides ability to manage physical machines running hypervisor
- Enables centralized management of resources from a management server
- Enables pooling of resources and allocates capacity to VMs
  - Communicates with hypervisors to perform management
- Provides operational automation



A Virtualized Data Center (VDC) can have several physical machines running

hypervisors. The resource management tool provides the ability to manage these physical machines centrally. Resource management tools, which run on a management server, enable pooling of resources and allocating capacity to VMs. It communicates with hypervisors to perform management operations. It provides administrators better control, simplified day-to-day tasks, and reduced complexity and cost of managing an IT environment. It also provides operational automation that allows task scheduling and alerting to improve responsiveness to business needs and prioritizes actions requiring the most urgent attention.

Module 3: Virtualized Data Center – Compute

Lesson 5: Physical to Virtual Conversion

Topics covered in this lesson:

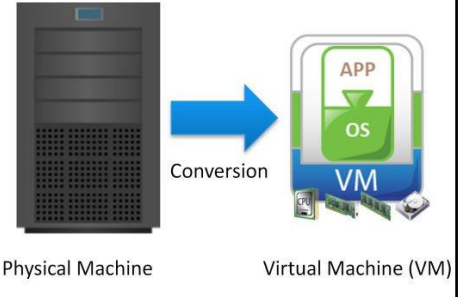
- Converter components
- Conversion options
- Conversion process
- Conversion considerations

Physical to Virtual Machine (P2V) Conversion

P2V Conversion

It is a process through which physical machines are converted into virtual machines (VMs).

- Clones data from physical machine's disk to VM disk
- Performs system reconfiguration of the destination VM such as:
  - Change IP address and computer name
  - Install required device drivers to enable the VM to boot

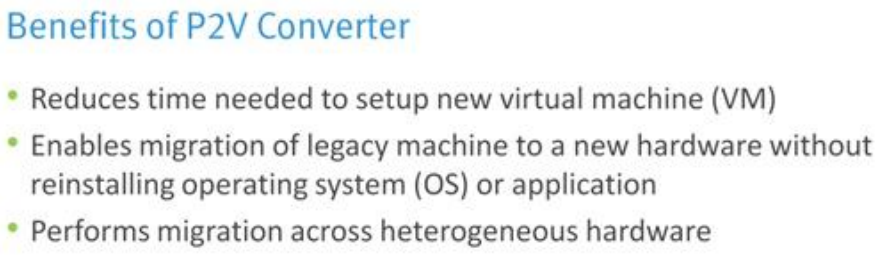


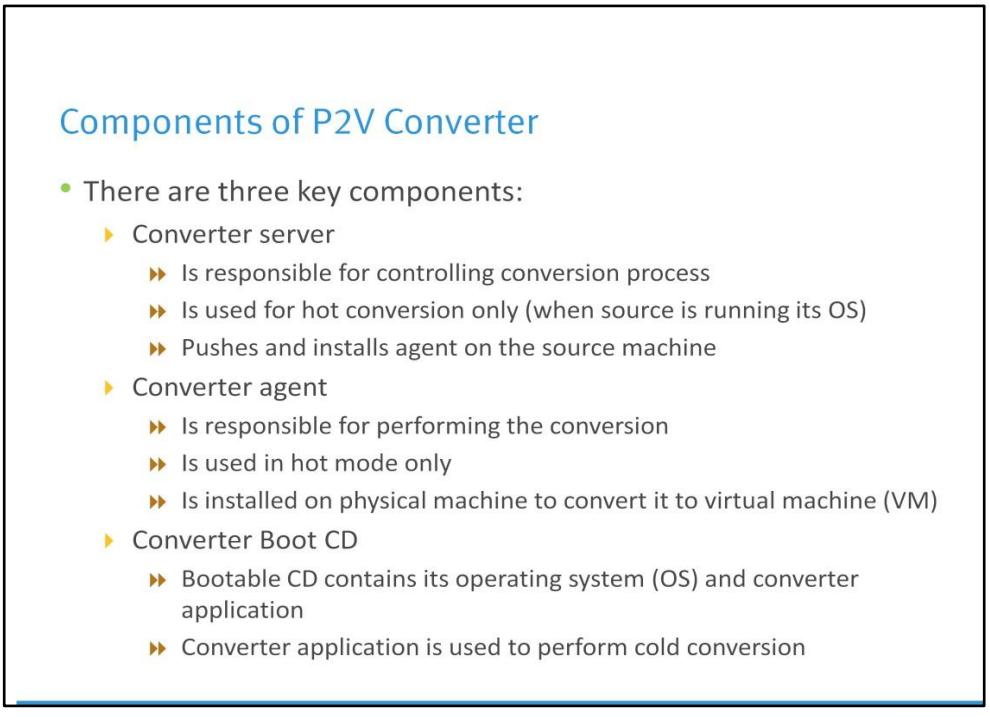
Physical to virtual machine (VM) conversion is a process through which a physical machine is converted into a virtual machine. When converting a physical machine, the "Converter Application" (Converter) clones data on the hard disk of the source machine and transfers that data to the destination virtual disk. Cloning is the process of creating a cloned disk, where the cloned disk is a virtual disk that is an exact copy of the source physical disk.

After cloning is complete, system reconfiguration steps are performed to configure the destination virtual machine. System reconfiguration is the process of configuring the migrated

operating system to enable it to function on a virtual hardware. This configuration is performed on the target virtual disk after cloning, and enables the target virtual disk to function as a bootable system disk in a virtual machine.

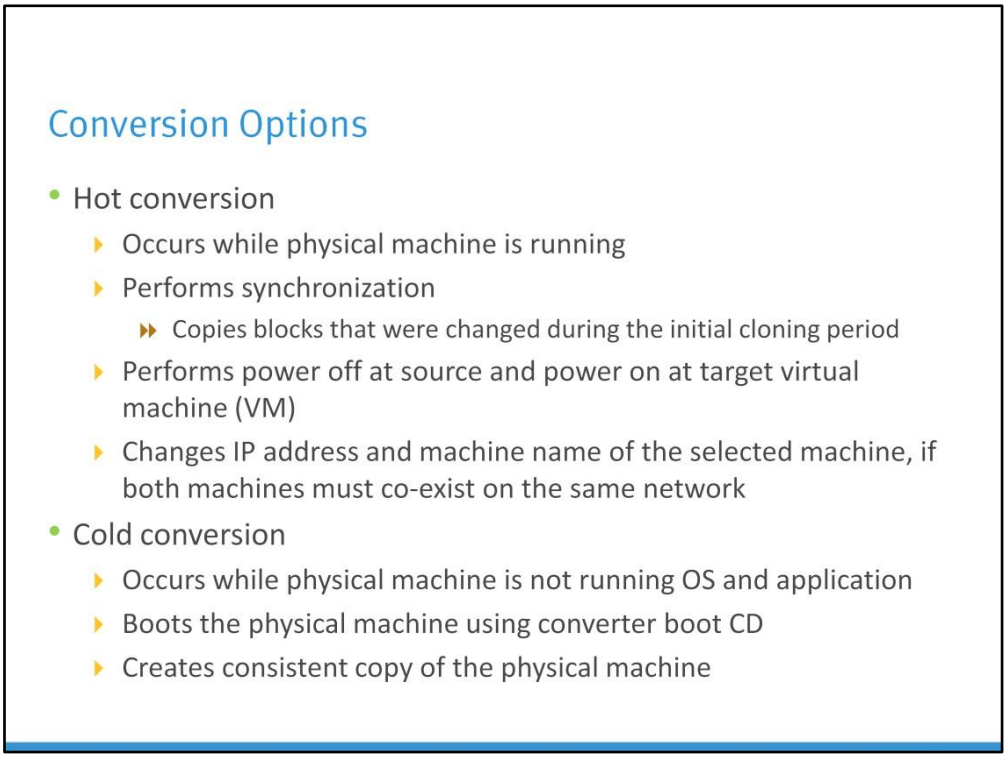
Because the migration process is non-destructive to the source, it can continue to be in use after the conversion is complete.





The P2V "converter application" consists of three components: converter server, converter agent,
and converter boot CD.

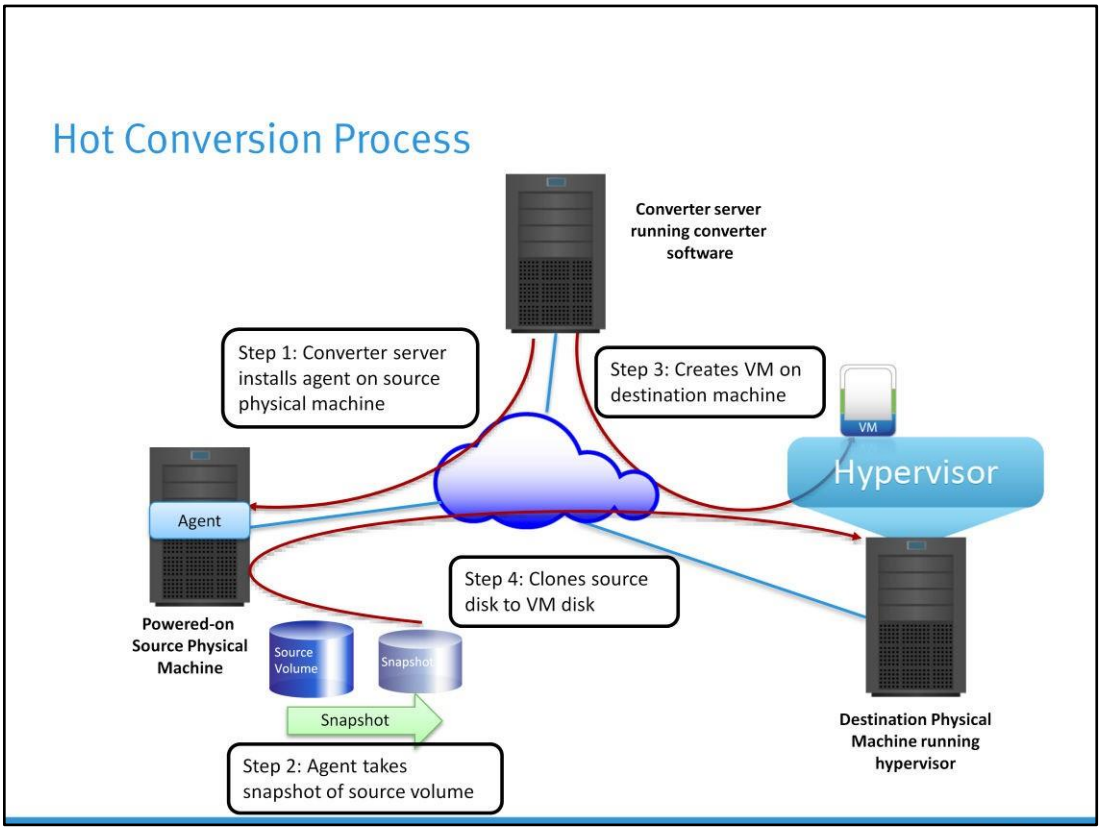
1. A converter server is an application that is loaded on a separate physical machine. It controls the conversion process in a hot mode (when the source machine is running its operating system). While performing the conversion, the converter server pushes and installs a converter agent on the source physical machine that needs conversion.
2. A converter agent is responsible for performing the physical to virtual machine conversion. This agent is installed on the physical machine only for hot conversion.
3. A converter boot CD is a bootable CD with its operating system and converter application on it. This CD is used to perform cold conversion (when the source machine is not running its operating system).



There are two ways to migrate from physical machine to virtual machine (VM). These are hot migration and cold migration.

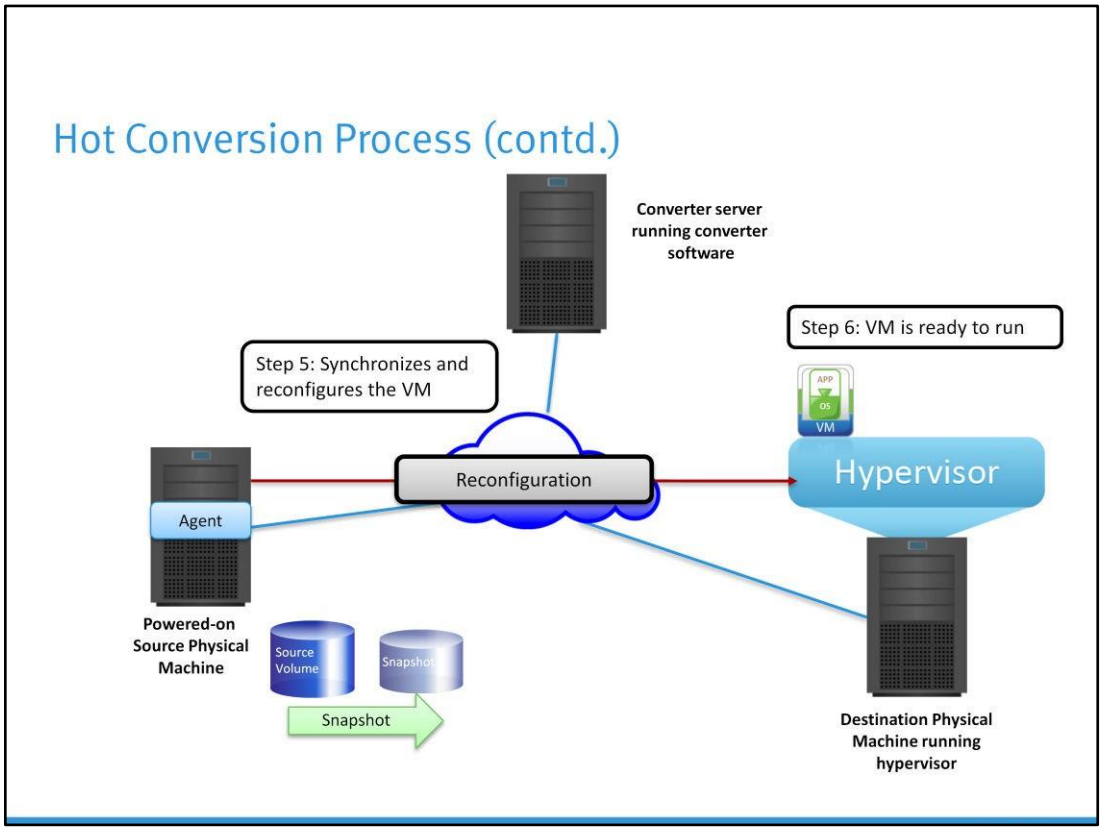
Hot conversion involves converting the source physical machine while it is running its operating system (OS). Because processes continue to run on the source machine during conversion, the resulting virtual machine is not an exact copy of the source physical machine. After conversion is complete, the destination virtual machine is synchronized with the source machine. During synchronization, blocks that were changed during the initial cloning period are transferred from the source to the destination. After conversion is completed, source machine may be powered off and destination virtual machine is commissioned for production. If the source physical machine and the destination virtual machine coexist on the same network, then the machine name and the IP address of the selected machine must be changed.

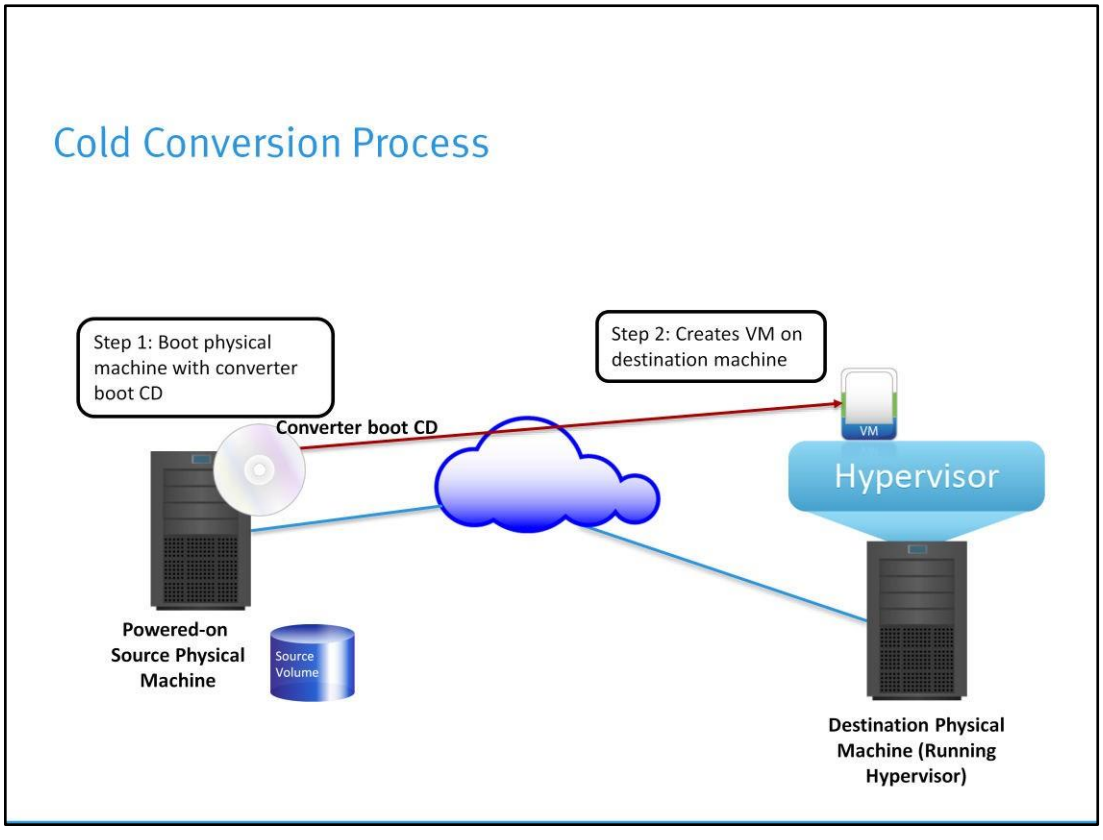
Cold conversion, also called offline conversion, is an option in which conversion of the source physical machine is performed when it is not running its operating system. When performing cold conversion of a physical machine, the source machine is rebooted using a converter boot CD that has its own operating system and converter application. Cold conversion creates a consistent copy of the source physical machine because no changes occur on the source machine during the conversion.



Hot conversion of a physical machine to virtual machine (VM) involves following steps.

1. The converter server prepares the source machine for the conversion by installing the agent on the source physical machine.
2. The agent takes a snapshot of the source volume.
3. The converter server creates a virtual machine on the destination machine.
4. The agent clones the physical disk of source machine (using snapshot) to the virtual disk of the destination virtual machine.
5. The agent synchronizes the data and installs the required drivers to allow the operating system (OS) to boot from a virtual machine (VM) and personalize the virtual machine (changes the IP address and machine name, for example).
6. The virtual machine is ready to run on the destination server.



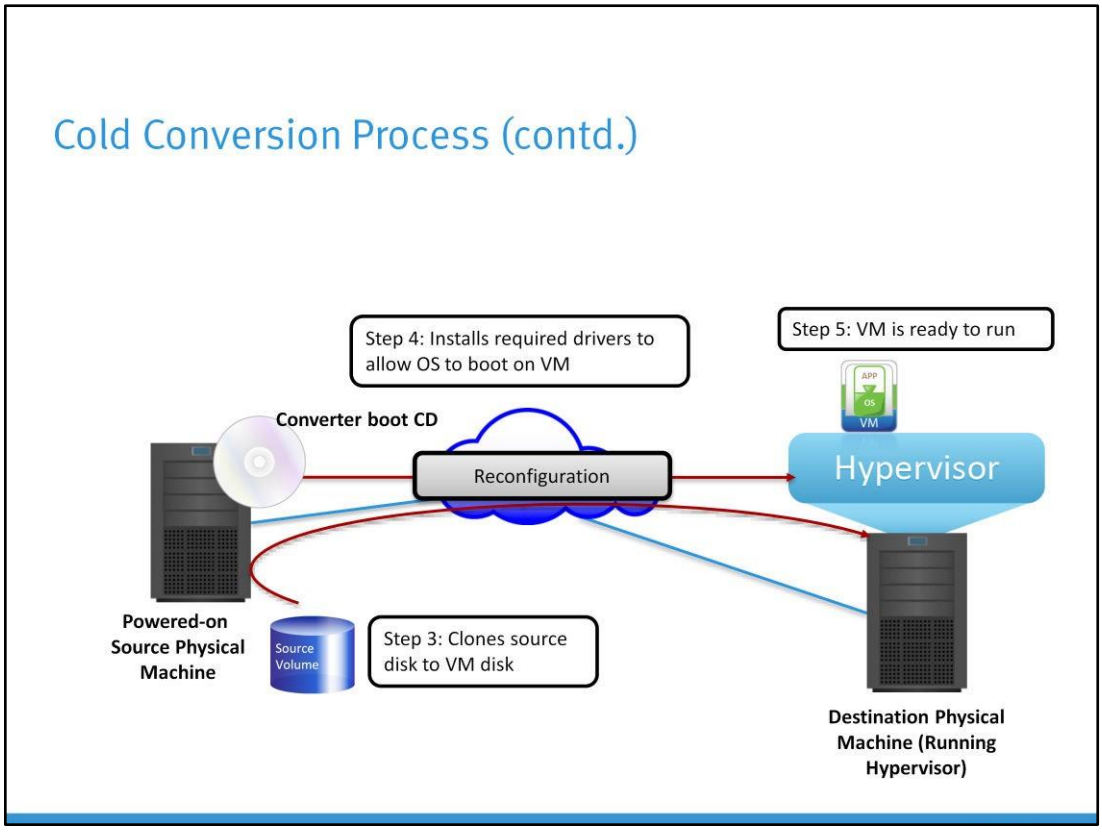


While performing cold conversion of a machine, reboot the source physical machine from a converter boot CD that has its own operating system (OS) and include the converter

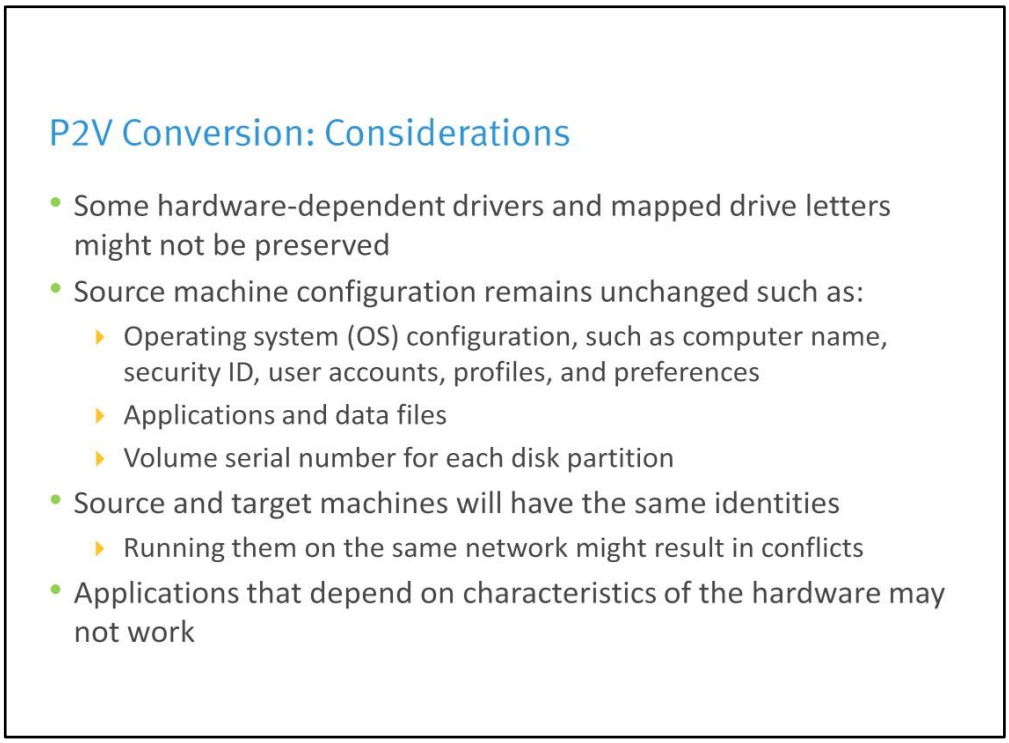
application.

Cold conversion of a physical machine to virtual machine (VM) involves following steps:

1. Boot the source machine from the converter boot CD and use the converter software to define conversion parameters and start the conversion.
2. The converter application creates a new virtual machine on the destination physical machine.



3. The converter application copies volumes from the source machine to the destination machine.
4. The converter application installs the required drivers to allow the operating system to boot in a virtual machine and personalizes the virtual machine (for example, changing the IP address and machine name).
5. The virtual machine is ready to run on the destination server.



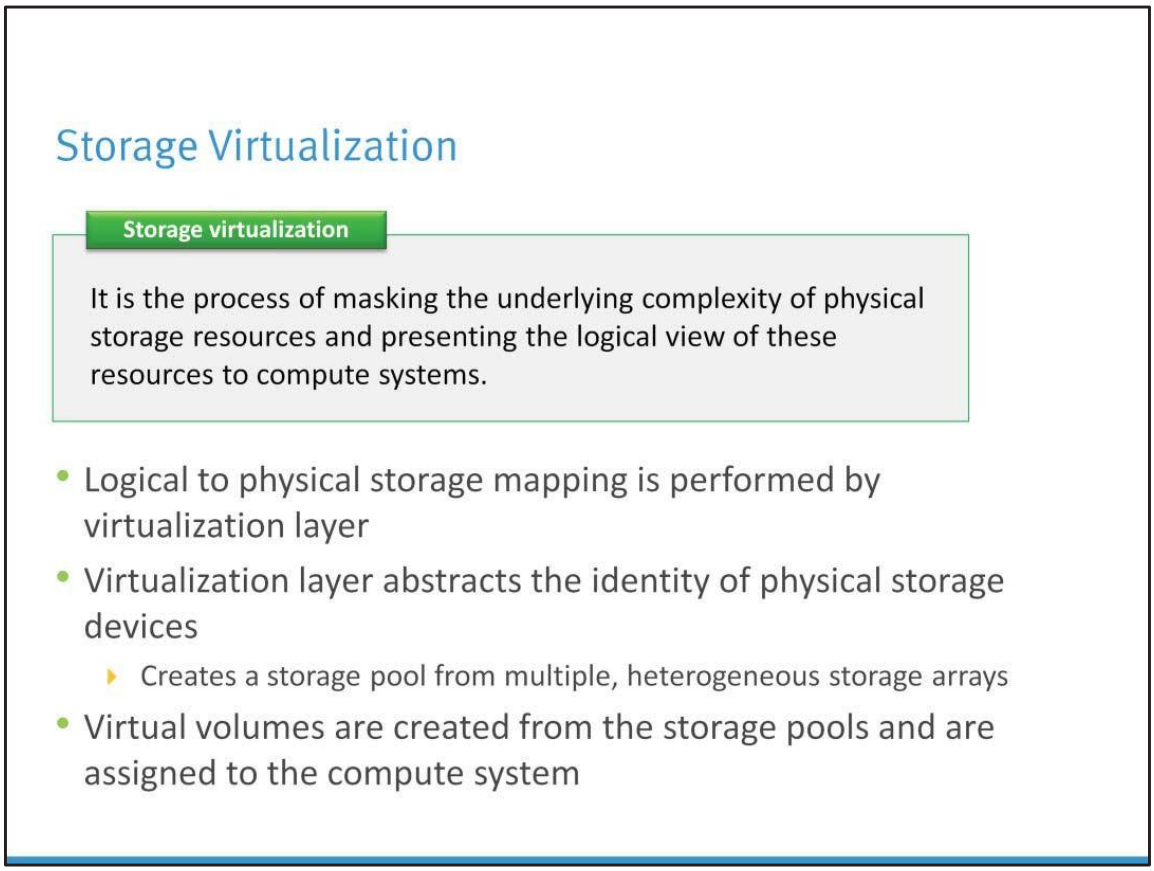
While performing conversion of physical to virtual machine (VM), you should consider the following key points:

1. A virtual machine created by the converter application contains a copy of the disk state of the source physical machine. During this conversion, some hardware-dependent drivers and sometimes the mapped drive letters may not be preserved.
2. The following source machine configurations remain unchanged:
    - Operating system configuration (computer name, security ID, user accounts, profiles, preferences, and so on)
    - Applications and data files
    - Volume serial number for each disk partition

    Because the target and the source virtual machines have the same identities, running them on the same network might result in conflicts.
3. After conversion, most applications function correctly on the virtual machine because their configuration and data files have the same location as on the source machine. However, applications might not work if they depend on specific characteristics of the underlying hardware, such as the serial number or the device manufacturer.

This module focuses on storage virtualization implementation, key underlying technologies, and methods for providing virtual storage to compute systems in a VDC environment.

This lesson covers the key benefits of storage virtualization and the various layers at which the storage virtualization technologies are implemented.



Storage virtualization is the process of masking the underlying complexity of physical storage resources and presenting the logical view of these resources to compute systems in a VDC environment. Storage virtualization enables creating one or more logical storage on the physical storage resources. This logical or virtual storage appears as physical storage to the compute systems. The logical to physical storage mapping is performed by storage virtualization layer. The virtualization layer abstracts the identity of physical storage devices and creates a storage pool by aggregating storage resources from multiple heterogeneous storage arrays.

Virtual volumes are created from these storage pools and are assigned to the compute system. Compute system remains unaware of the mapping operation and access the virtual volumes, as if accessing physical storage attached to them.

Benefits of Storage Virtualization

- Adds or removes storage without any downtime
- Increases storage utilization thereby reducing TCO
- Provides non-disruptive data migration between storage devices
- Supports heterogeneous, multi-vendor storage platforms
- Simplifies storage management

Storage virtualization enables adding or removing storage without affecting application availability. It increases storage utilization by consolidating multiple heterogeneous storage resources and creating storage pools. Storage pools provide flexibility in storage resources allocation to the compute system and increase storage utilization. This considerably reduces investment in new storage resources and thereby lowers the Total Cost of Ownership (TCO).

Data migration is required during technology refresh initiatives where newer storage systems replace the legacy storage systems. It may also be required to move data between different storage systems when performance and availability requirements change. Storage virtualization enables non-disruptive data migration (data is being accessed while migrations are in progress) between storage systems since it masks the complexity of underlying physical storage resources. It also enables support of heterogeneous, multi-vendor storage platforms.

In a virtual environment, different virtual machines running different applications can exist on a single compute system. This creates a very complex environment as there exist multiple sets of workloads and requirements. It can be extremely challenging to manage such an environment in a traditional storage environment. However, by implementing storage virtualization, the complexity in storage provisioning is removed from the environment.
Features such as storage pools, storage tiering, and virtual provisioning are key advances in storage technologies that provide this simplification.

Storage Virtualization at Different Layers

Layers	Examples
Compute	• Storage provisioning for VMs
Network	• Block-level virtualization • File-level virtualization
Storage	• Virtual Provisioning • Automated Storage Tiering

Storage virtualization may be implemented at compute, network, and storage layers. At the compute layer, hypervisor allocates storage space for VMs without exposing the complexity of the physical storage. Block and file level virtualization are network-based virtualization techniques which embed intelligence of virtualizing storage resources at network layer. At storage layer, both

virtual provisioning and automated storage tiering simplify the storage management and help to optimize the storage infrastructure. You will learn about each of these concepts later in this module.

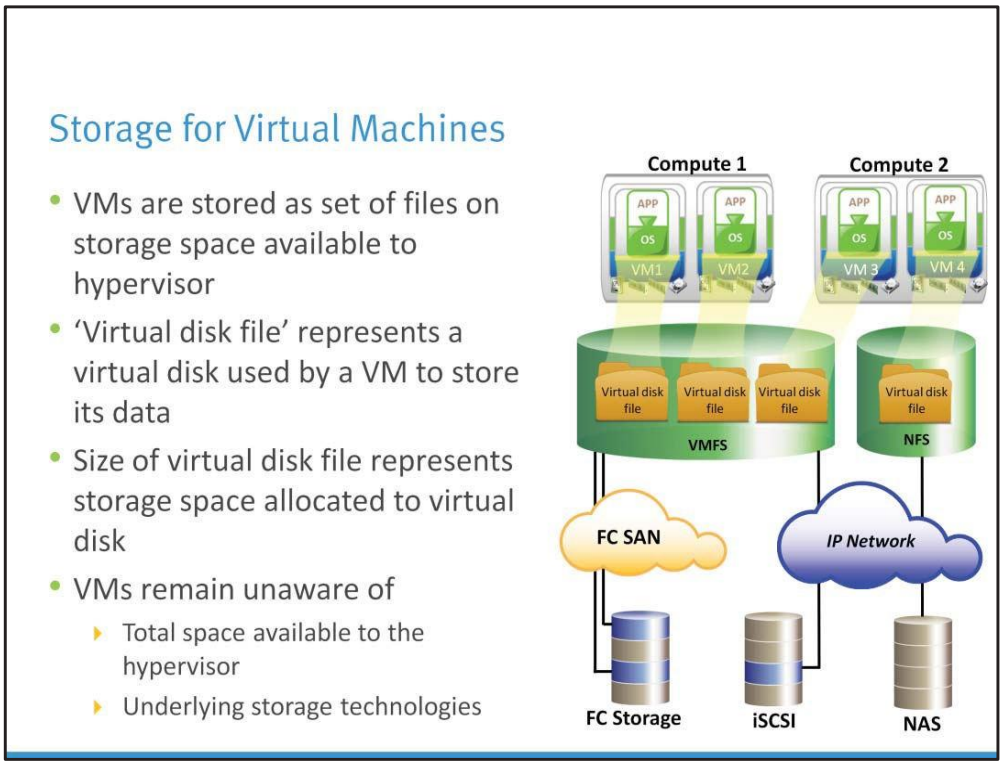
Module 4: Virtualized Data Center – Storage

Lesson 2: Virtual Machine Storage

Topics covered in this lesson:

- Virtual machine storage options
- Virtual machine storage considerations

This lesson covers the various virtual machine storage options that are used to provide storage for VMs running on different compute systems and also covers considerations for provisioning storage to the VMs.



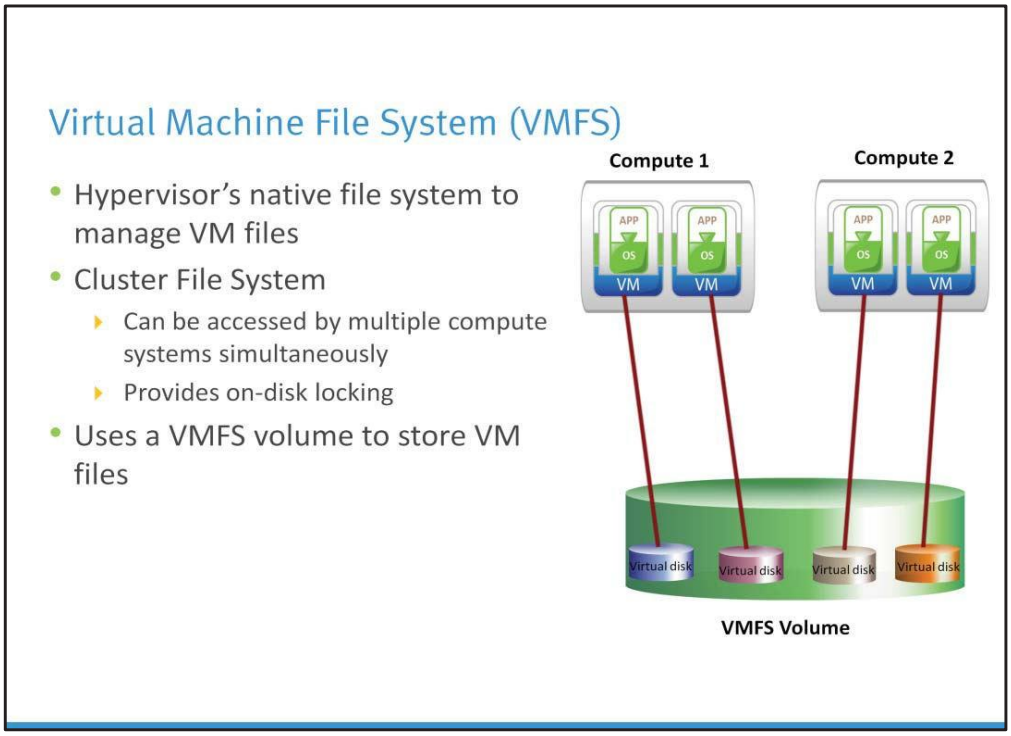
A 'Virtual Machine' is stored as a set of 'files' on the storage devices assigned to the hypervisor. One of the file called 'virtual disk file' represents a 'virtual disk' which is used by a VM to store its data. The virtual disk appears as a local physical disk drive to the VM. The size of the 'virtual disk file' represents the storage space allocated to the 'virtual disk'.

Hypervisor may access FC storage device, or IP storage devices such as iSCSI, and NAS devices. Virtual machines remain unaware of the total storage space available to the hypervisor and the underlying storage technologies.

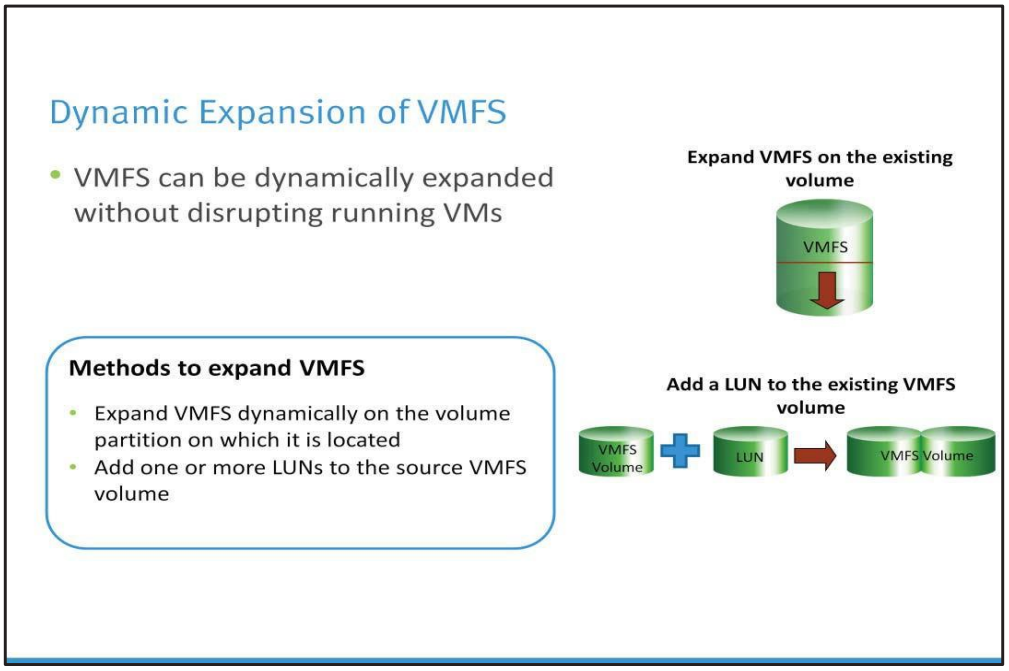
File System for Managing VM Files

- Hypervisor uses two file systems to manage the VM files
  - Hypervisor's native file system called Virtual Machine File System (VMFS)
  - Network File System (NFS) such as NAS file system

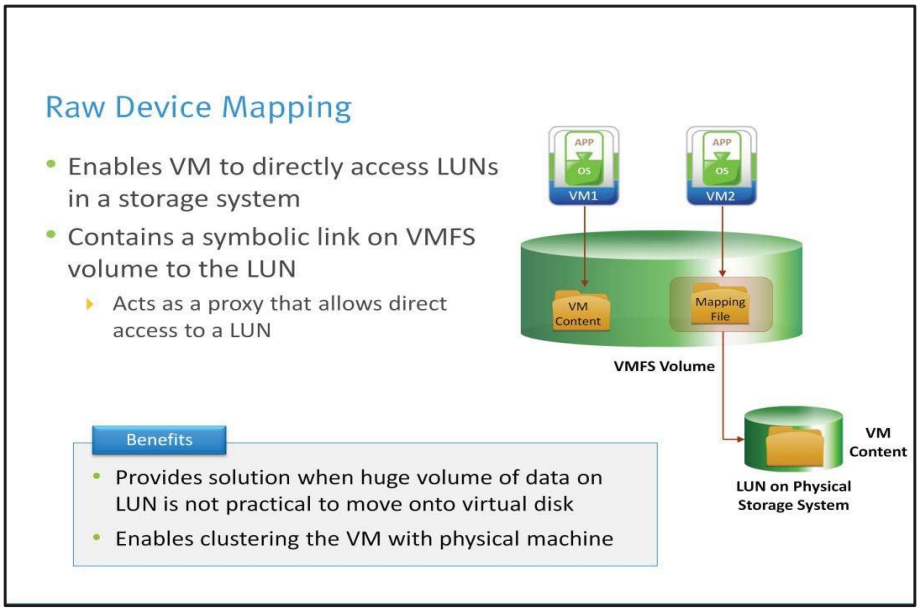
Virtual machine files could be managed by hypervisor's native file system, called Virtual Machine File System (VMFS), or Network File System (NFS) such as NAS file system.



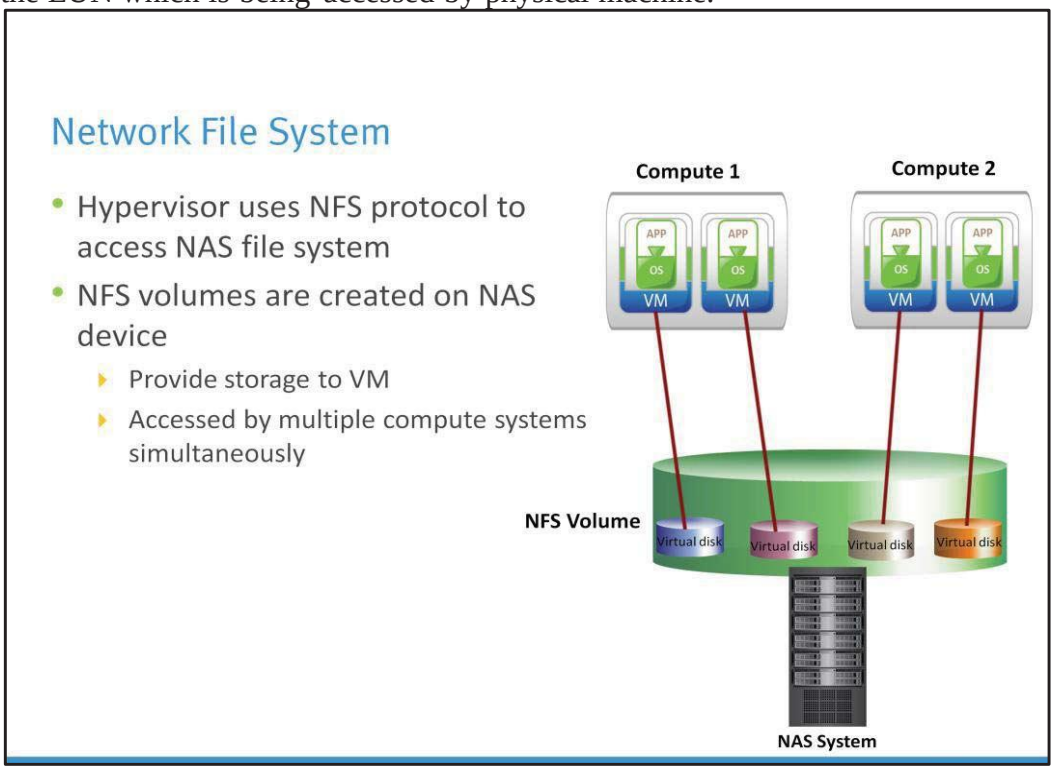
Virtual Machine File System (VMFS) is the native file system for hypervisor. It is a simple, efficient, cluster file system that allows multiple compute systems to read and write to the same storage simultaneously. It provides on-disk locking to ensure that the same virtual machine is not powered on by multiple compute systems at the same time. If a compute system fails, the on-disk lock for each virtual machine running on the failed compute system can be released so that virtual machines may be restarted on other compute system. VMFS volume is used for providing storage space for creating Virtual Machine File System to store virtual machine files.



Hypervisors support Logical Volume Manager (LVM) for extending VMFS dynamically without disrupting the running VMs across compute systems. There are two ways by which the VMFS can be dynamically expanded. The first method allows to expand VMFS dynamically on the volume partition on which it is located. The second method allows to expand the capacity of VMFS volume by adding one or more LUNs to the source VMFS volume to create a large one.

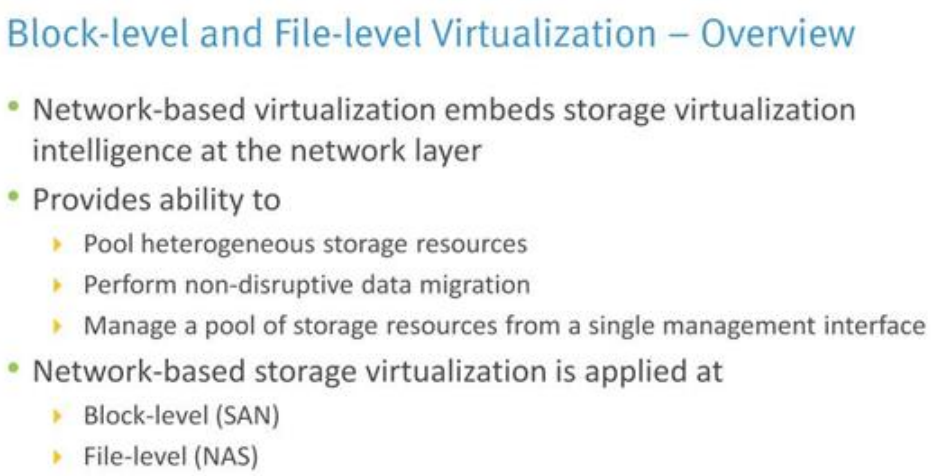


Virtual machines can store data directly on a LUN in the storage system instead of storing its data in a virtual disk file on a VMFS volume. Storing data in this way is useful when the applications running on the VMs are required to know the physical characteristics of the storage device. Raw Device Mapping provides a mechanism for a virtual machine to have direct access to a LUN on the physical storage subsystem (FC or iSCSI). It is a special file in a VMFS volume that acts as a proxy for the LUN in the storage system. RDM is recommended when there is large amount of data on the LUN in the storage system and it is not practical to move onto a virtual disk. It is also used when clustering virtual machine with physical machine. In this case, the virtual machine is required to access the LUN which is being accessed by physical machine.

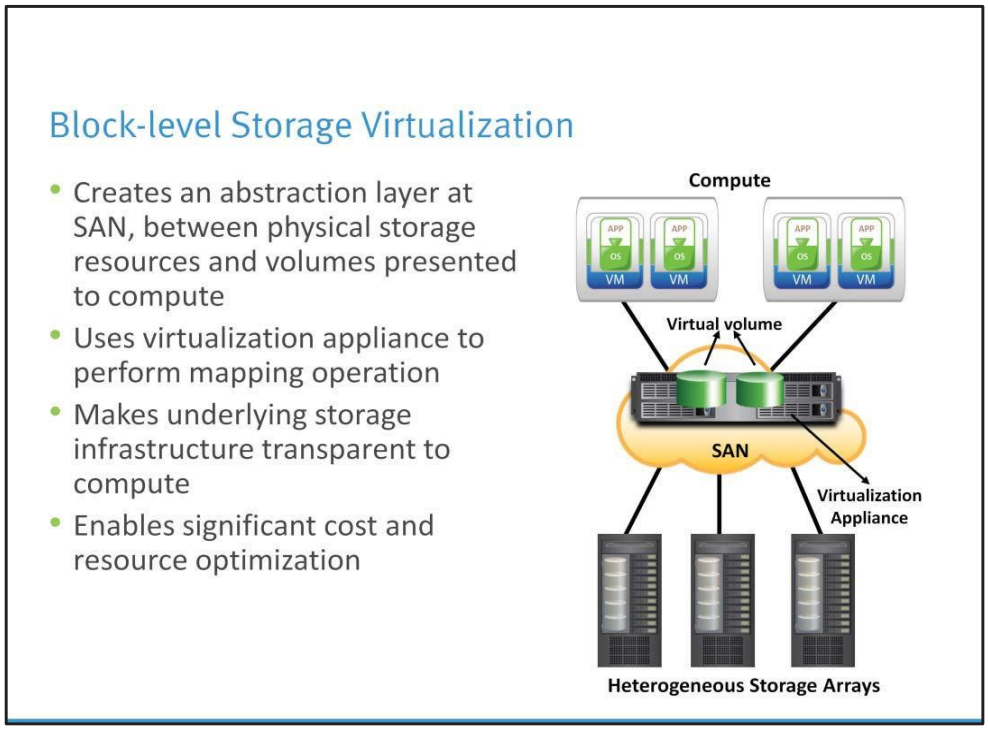


Hypervisor supports NAS system through the NFS protocol. The NFS protocol enables communication between NFS client and NFS server. Hypervisors come with NFS client software for NFS server (NAS) access. The NFS file system must be mounted on the compute system in order to allow NFS volumes to provide storage for all VMs. Provisioning of storage to virtual machines from NFS volume(s) is similar to provisioning from VMFS volume. NFS volume is managed entirely by the NAS system.

This lesson covers network-based block-level and file-level storage virtualization.

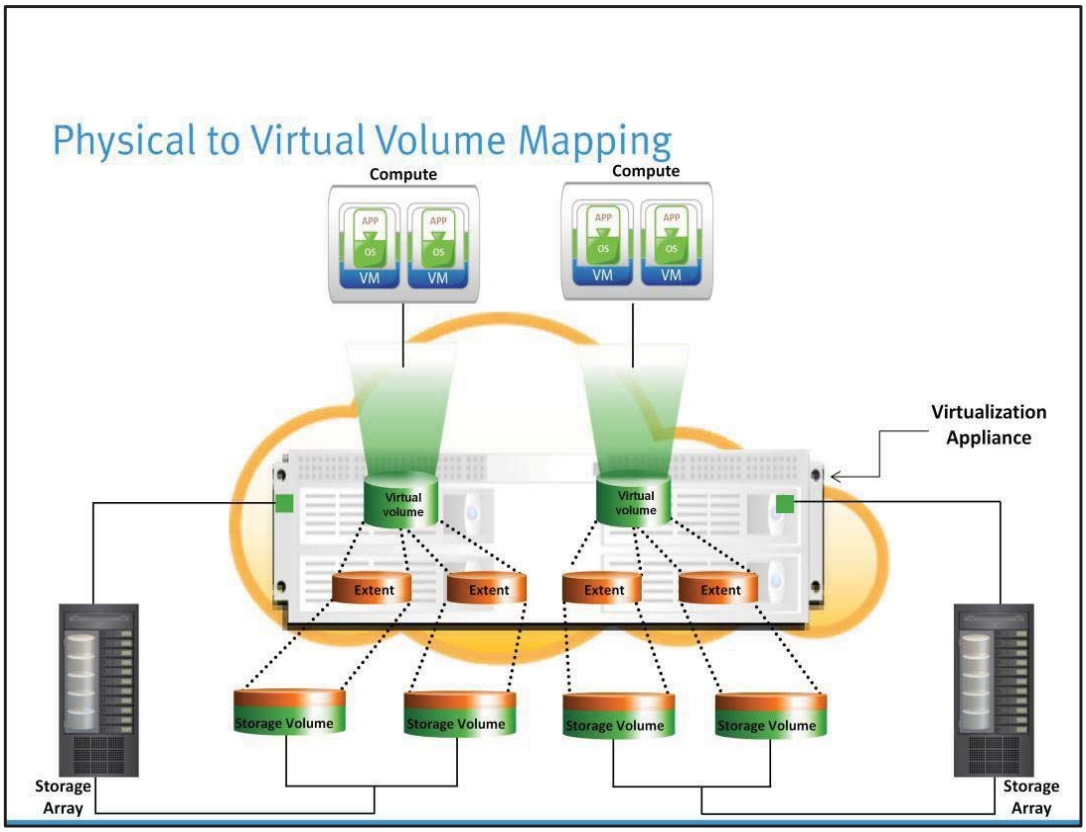


Network-based storage virtualization embeds the intelligence of virtualizing storage resources at network layer and provides an abstract view of physical storage resources. When an I/O is sent from the compute system, it is redirected through the virtualization layer at the network to the mapped physical storage. Virtualization applied at the network enables to pool multi-vendor storage resources and to manage these pools from a single management interface. It also enables non-disruptive data migration between arrays. Network-based storage virtualization can be implemented in both SAN and NAS environments. In a SAN, virtualization is applied at the block level, whereas in NAS, it is applied at the file level.



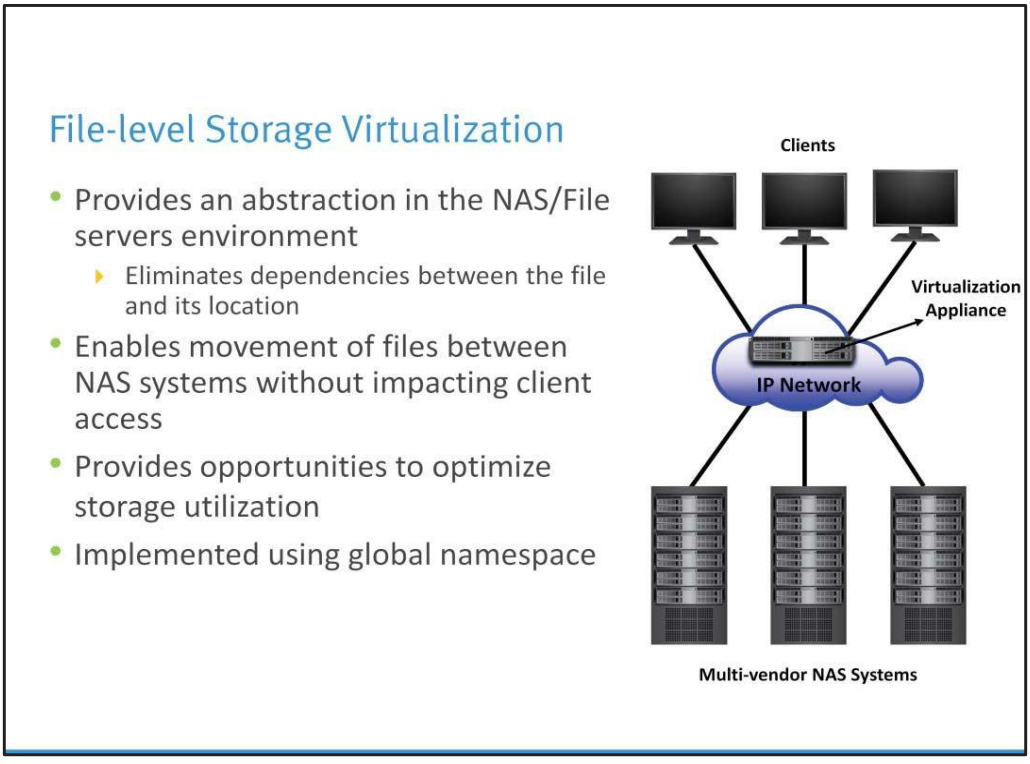
Block-level storage virtualization creates an abstraction layer in the SAN, between physical storage resources and virtual volumes presented to the compute systems. Instead of being directed to the LUNs on the individual storage arrays, the compute systems are directed to the virtual volumes on the virtualization appliance at the network. The virtualization appliance performs mapping between the virtual volumes and the LUNs on the arrays. Block- level storage virtualization enables us to combine several LUNs from one or more arrays into a single virtual volume before presenting it to the compute systems. It also takes a single large LUN from an array, slices it into smaller virtual volumes, and presents these volumes to the compute systems. Block-level storage virtualization supports dynamic increase of storage volumes, consolidation of heterogeneous storage arrays, and transparent volume access.

With block-level virtualization solution in place, the virtualization appliance at the network handles the migration of data. The virtualization appliance enables storage volumes to remain online and accessible while data is being migrated. After data is migrated from one array to another, no physical changes are required because the compute system still points to the same port on the virtualization appliance. However, the mapping on the virtualization appliance should be changed to point to the new location. These changes can be performed online with no impact to end the user data access. Deploying block-level storage virtualization in heterogeneous arrays environment facilitates an Information Lifecycle Management (ILM) strategy, enabling significant cost and resource optimization. Low-value data can be migrated from higher performance to appropriate performance arrays or disks.



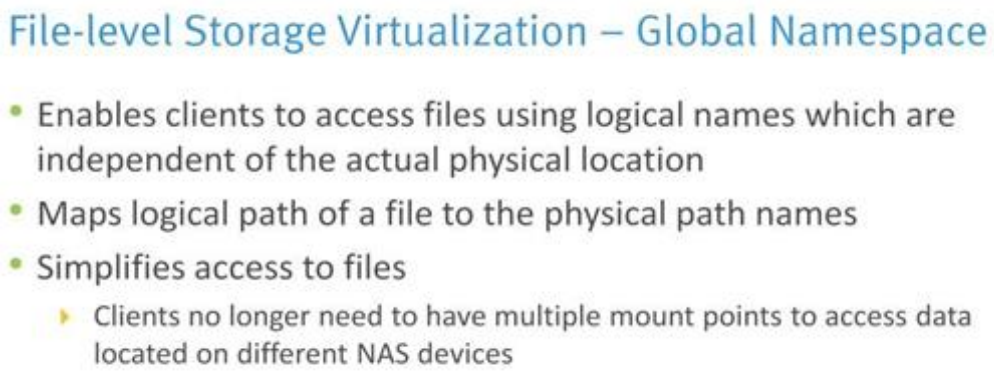
The virtualization appliance encapsulates physical storage devices and applies layers of logical abstraction to create virtual volumes. These virtual volumes are presented to the compute system.

Storage volume is a device or LUN on an attached storage system that is visible to the virtualization appliance. The available capacity on a storage volume is used to create extent and virtual volumes. Extents are mechanisms a virtualization appliance uses to divide storage volumes. Extents may be all or part of the underlying storage volume. The virtualization appliance aggregates these extents and applies RAID protection to create virtual volumes.



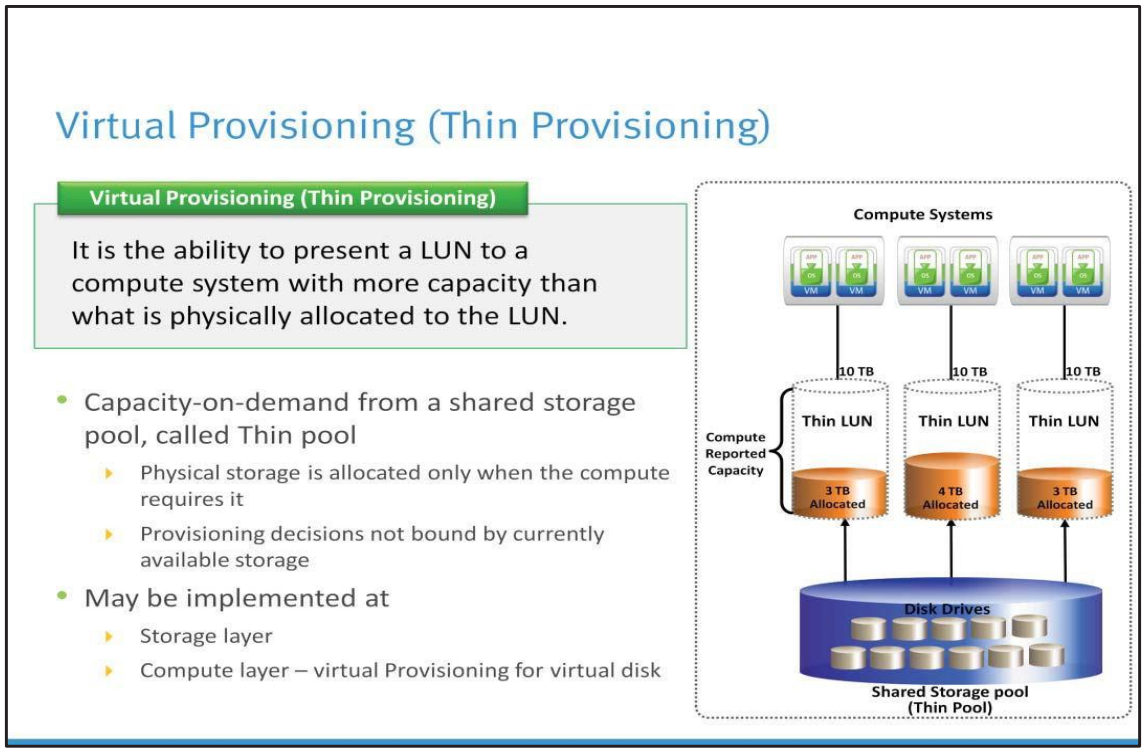
File-level storage virtualization provides an abstraction in the NAS environment and eliminates dependencies between the file and its physical location. Before file-level virtualization, each client knows the exact location of its file-level resources. In a data center, migrating data from a NAS to another NAS may be required for technology refresh, performance requirements, and non-availability of additional storage capacity. However, it is not easy to move files across this environment, and this requires downtime for NAS systems. Moreover, clients need to be reconfigured with the new path. This makes it difficult for storage administrators to improve storage efficiency, while maintaining the required service level.

File-level virtualization simplifies file mobility. File virtualization appliance at the network creates a logical pool of storage and enables users to use a logical path, rather than a physical path to access files. File virtualization facilitates the movement of files between the NAS systems without any downtime i.e., clients can access their files non-disruptively while the files being migrated. Global namespace is used to map the logical path of a file to the physical path names.



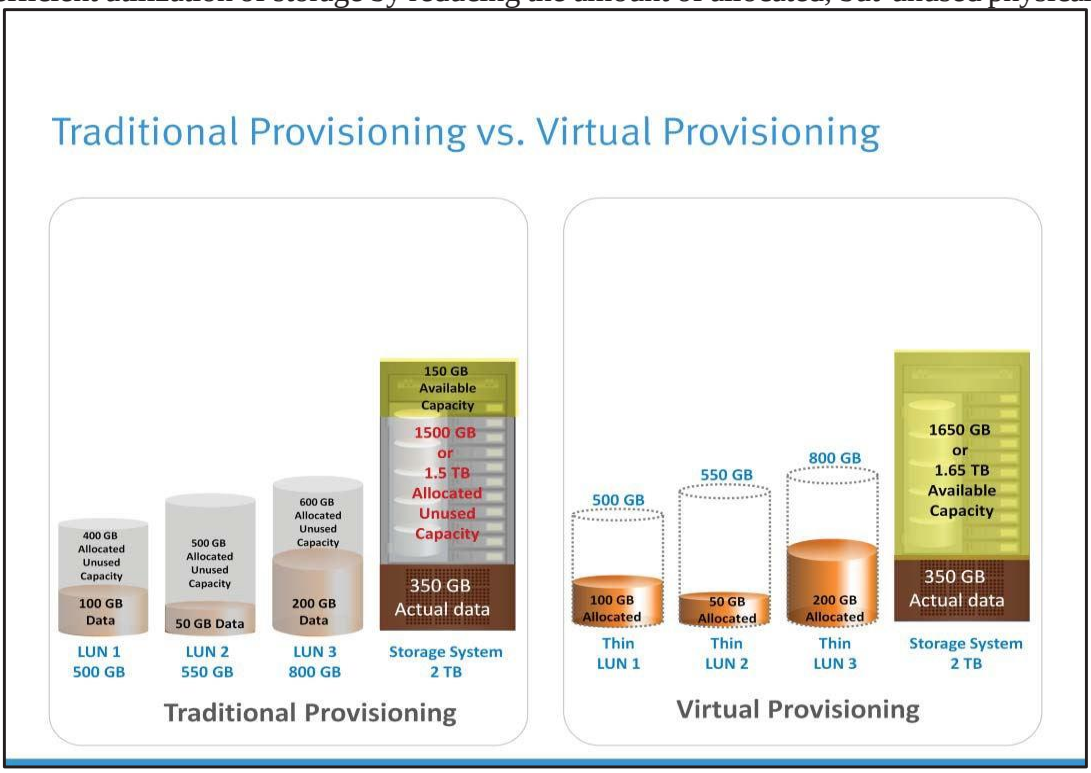
Namespace provides an abstraction layer, enabling clients to use a logical name that is independent of the actual physical location. Typically, with a standard file system such as NTFS, a namespace is associated with a single machine or file system. By bringing multiple file systems under a single namespace, global namespace provides a single view of the directories and files. It also provides administrators a single control point for managing files.

This lesson covers the concept of virtual provisioning and automated storage tiering.



One of the biggest challenges for storage administrators is balancing the storage space required by various applications in their data centers. Administrators typically allocate storage space based on anticipated storage growth. They do this to reduce the management overhead and application downtime required to add new storage later on. This generally results in the over-provisioning of storage capacity, which leads to higher costs, increased power, cooling, and floor space requirements, and lower capacity utilization. These challenges are addressed by Virtual Provisioning.

Virtual Provisioning is the ability to present a logical unit (Thin LUN) to a compute system, with more capacity than what is physically allocated to the LUN on the storage array. Physical storage is allocated to the application "on-demand" from a shared pool of physical capacity. This provides more efficient utilization of storage by reducing the amount of allocated, but unused physical storage.



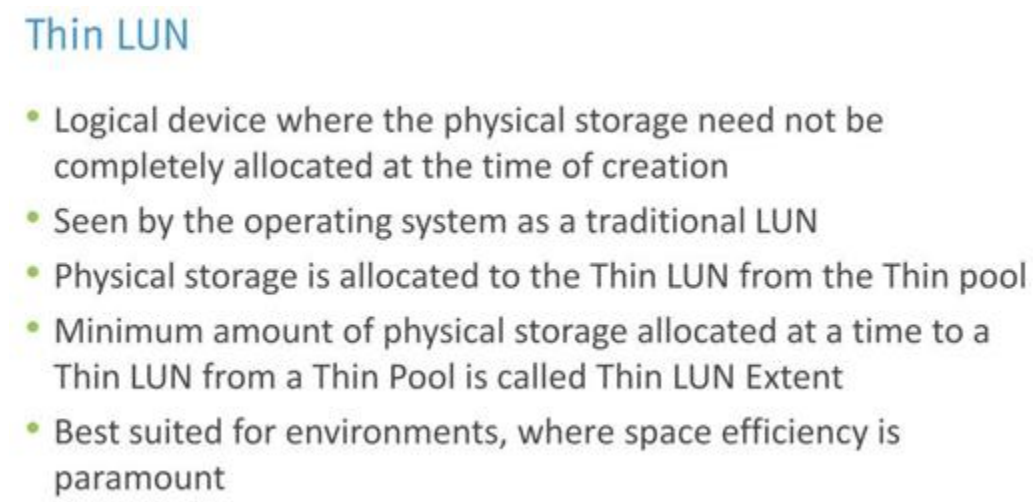
The example shown on this slide compares virtual provisioning with traditional storage provisioning. The example demonstrates the benefit of better capacity utilization.

Let us assume that three LUNs are created and presented to one or more compute systems using traditional provisioning methods. The total usable capacity of the storage system is 2 TB.

- The size of LUN 1 is 500 GB, of which 100 GB contains data and 400 GB is allocated, but unused.
- The size of LUN 2 is 550 GB, of which 50 GB contains data and 500 GB is allocated, but unused.
- The size of LUN 3 is 800 GB, of which 200 GB contains data and 600 GB is allocated, but unused.

In total, the storage system contains 350 GB of actual data, 1.5 TB of allocated, but unused capacity, and only 150 GB of capacity available for other applications. Now, let us assume that a new application is installed in the data center and requires 400 GB storage capacity. The storage system has only 150 GB of available capacity. So, it is not possible to provide 400 GB storage to the new application even though 1.5 TB of unused capacity is available. This shows the under utilization of storage in a traditional storage provisioning environment.

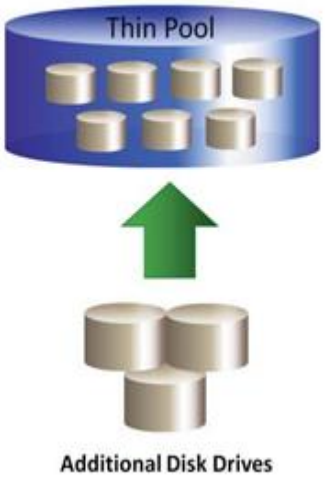
If we consider the same 2 TB storage system with Virtual Provisioning, the differences are quite dramatic. Although the system administrator creates the same size LUNs, there is no allocated unused capacity. In total, the storage system with Virtual Provisioning has 350 GB of actual data and 1.65 TB of capacity available for other applications, versus only 150 GB available in the traditional storage provisioning method.



Thin LUNs are logical devices. Physical storage need not be completely allocated to them at the time of creation. Physical storage is allocated to the Thin LUNs from the Thin pool. ‘Thin LUN extent’ is the minimum amount of physical storage that is consumed at a time by a ‘Thin LUN’ from a ‘Thin pool’. From the operating system’s perspective, Thin LUNs appear as traditional LUNs. Thin LUNs are best suited for situations where space efficiency is paramount. They are used for applications when the storage space consumption is difficult to predict.

Thin Pool

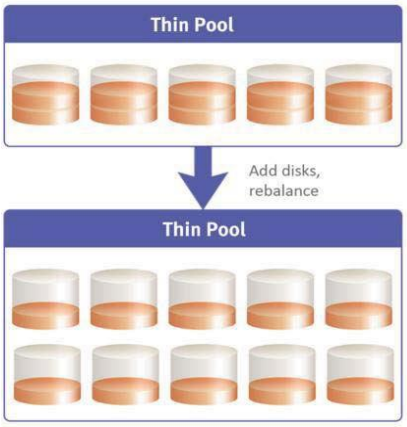
- Collection of physical drives that provide the actual physical storage used by Thin LUNs
- Multiple pools may be created within a storage array
- Can be expanded dynamically
  - Drives can be added to a Thin pool while pool is being used in production
- Allocated capacity is reclaimed by the pool when Thin LUNs are destroyed



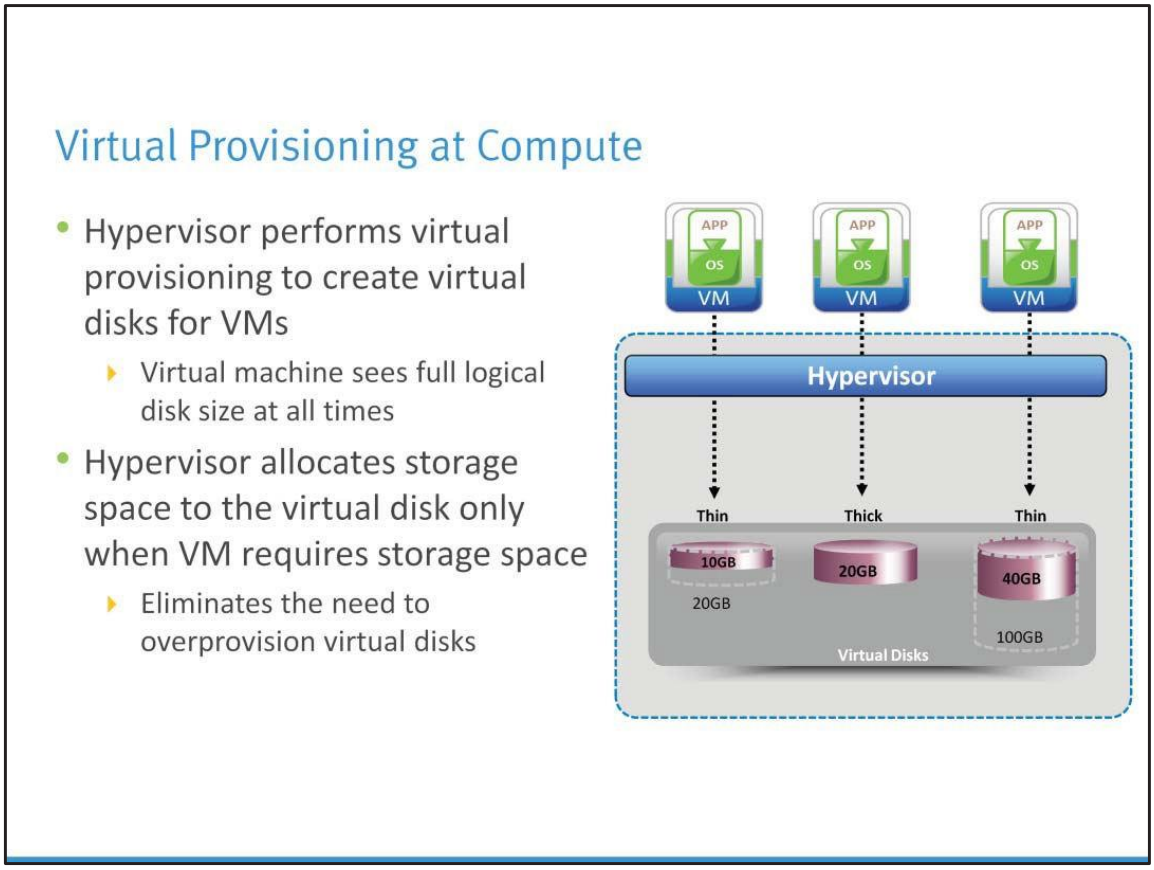
A Thin pool comprises physical drives that provide the actual physical storage used by Thin LUNs. A Thin pool is created by specifying a set of drives and a RAID type for that pool. Thin LUNs are then created out of that pool (similar to traditional LUN created on a RAID group). All the Thin LUNs created from a pool share the storage resources of that pool. Multiple pools can be created within a storage array. Adding drives to a Thin pool increases the available shared capacity for all the Thin LUNs in the pool. Drives can be added to a Thin pool while the pool is used in production. The allocated capacity is reclaimed by the pool when Thin LUNs are destroyed.

Thin Pool Rebalancing

- Balances the used capacity of physical disk drives over the entire pool when new disk drives are added
- Restripes data across all disk drives



Thin pool rebalancing is a technique that provides the ability to automatically rebalance allocated extents on physical disk drives over the entire pool when new disk drives are added to the pool. Thin pool rebalancing restripes data across all the disk drives(both existing and new disk drives) in the thin pool. This enables spreading out the data equally on all the physical disk drives within the Thin pool, ensuring that the used capacity of each disk drive is uniform across the pool.



Virtual provisioning may occur at the compute level. Hypervisor performs virtual provisioning to create virtual disks for VMs.

Hypervisor offers two options for provisioning storage to virtual disk:

- Provisioning thick disk
- Provisioning thin disk

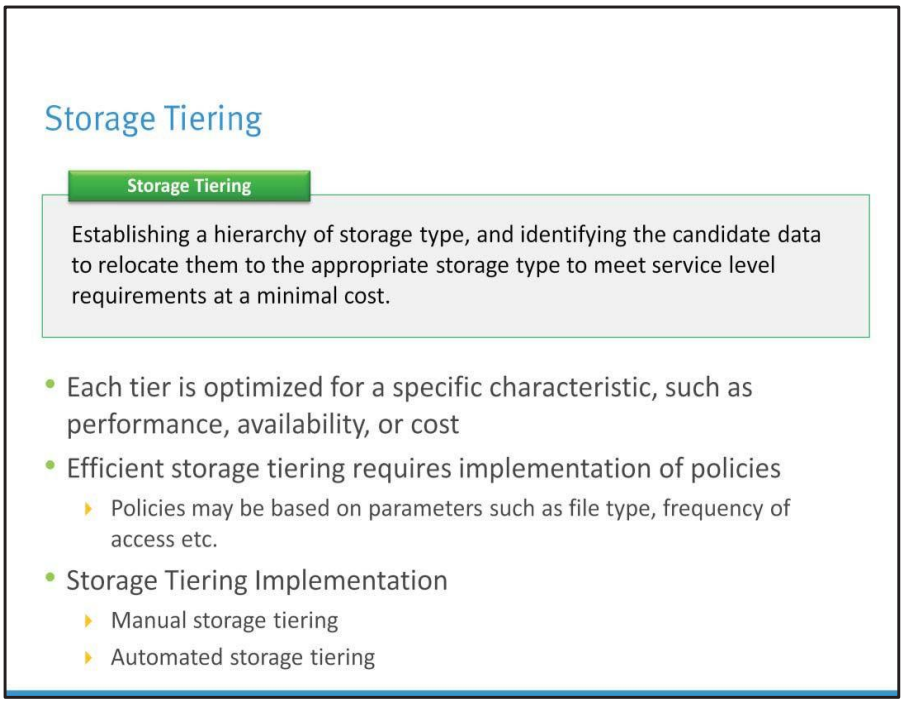
When Thick disk is provisioned, the entire provisioned space is committed to the virtual disk. Creating virtual disks in Thick format can lead to underutilization of virtual disks. In this case, large amounts of storage space, allocated to individual virtual machines, may remain unused.

When Thin disk is provisioned, the hypervisor allocates storage space to the virtual disk only when VM requires storage space. This eliminates the allocated, but unused storage capacity at the virtual disk.

Virtual Provisioning reduces administrative overhead by simplifying storage provisioning to the compute systems. Storage provisioning can be done independent of the physical storage capacity. Virtual provisioning can reduce the time required to repeatedly add storage capacity to the compute systems. It improves capacity utilization by reducing the amount of allocated, but unused physical

storage and also avoids over-allocation of storage to the compute systems. Virtual provisioning reduces storage and operating costs. Storage costs are reduced through increased space efficiency in primary storage because the storage is allocated as required. Virtual provisioning eliminates additional investment on high-end primary storage drives due to effective storage utilization. Operating costs are reduced considerably because fewer disks consume less power, cooling, and floor space. Virtual Provisioning is less disruptive to applications, and so, administrators do not have to continually take applications off-line to increase the storage capacity.

Drives in Thin pool should have the same RPM. If there is a mismatch, then the required performance may vary. All the drives in a pool must be of the same size because drives of different sizes may result in unutilized drive capacity. It is noted that all applications are not suited to Thin LUNs. Thin LUNs are most appropriate for applications that can tolerate some variation in performance. For applications demanding higher service levels, traditional LUNs on RAID groups would be a more suitable choice.

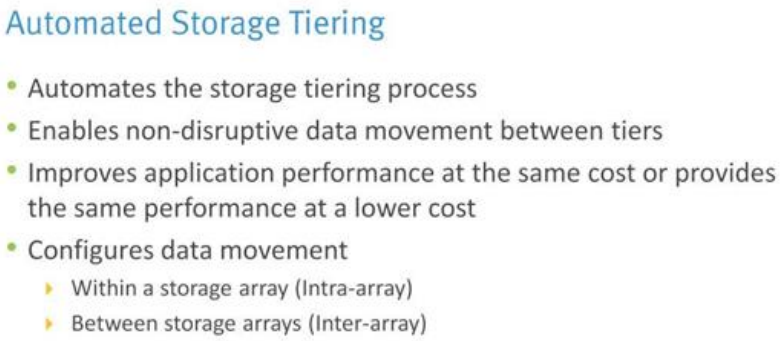


Organizations are experiencing tremendous data growth, which increases their storage requirements. They are also required to meet regulatory requirements. The cost of storing data and meeting SLA is a serious concern. Buying more high-end storage is not a cost-efficient solution for the growing data storage needs. Organizations require solutions that enable storing the right data, at the right cost, with the right access.

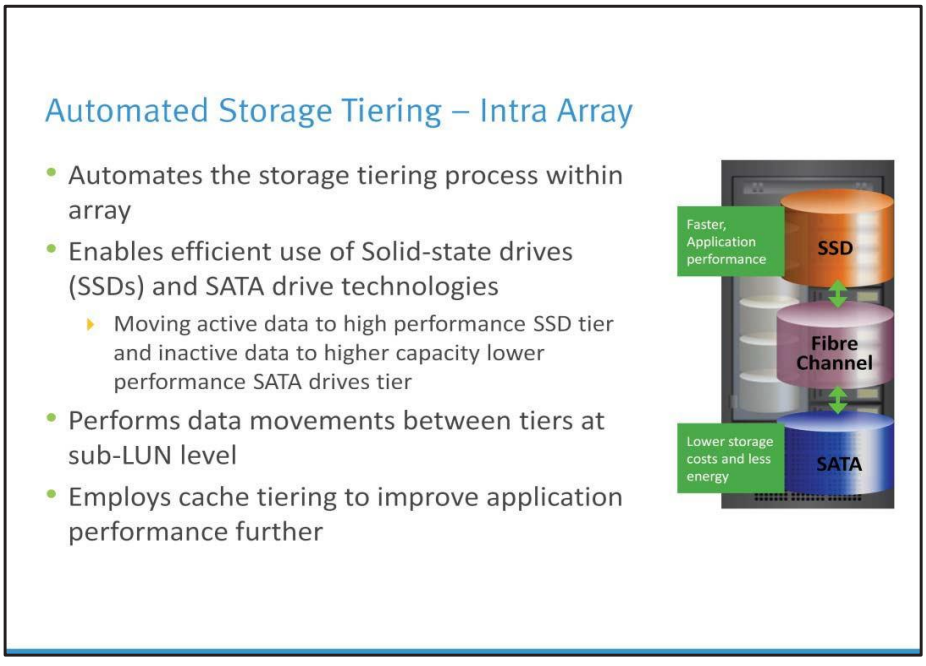
Storage tiering has emerged as a means to address these challenges. It is an approach to establish a hierarchy of storage types. It helps identify active or inactive data to relocate them to an appropriate storage type. This enables in meeting service level requirements at an optimal cost. Each tier has different levels of protection, performance, data access frequency, and other considerations. For example, high performance FC drives may be configured as tier 1 storage to keep frequently accessed

data to improve performance and low cost SATA drives as tier 2 storage to keep the less frequently accessed data. Moving the active data (frequently used data) to Solid-state drive (SSD) or FC improves the application performance. Moving the inactive data (less frequently used) to SATA can free up storage capacity in high performance drives and reduce the cost of storage. This movement of data happens based on defined tiering policies. The tiering policy may be based on parameters such as file type, frequency of access, performance, etc. For example, if a policy states "move the files which are not accessed for last 30 days to lower tier", then the files matching this condition are moved to the lower tier.

There are two types of Storage tiering: manual storage tiering and automated storage tiering. The manual storage tiering is the traditional method where the storage administrator has to monitor the storage workloads periodically and move the data between the tiers. A traditional storage tiering process is manual, repetitive, and takes few hours to few days to complete.

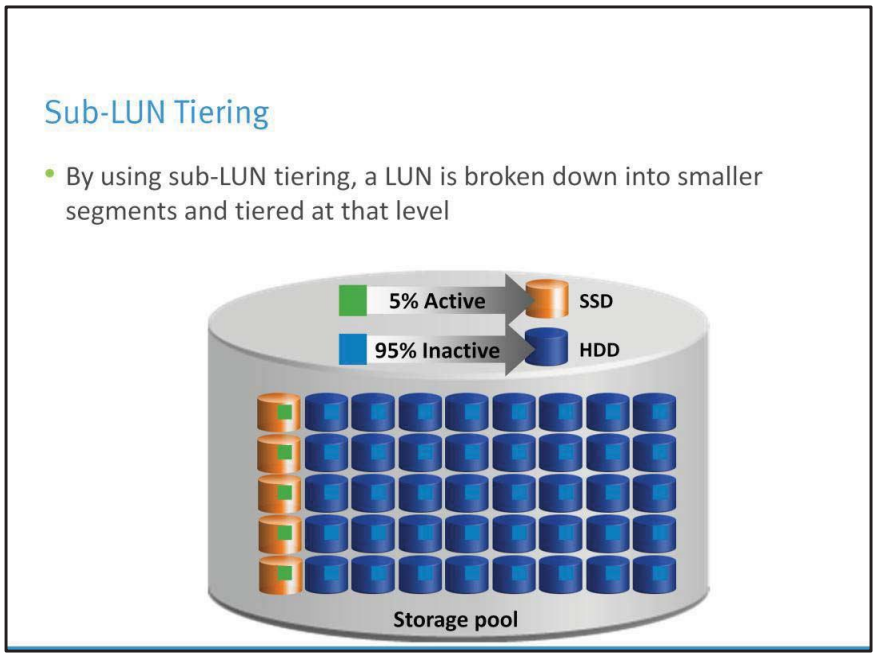


Automated storage tiering automates the storage tiering process. Data movement between tiers is performed non-disruptively, without affecting business continuity. Automated storage tiering eliminates manual tiering when the application workload characteristic changes over time. It improves application performance at the same cost or provides the same application performance at a lower cost. Data movements between tiers can happen within or between storage arrays.

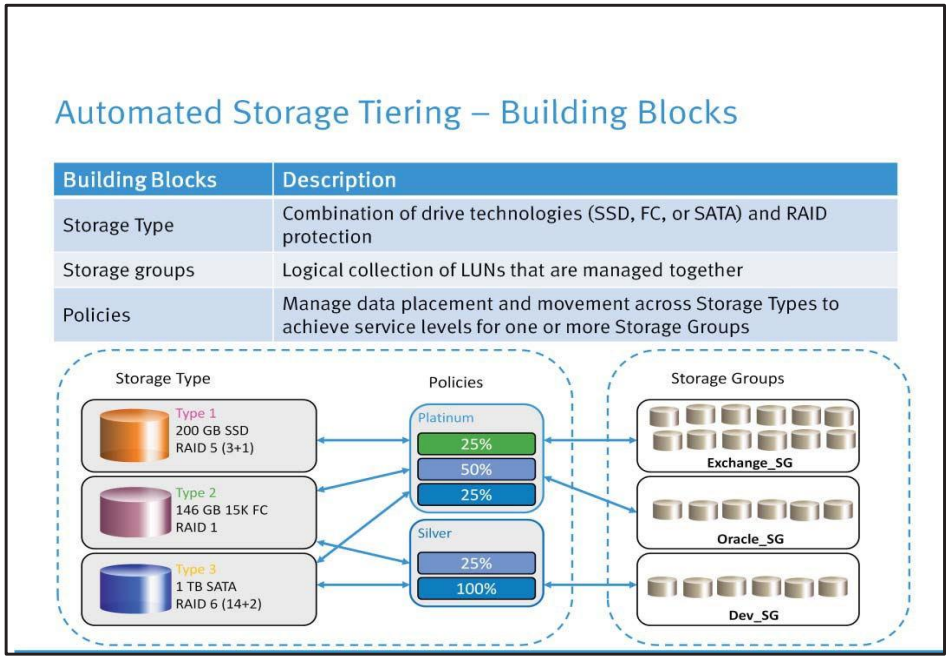


Automated storage tiering (Intra-array) automates the process of storage tiering within a storage array. It enables efficient use of SSDs and SATA drive technologies and provides performance and cost optimization. Automated storage tiering proactively monitors application workload and automatically moves the active data to higher performing SSDs tier, and inactive data to higher

capacity, lower performance SATA drives tier. The goal is to keep the SSDs busy by storing the most frequently accessed data on them, while moving out less frequently accessed data to SATA drives. Data movements executed between tiers can be performed at the sub-LUN level. The performance can be further improved by implementing tiered cache.



Traditional storage tiering moves an entire LUN from one tier of storage to another. This movement includes both active and inactive data in that LUN. This method does not give effective cost/performance benefits. In sub-LUN tiering, a LUN is broken down into smaller segments and tiered at that level. Movement of data with much finer granularity (ex., 8MB) greatly enhances the value proposition of automated storage tiering. Tiering at the sub-LUN level moves more active data to faster drives and less active data to slower drives effectively.

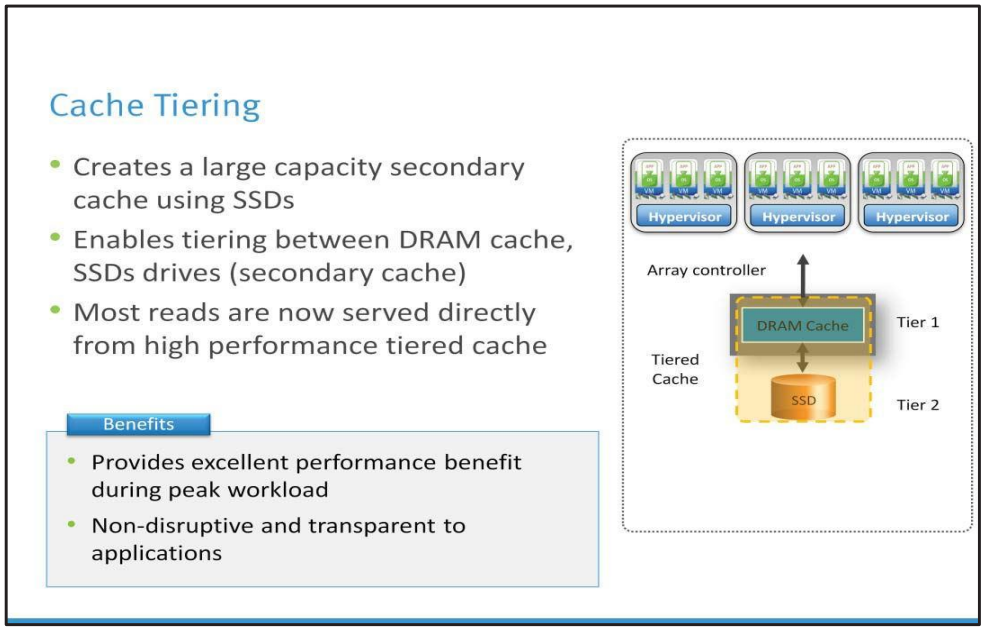


There are three major building blocks of automated storage tiering:

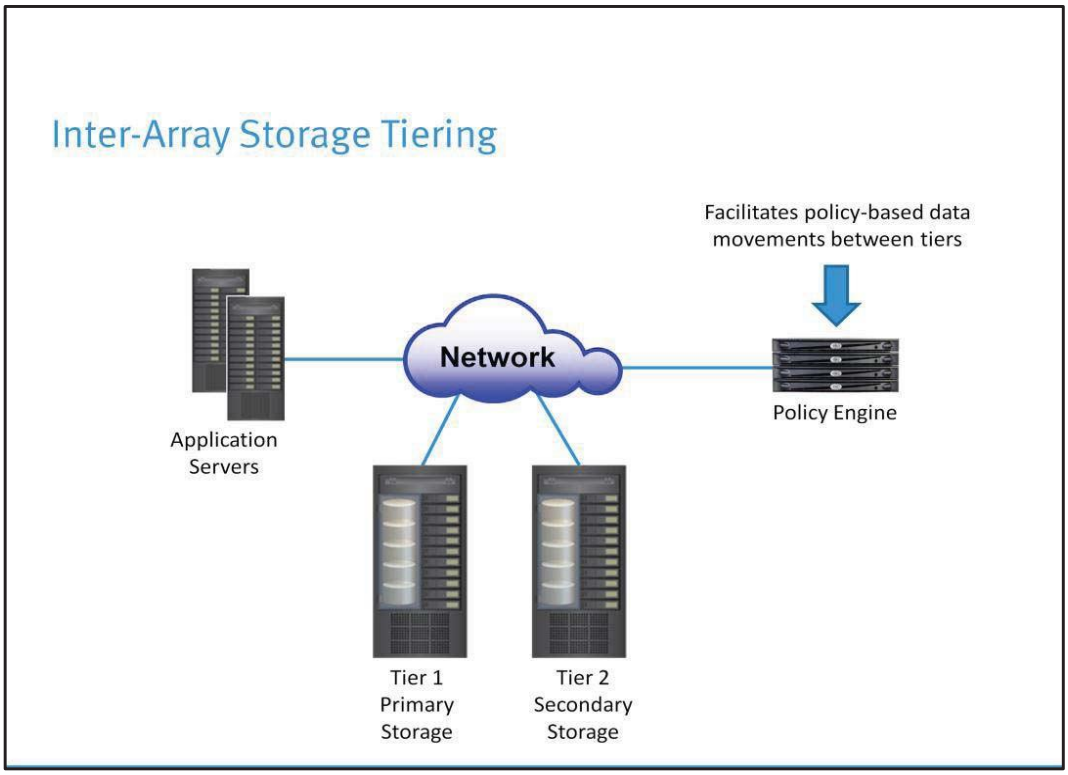
Storage Type is a combination of a drive technology (SSD, FC, or SATA) and a RAID protection type. This slide displays a simple example with three storage types – one for SSD, one for FC, and one for SATA.

Storage Groups are a logical collection of LUNs that are to be managed together. In this example, three storage groups represent three different business applications. Typically, each application will have its own service level requirements.

Policies manage data movement across storage types. The data movement is done by specifying a set of tier usage rules. A policy may be applied to one or more previously defined storage groups. In our example, the Platinum policy implies that up to 25% of the Storage Group's capacity is allowed to reside on SSD, up to 50% on Fibre Channel, and up to 25% on SATA. Note that the percentages for tier usage within a given policy need not add to 100%. For example, the "Silver" policy implies that all of the Storage Group's capacity (i.e. up to 100% of its capacity) may reside on SATA. Only the Fibre Channel capacity is restricted to 25% of the total capacity allocated to this Storage Group. This gives the flexibility to move the volumes included in these Storage Groups entirely to SATA, if that can produce optimal results.



A large cache in a storage array improves performance by retaining frequently accessed data for long period of time. However, configuring large cache in the storage array involves more cost. An alternate way to increase the size of the cache is by utilizing the existing SSDs on the storage array to extend the size of the cache configured in the array. Cache tiering uses SSDs to create a large capacity secondary cache. It enables tiering between DRAM (primary cache) and SSDs (secondary cache). It also enables to store large volumes of frequently accessed data on the cache tier. So, most reads are now served directly from cache tiering, and provides excellent performance benefit during peak workload.



Inter-array storage tiering automates the identification of active or inactive data to relocate them to different performance/capacity tiers between the arrays. This slide shows an example of a two-tiered storage environment. This environment optimizes primary storage for performance and secondary storage for capacity and cost. The policy engine facilitates moving inactive or infrequently accessed data from the primary to secondary storage. Some of the prevalent reasons to tier data across arrays is for archival or compliance requirements. As an example, the policy engine may be configured to locate all files in the primary storage that have not been accessed in one month, and archive those files to the secondary storage. For each file it archives, the policy engine leaves behind a small space-saving stub file that points to the real data on the secondary storage. When a user tries to access the file at its original location on the primary storage, the user is transparently provided with the actual file that the stub points to, from the secondary storage.

UNIT-4

Virtualized Data Center (VDC) – Networking and desktop applications

Upon completion of this module, you should be able to:

- Describe network virtualization in VDC
- Describe VDC network infrastructure and components
- Describe Virtual LAN (VLAN) and Virtual SAN (VSAN) and their benefits
- Describe the key network traffic management techniques in VDC

This module focuses on networking in VDC environment. It covers network virtualization in VDC, VDC network infrastructure and components, virtual LAN, and virtual SAN. It also describes the key network traffic management techniques.

Lesson 1: VDC Networking Overview

Topics covered in this lesson:

- Overview of network virtualization
- Overview of network that is virtualized
- Virtualization tools that enable network virtualization
- Benefits of network virtualization

This lesson covers the overview and benefits of network virtualization. It also includes an overview of network that is virtualized in VDC and the virtualization tools that enable network virtualization.

Network Virtualization

Network Virtualization

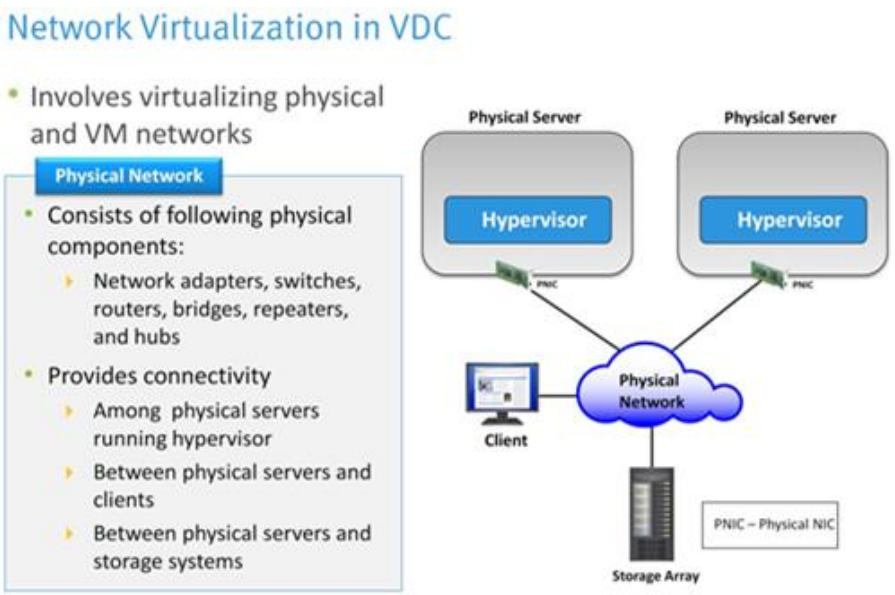
It is a process of logically segmenting or grouping physical network(s) and making them operate as single or multiple independent network(s) called "Virtual Network(s)".

- Enables virtual networks to share network resources
- Allows communication between nodes in a virtual network without routing of frames
- Enforces routing for communication between virtual networks
- Restricts management traffic, including 'Network Broadcast', from propagating to other virtual network
- Enables functional grouping of nodes in a virtual network

Network virtualization involves logically segmenting or grouping physical network(s) into discrete logical entities called "virtual network(s)", and making them behave as single or multiple independent network(s). Network virtualization allows multiple virtual networks to share network resources without leaking information among them.

A virtual network appears as a physical network to the nodes connected to it. Two nodes connected to a virtual network can communicate among themselves without routing of frames, even if they are in different physical networks. Network traffic must be routed when two nodes in different virtual networks are communicating, even if they are connected to the same physical network. Network management traffic, including 'network broadcast' within a virtual network, does not propagate to any other nodes that belong to a different virtual network. This enables functional grouping of nodes with a common set of requirements in a virtual network, regardless

of the geographic location of the nodes.

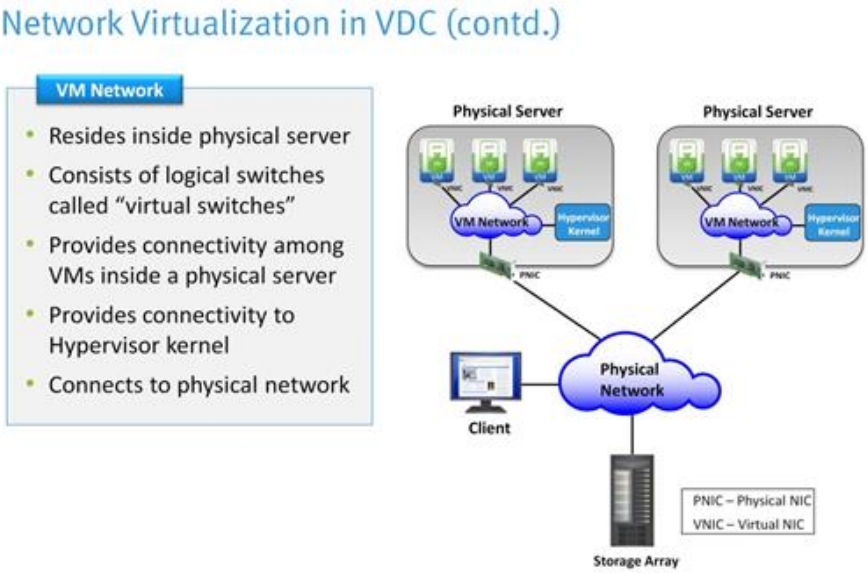


In VDC, network virtualization involves virtualization of both physical and 'Virtual Machine (VM)' networks.

The physical network may consist of network adapters, switches, routers, bridges, repeaters, and hubs.

The physical network provides connectivity:

- Among physical servers running hypervisor
- Between physical servers and clients
- Between physical servers and storage systems



A VM network resides inside a physical server. It includes logical switches, called 'virtual switches', which function similar to physical switches. The VM network enables communication among VMs within a physical server. For example, a VM which is running a business application may need to filter its traffic via a firewall server which could be another VM within the same physical server. It is beneficial to connect these VMs internally through the VM network. Connecting them through a physical network will add more delay to the VM traffic because it travels over the external physical network

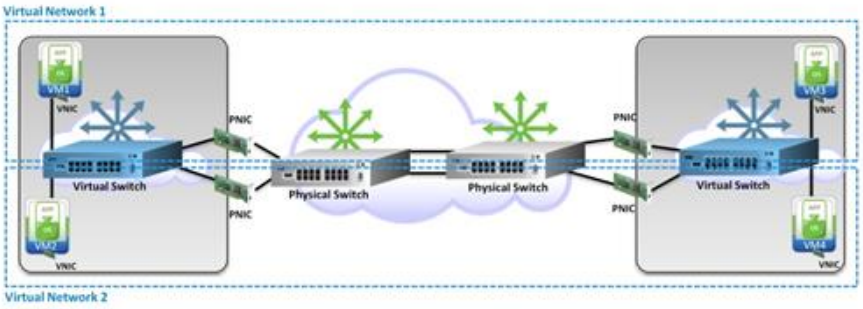
Hypervisor kernels are connected to the VM network. Hypervisor kernels communicate with the management server and storage systems using the VM network. The management server could

be a VM hosted in a physical server.

For communication between two VMs residing in different physical servers and between a VM and its clients, the VM traffic must travel through both the VM and physical networks. Hypervisor traffic is also required to transfer between the VM and physical networks. Hence, the VM network must be connected to the physical network.

Network Virtualization in VDC (contd.)

- VM and physical networks are virtualized to create virtual networks; for example: virtual LAN, virtual SAN



Network virtualization allows an administrator to create multiple virtual networks in VDC. These virtual networks may span across both VM and physical networks and share physical and virtual switches. A virtual network provides grouping of all the nodes that belong to the same functional unit in an organization. Virtual LAN and virtual SAN are examples of virtual networks.

In the figure shown on this slide, two virtual networks are created on both virtual and physical switches. Virtual network 1 provides connectivity to VM1 and VM3 and enables communication between them without routing of frames. Similarly, VM2 and VM4 belong to virtual network 2 and are allowed to communicate without routing.

Network Virtualization Tools

- Physical switch Operating System (OS)
  - OS must have network virtualization functionality
- Hypervisor
  - Uses built-in networking and network virtualization functionalities
    - To create virtual switch and configuring virtual networks on it
  - Or, uses third-party software for providing networking and network virtualization functionalities
    - Third-party software is installed onto the hypervisor
    - Third-party software replaces the native networking functionality of the hypervisor

Network virtualization is performed by hypervisor and physical switch Operating System (OS). These types of software allow an administrator to create virtual networks on physical and VM networks.

A physical switch runs an Operating System which performs network traffic switching. The Operating System must have network virtualization functionality to create virtual networks on the switch. Hypervisor has built-in networking and network virtualization functionalities. These functionalities can be leveraged to create a virtual switch and configure virtual networks on it. These functionalities are also provided by third-party software, which may be installed onto the hypervisor. Then, the third-party software module replaces the native networking functionality of the hypervisor.

Benefits of Network Virtualization

Benefit	Description
Enhances security	• Restricts access to nodes in a virtual network from another virtual network • Isolates sensitive data from one virtual network to another
Enhances performance	• Restricts network broadcast and improves virtual network performance
Improves manageability	• Allows configuring virtual networks from a centralized management workstation using management software • Eases grouping and regrouping of nodes
Improves utilization and reduces CAPEX	• Enables multiple virtual networks to share the same physical network, which improves utilization of network resource • Reduces the requirement to setup separate physical networks for different node groups

Network virtualization provides enhanced security by restricting access to nodes located within a virtual network from another virtual network. Therefore, sensitive data of one virtual network is isolated from other virtual networks.

Network broadcasts within a virtual network are not allowed to propagate to other virtual networks. Restricting broadcast preserves network bandwidth, which consequently improves virtual network performance for usual network traffic.

Virtual network allows grouping of nodes based on an organization's requirement. When new requirements come, an administrator changes the virtual network configuration using a management software and regroups nodes. The management software provides an interface to configure virtual networks from a centralized management workstation. The interface enables an administrator to send configuration commands to physical switch OS and/or hypervisor. As regrouping of nodes does not require re-cabling or physical movement of equipments, network management becomes easy.

Network virtualization allows multiple virtual networks to share the same physical network. This improves utilization of network resources. Network virtualization also cuts down the capital expenditure (CAPEX) in procuring network equipments for different node groups.

Module 5: Virtualized Data Center – Networking

Lesson 2: VDC Network Infrastructure

Topics covered in this lesson:

- Network infrastructure and components
- Network connectivity and traffic flow
- Features and functions of network components

This lesson covers features, functions, and connectivity of VDC network components.

Components of VDC Network Infrastructure

- VDC network infrastructure includes both virtual and physical network components
  - Components are connected to each other to enable network traffic flow

Component	Description
Virtual NIC	• Connects VMs to the VM network • Sends/receives VM traffic to/from VM network
Virtual HBA	• Enables a VM to access FC RDM disk/LUN assigned to the VM
Virtual switch	• Is an Ethernet switch that forms VM network • Provides connection to virtual NICs and forwards VM traffic • Provides connection to hypervisor kernel and directs hypervisor traffic: management, storage, VM migration
Physical adapter: NIC, HBA, CNA	• Connects physical servers to physical network • Forwards VM and hypervisor traffic to/from physical network
Physical switch, router	• Forms physical network that supports Ethernet/FC/iSCSI/FCoE • Provides connections among physical servers, between physical servers and storage systems, and between physical servers and clients

Similar to Classic Data Center (CDC) networking, there are basic building blocks to perform networking in a VDC. A VDC network infrastructure consists of both physical and virtual components, as listed on this slide. These components are connected to each other to enable network traffic flow.

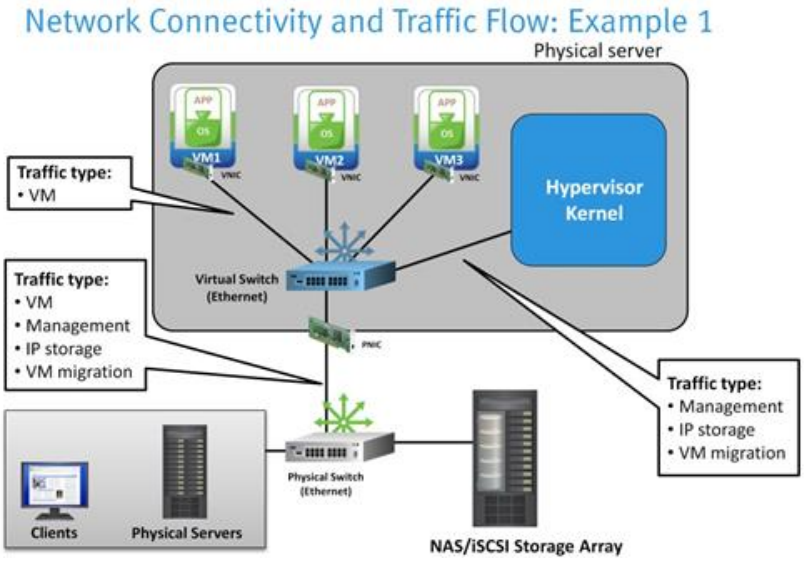
Network components such as virtual NIC, virtual HBA, and virtual switch are created inside a physical server using hypervisor. Virtual NICs enable VMs to connect to VM network. They send and receive VM traffic to and from the VM network. Virtual HBA enables a VM to access FC RDM disk/LUN assigned to the VM.

Virtual switches form VM networks and support the Ethernet protocol. They provide connection to the virtual NICs and forward the VM traffic. They also direct management, storage, and VM migration traffic to/from hypervisor kernel.

Physical adapters, such as Network Interface Card (NIC), Host Bus Adapter (HBA), and Converged Network Adapter (CNA), allow physical servers to connect to physical network. They forward VM and hypervisor traffic to/from a physical network.

A physical network includes physical switches and routers. Physical switches and routers provide connections among physical servers, between physical servers and storage systems, and between physical servers and clients. Depending on the network technology and protocol supported, these switches direct Ethernet, FC, iSCSI, or FCoE traffics.

<u>Note:</u> VM migration enables moving a VM from one physical server to another, and is controlled from management server. If a resource crunch occurs at a physical server, VMs are migrated to another physical server, which has sufficient resource to run these VMs.



The connectivity among VDC network components varies based on the type of protocol and the physical adapter used to enable physical server access to the storage system.

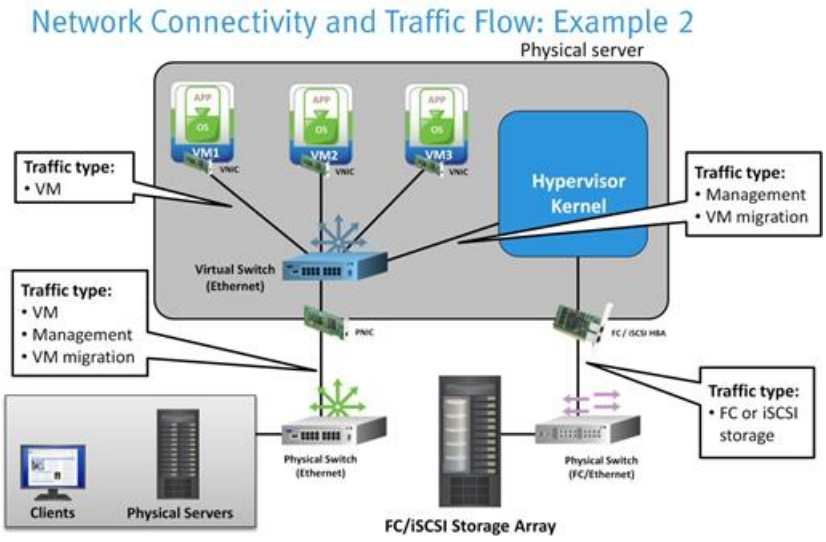
In this example, physical servers are connected to the IP storage system, such as a NAS or an iSCSI storage array. A physical Ethernet switch is used to connect physical servers and the storage system.

Each physical server hosts multiple VMs that are connected to a virtual switch. Each VM has at least one virtual NIC which transfers/receives VM I/Os in the form of Ethernet frames.

These Ethernet frames travel through virtual and/or physical switches before reaching their destination (Clients and any other VMs residing in other physical servers.)

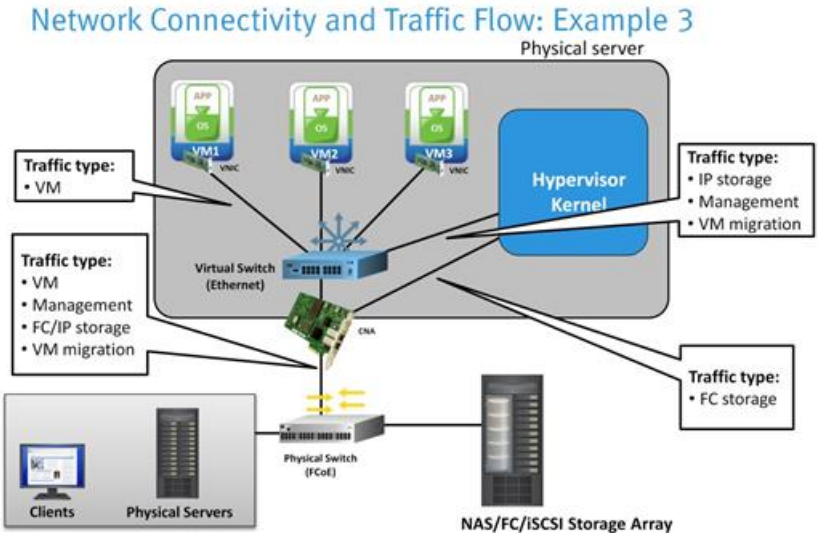
Hypervisor kernel is also connected to the virtual switch. Hypervisor kernel leverages the virtual and physical switches to send the IP storage, management, and VM migration traffic.

A physical server has one or more physical NICs (one NIC in this example). NICs provide a link between the virtual and physical switches and forwards VM and hypervisor kernel traffic between the switches.



The connectivity shown on this slide is similar to the previous example. However, in this case, the hypervisor kernel uses an FC or iSCSI HBA to access FC or iSCSI storage array. A hypervisor kernel is directly connected to the HBA. The HBA sends or receives storage traffic via an FC or Ethernet switch (This could be the same physical Ethernet switch that is connected to the virtual switch).

- A hypervisor kernel still uses the virtual switch to send/receive management and VM migration traffic.



In this scenario, a physical server uses a CNA card instead of separate HBA and NIC. The CNA card provides the connection between a virtual switch and a physical FCoE switch. CNA has the capability to converge both FC and Ethernet traffic over an Ethernet connection. This allows hypervisor kernel to access FC and IP storage system using a single network adapter.

Hypervisor kernel recognizes CNA as an NIC and an FC HBA. To access FC storage, hypervisor kernel directly sends storage traffic to CNA. To access the IP storage or to forward management and VM migration traffic, the hypervisor kernel sends traffic through the virtual switch.

Virtual Network Component: Virtual NIC

- Connects VMs to virtual switch
- Forwards Ethernet frames to virtual switch
- Has unique MAC and IP addresses
- Supports Ethernet standards similar to physical NIC

A VM can have one or more virtual NICs. The working of a virtual NIC is similar to a physical NIC, even though the virtual NIC is used for connecting VMs and virtual switches. The guest Operating System sends network I/Os to the virtual NIC through a device driver similar to a physical NIC. Virtual NIC forwards I/Os in the form of Ethernet frames to a virtual switch for further transmission to destination. Each virtual NIC has unique MAC and IP addresses and responds to the standard Ethernet protocol exactly as how a physical NIC would. Hypervisor generates these MAC addresses and allocates a MAC address to a virtual NIC at the time of VM creation.

Virtual Network Component: Virtual HBA

- Enables a VM to access FC RDM disk/LUN assigned to the VM
- Configured using N_Port ID Virtualization (NPIV) technology
  - Single physical FC HBA or CNA port (N_port) to function as multiple virtual N_ports, each with its own WWN
  - A virtual N_port acts as a virtual HBA port
- Hypervisor kernel leverages NPIV to instantiate virtual N_ports
  - Assigns the virtual N_ports to the VMs
- Enables zoning and LUN masking at VM level

Virtual HBA enables a VM to access a FC RDM disk/LUN assigned to the VM. Virtual HBAs are configured using N_Port ID Virtualization (NPIV) technology, which is based on an ANSI T11 standard. NPIV enables a single physical FC HBA or CNA port (N port) to function as multiple virtual N_ports with each virtual N_port having a unique WWN identity in the FC SAN. This allows a single physical FC HBA port to register several unique WWNs with the fabric and obtain multiple fabric addresses. A virtual N_port acts as a virtual HBA port that enables a VM to directly access an LUN assigned to it. Hypervisor kernel leverages NPIV to instantiate virtual N_ports on the physical HBA or CNA and then assigns the virtual N_ports to the VMs.

A VM with a virtual HBA is recognized as a node in the fabric. This allows an administrator to restrict access to specific LUNs to specific VMs using zoning and LUN masking, in the same way that they can be restricted to specific physical servers. In the absence of virtual HBA, VMs share the WWN identity of a physical HBA or CNA to access RDM disks. There is no option to uniquely secure and manage storage for an individual VM. For VM with a virtual HBA, the virtual HBA WWN can be included as a zone member or storage group member in LUN masking. This allows access control and storage management per VM basis.

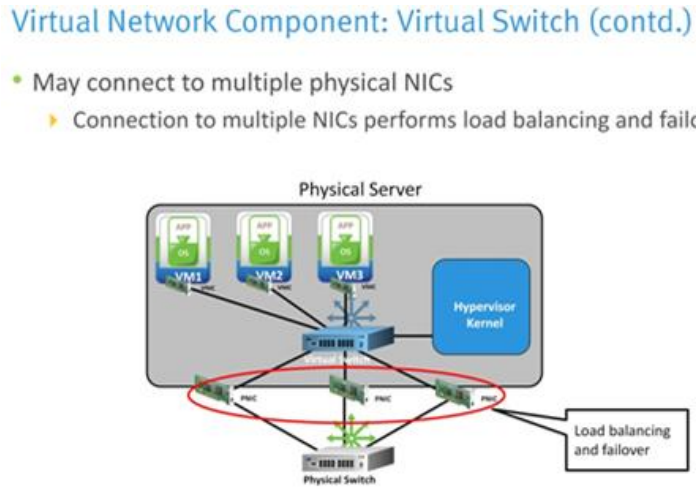
Virtual Network Component: Virtual Switch

- Is a logical OSI layer 2 switch that supports Ethernet protocol
- Resides inside a physical server
- Is created and configured using hypervisor
- Maintains MAC address table for frame forwarding
- Directs network traffic to/from VMs and hypervisor kernel
  - VM to VM within physical server
  - VM to physical network
  - Hypervisor kernel: IP storage, VM migration, and management

A virtual switch is a logical layer 2 (OSI model) Ethernet switch that resides inside a physical server that uses a hypervisor. Virtual switches are created and configured using hypervisor. Virtual switches provide traffic management for VMs and hypervisor kernel. Each virtual switch maintains a MAC address table, which includes a list of MAC addresses and corresponding virtual

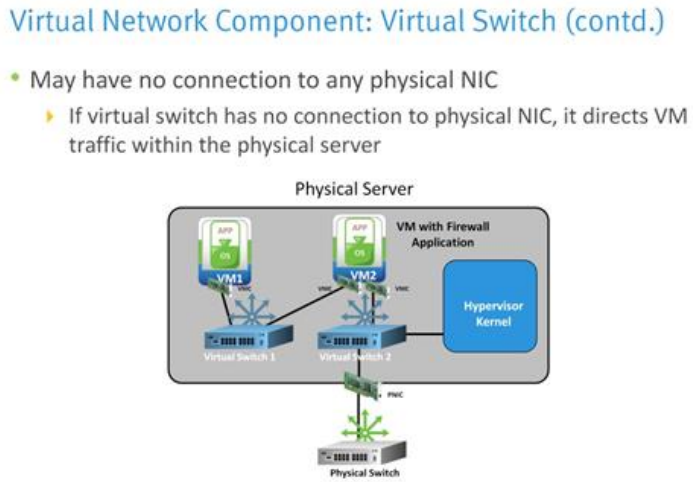
switch ports for frame forwarding. The virtual switch forwards frames to a virtual switch port based on the destination MAC address of the frame.

Virtual switches enable communication among VMs within a physical server and direct VM traffic to a physical network. Switching of VM traffic to physical network allows VMs to communicate with their clients or with VMs hosted on another physical server. A virtual switch also handles the hypervisor kernel traffic so as to enable the management server access the physical server, hypervisor kernel access the IP storage, and migrate VMs from one physical server to another.



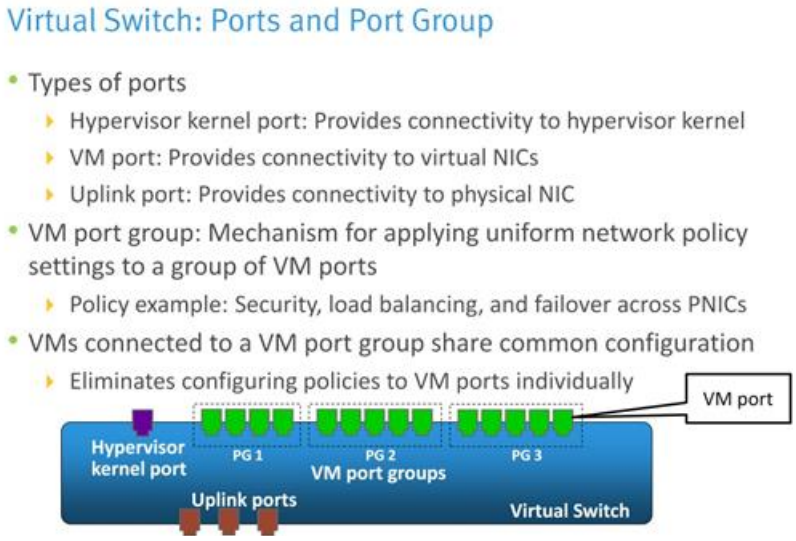
A virtual switch may be connected to one or more physical NICs. If a virtual switch is connected to more than one physical NIC, it allows the virtual switch to distribute outbound traffic across multiple physical NICs. Some of the physical NICs may be configured as standby. In the event of an active physical NIC hardware failure or its network link outage, virtual switches failover the traffic to a standby physical NIC.

A virtual switch has no control over the inbound traffic. The load balancing and failover of inbound traffic is performed by supported physical switches that are connected to the virtual switch via physical NICs.



Virtual switches can also be created without a connection to any physical NIC. In this case, all VMs that are connected to that virtual switch will only be able to send traffic among themselves locally; for example, a virtual switch connects a VM running firewall application to another VM protected by the firewall.

There is no direct connection between virtual switches within a compute system. However, frames may be transferred between virtual switches through VMs. Physical NICs provide links for virtual switches to connect to the physical network. Hence, they are not shared between virtual switches.



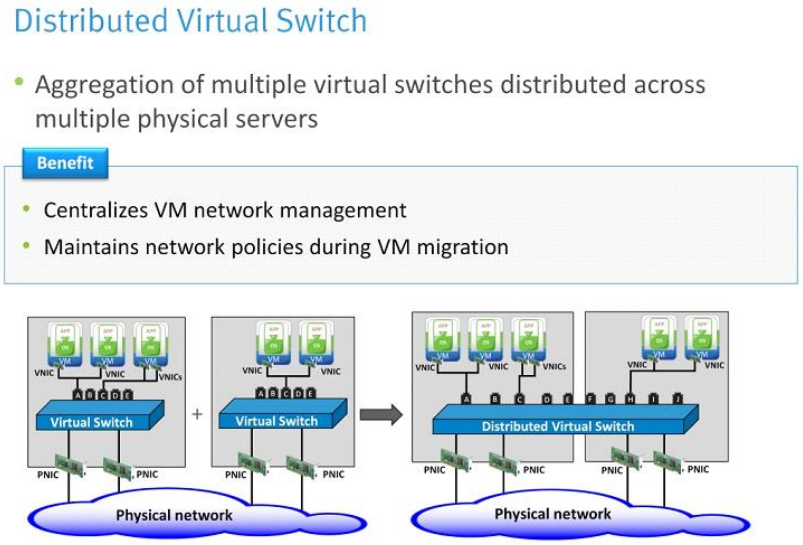
Virtual switches direct both hypervisor kernel traffic and VM traffic. For different types of traffic, different types of virtual ports are configured on a virtual switch. Alternatively, multiple virtual switches may be created; each with its own virtual port type. Virtual ports are classified as hypervisor kernel port, VM port, and uplink port.

- Uplink ports connect a virtual switch to physical NICs of the physical server where the virtual switch resides. A virtual switch can transfer data to a physical network only when one or more physical NICs are attached to its uplink ports.
- VM ports allow virtual NICs to connect to a virtual switch.
- A hypervisor kernel port enables the hypervisor kernel to connect to a virtual switch.

Port group is a mechanism for applying network policy settings to a group of VM ports and hence to a group of VMs. This allows an administrator to apply identical network policy settings across a group of VMs, rather than configuring the policies to VM ports individually. Examples of network policies are:

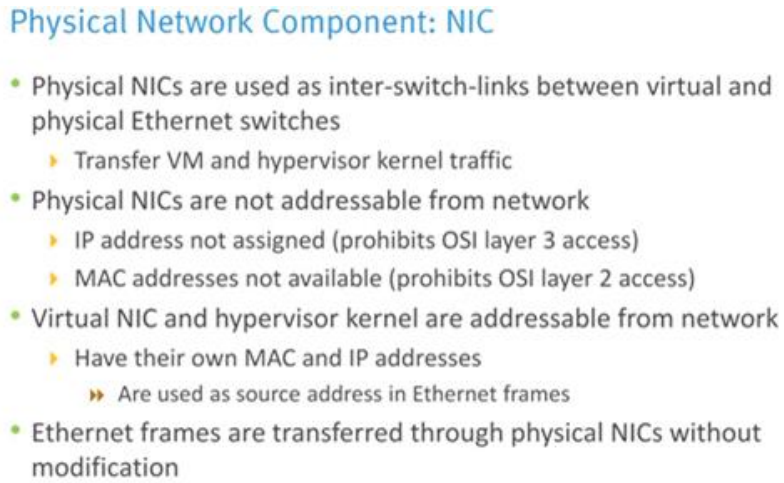
- Security
- Load balancing and failover across physical NICs
- Limiting network bandwidth for VMs
- Virtual LAN assignment to a VM port group to transfer the VM traffic

A virtual switch may have multiple port groups. VMs that are connected to a VM port group share a common configuration applied to the port group. An example of security policy setting for a port group is to stop receiving frames, if guest Operating System attempts to change the MAC address assigned to the virtual NIC. This helps protect against attacks launched by a rogue Operating System.



A distributed virtual switch is an aggregation of multiple virtual switches distributed across multiple physical servers. It functions as a single virtual switch and is created at the management server, which provides a single interface for managing a group of physical servers hosting VMs. Virtual switches that form the VM network still exist. However, virtual switch configuration and control is moved into the management server. The standalone virtual switches execute the frame forwarding based on the configuration made at the distributed virtual switch.

A distributed virtual switch brings up the next generation virtual networking concept. Instead of per physical server configuration, a distributed virtual switch brings in a centralized point of a VM network configuration. This simplifies configuration and ongoing administration of the VM network. Virtual switch port and port groups are configured at the distributed virtual switch level. Therefore, network policy settings remain consistent as VMs within a port group migrate across multiple physical servers. A distributed virtual switch allows movement of port group policies with VM.



A physical NIC provides an inter-switch-link between a virtual switch and a physical Ethernet switch. It is used to transfer VM and hypervisor kernel traffic between the VM and physical networks. It is called a link because it is not addressable from the network. Physical NICs are neither assigned an IP address (for OSI layer 3 access), nor are their built-in MAC addresses (for OSI layer 2 access) available to VMs or physical servers on the network.

Virtual NICs and hypervisor kernel are addressable from a network. A virtual NIC adds its MAC and IP addresses as source addresses to the Ethernet frames it transfers. Similarly, hypervisor inserts its own MAC and IP addresses before sending an Ethernet frame to a virtual switch. These Ethernet frames are transferred without modification through physical NICs.

Physical Network Component: HBA and CNA

Type of Adapter	Description
iSCSI HBA	• Transfers hypervisor storage I/Os (SCSI I/Os) to iSCSI storage systems • Has built-in iSCSI initiator • Encapsulates SCSI I/O into iSCSI frames and then encapsulates iSCSI frames into Ethernet frames • Uses its own MAC and IP addresses for transmission of Ethernet frames over the Ethernet network • Offloads iSCSI processing (SCSI to iSCSI) from hypervisor
FC HBA	• Transfers hypervisor storage I/Os (SCSI I/Os) to FC storage systems • Encapsulates SCSI data into FC frame • Uses its own FC address for transmission of frames over FC network
CNA	• Hypervisor recognizes as an FC HBA and as an NIC ▸ NIC : Used as a link between virtual and physical switches ▸ FC HBA : Provides hypervisor access to the FC storage

An iSCSI HBA transfers hypervisor storage I/Os (SCSI I/O) to iSCSI storage systems. iSCSI HBA has a built-in iSCSI initiator. The iSCSI initiator encapsulates SCSI I/O into iSCSI frames. Then, iSCSI HBA encapsulates iSCSI frames into Ethernet frames before sending them to iSCSI storage systems over the Ethernet network. Unlike physical NIC, iSCSI HBA inserts its own MAC and IP addresses into the Ethernet frames.

Hypervisor kernel has a built-in software iSCSI initiator. This software initiator is used to perform iSCSI processing when hypervisor accesses the iSCSI storage via physical NICs. If iSCSI HBAs are used instead of physical NICs to access iSCSI storage, the iSCSI processing is offloaded from the hypervisor kernel.

An FC HBA transfers hypervisor storage I/Os (SCSI I/O) to FC storage systems. Similar to an iSCSI HBA, an FC HBA has SCSI to FC processing capability. It encapsulates hypervisor storage I/Os into FC frames before sending the frames to FC storage systems over the FC network. FC HBA has its own WWN and FC addresses. The FC address is added as the source address to each FC frame.

- A Converged Network Adapter (CNA) is a single physical network adapter; although, to the hypervisor, it is recognized as an FC HBA and as an NIC. Therefore, a CNA provides the link between virtual and physical switches and also provides hypervisor access to the FC storage.

Module 5: Virtualized Data Center – Networking

Lesson 3: VLAN and VSAN Technologies

Topics covered in this lesson:
- Definition and benefits of VLAN and VSAN
- VLAN configuration
- VLAN and VSAN trunking and tagging
- Convergence of VLAN and VSAN traffic using FCoE

This lesson covers VLAN and VSAN technology including their benefits and configuration. It also includes VLAN and VSAN trunking, and convergence of VLAN and VSAN traffic in FCoE environment.

Virtual Local Area Network (VLAN)

VLAN

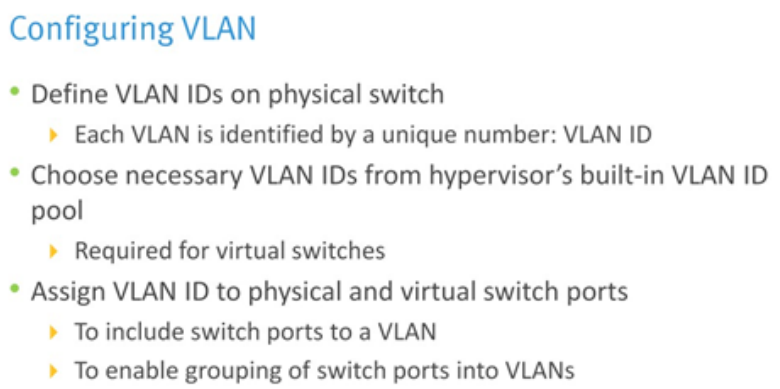
A logical network, created on a LAN or across LANs consisting of physical and virtual switches, enabling communication among a group of nodes, regardless of their location in the network.

Benefit

- Controls broadcast activity and improves network performance
- Simplifies management
- Increases security levels
- Provides higher utilization of switch and reduces CAPEX

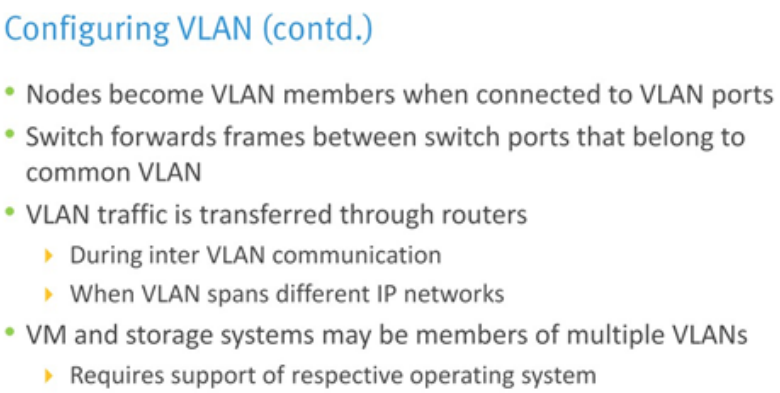
A VLAN is a logical network created on a LAN or across multiple LANs consisting of virtual and/or physical switches. The VLAN technology can divide a large LAN into smaller virtual LANs or combine separate LANs into one or more virtual LANs. A VLAN enables communication among a group of nodes based on functional requirements of an organization, independent of the nodes location in the network. All nodes in a VLAN may be connected to a single LAN or distributed across multiple LANs. The benefits of a VLAN are as follows:

- Broadcast traffic within the VLAN is restricted from propagating to another VLAN. For example, a node receives all broadcast frames within its associated VLAN, but not from other VLANs. Hence, the term VLAN is often used to convey a broadcast domain. Restricting broadcast using VLAN frees bandwidth for user traffic, which, thereby improves performance for the usual VLAN traffic.
- VLANs facilitate easy, flexible, and less expensive way to manage networks. VLANs are created using software. Therefore, they can be configured easily and quickly compared to building separate physical LANs for various communication groups. If it is required to regroup nodes, an administrator simply changes the VLAN configurations without moving nodes and re-cabling.
- VLANs also provide enhanced security by isolating sensitive data of one VLAN from any other VLANs. Restrictions may be imposed at the OSI layer 3 routing device to prevent inter VLAN routing.
- Since a physical LAN switch can be shared by multiple VLANs, the utilization of the switch increases. It reduces capital expenditure (CAPEX) in procuring network equipments for different node groups.



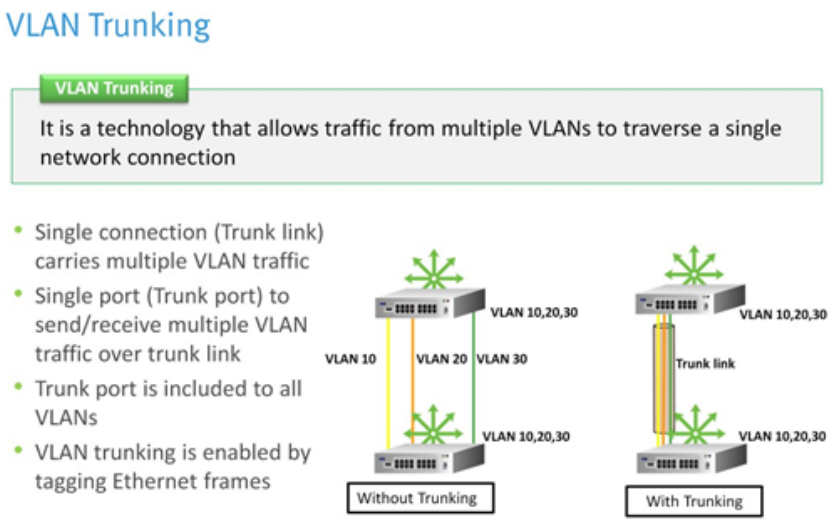
To create VLANs, an administrator first needs to define VLAN IDs on physical switches. Hypervisors have built-in VLAN ID pools. The administrator selects the necessary VLAN IDs from the pools. The next step is to assign a VLAN ID to a physical or virtual switch port or port group. By assigning a VLAN ID to a switch port, the port is included to the VLAN. In this manner, multiple switch ports can be grouped into a VLAN. For example, an administrator may group switch ports 1 and 2 into VLAN 101 (ID) and ports 6 to 12 into VLAN 102 (ID).

The technique to assign VLAN IDs to switch ports is called 'port-based VLAN' and is most common in VDC. Other VLAN configuration techniques are MAC-based VLAN, protocol-based VLAN, and policy-based VLAN.



When a node is connected to a switch port that belongs to a VLAN, the node becomes a member of that VLAN. Frames that are switched between ports of a switch must be within the same VLAN. Each switch makes forwarding decisions by frames and transfers the frames accordingly to other switches and routers. The VLAN traffic passes through a router for inter VLAN communication or when a VLAN spans across multiple IP networks.

Multiple VLANs may also be configured at the VM or storage system, provided guest Operating System or array Operating System supports such configurations. In this scenario, a node can be member of multiple VLANs.

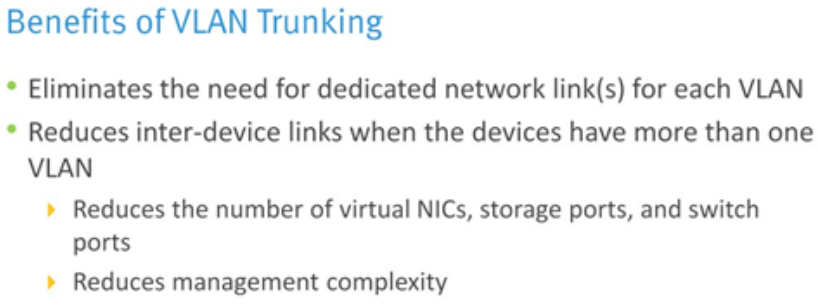


VLAN trunking allows traffic from multiple VLANs to traverse a single network connection. This technology allows for a single connection between any two networked devices, such as routers, switches, VMs, and storage systems with multiple VLAN traffic traversing the same path. The single connection through which multiple VLAN traffic can traverse is called a trunk link.

VLAN trunking enables a single port on a networked device to be used for sending or receiving multiple VLAN traffic over a trunk link. The port, capable of transferring multiple VLAN traffic, is called a trunk port. To enable trunking, the sending and receiving networked devices must have at least one trunk port configured on them. A trunk port on a networked device is included in all the VLANs defined on the networked device and transfers traffic for all those VLANs.

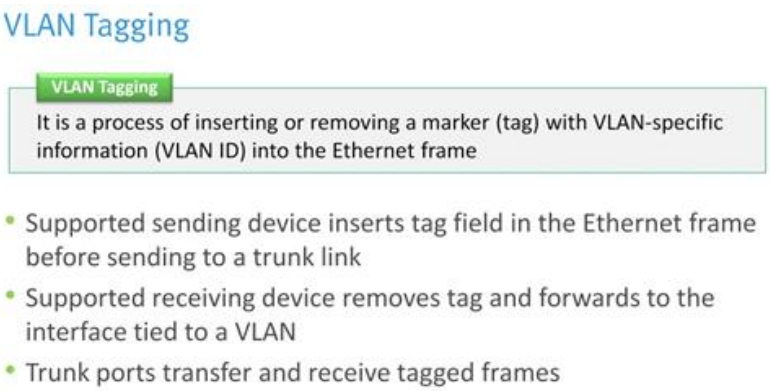
The mechanism used to achieve VLAN trunking is called VLAN tagging.

The diagram displayed on this slide illustrates the VLAN trunking against a network configuration without VLAN trunking. In both the cases, switches have VLAN 10, VLAN 20, VLAN 30 configurations. Without VLAN trunking, three inter switch links (ISLs) are used to transfer three different VLAN traffic. With trunking, a single trunk link is used to transfer the traffic of all VLANs.



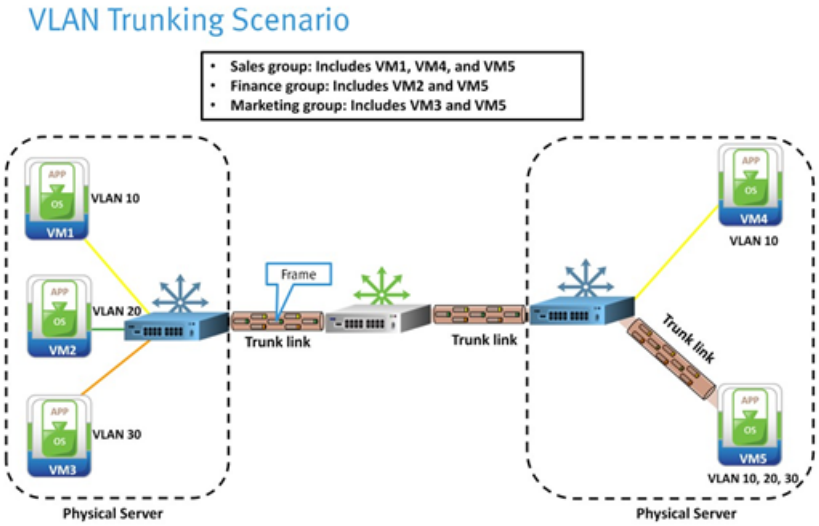
VLAN trunking allows for a single connection between any two networked devices (routers, switches, VMs, and storage systems) with multiple VLAN traffic traversing through the same network link. This eliminates the need to create dedicated network link(s) for each VLAN. It reduces the number of links between networked devices when the devices are configured with

multiple VLANs. As the number of inter device links decreases, the number of virtual and physical switch ports used for the inter device links reduces. It also cuts down the need for multiple virtual NICs and storage ports at VM and storage systems, respectively, to connect to the multiple VLANs. With reduced number of connections, the complexity of managing network links is also minimized.



The most common standard used to configure VLAN tagging in a network is IEEE 802.1Q. The standard uses a method of adding and removing a tag or VLAN ID (VID) to the Ethernet frame with VLAN-specific information, known as VLAN tagging. VLAN tagging allows multiple VLANs to share a trunk link without leakage of information between VLANs.

A networked device that supports IEEE 802.1Q VLAN tagging inserts a 4-byte tag field in the Ethernet frame before sending the frame down to a trunk link. At the receiving device, the tag is removed and the frame is forwarded to the interface, which is tied to a VLAN. To carry the traffic, belonging to multiple VLANs, between any two networked devices, device ports that are used for inter device link must be configured as trunk ports. A trunk port is capable of sending and receiving tagged Ethernet frames.



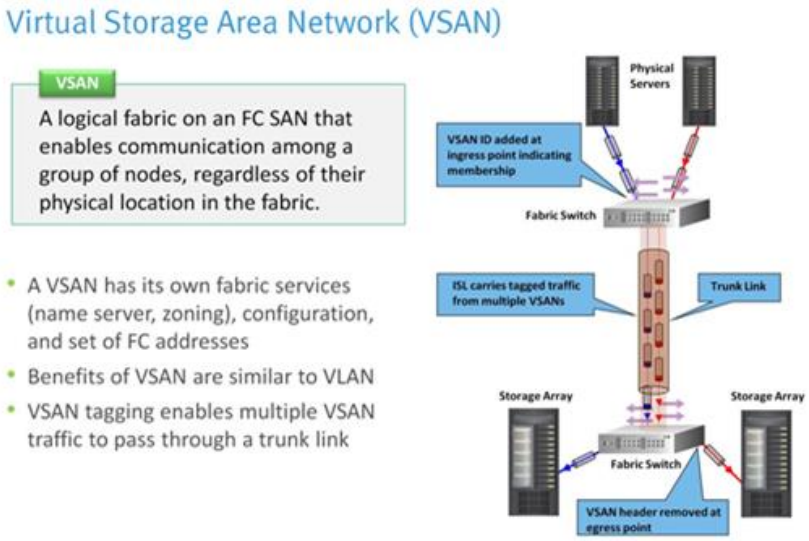
Consider a scenario where an organization has two physical servers with hypervisor. VM1, VM2, and VM3 reside in a physical server. VM4 and VM5 are hosted on another physical server. Each physical server has a virtual switch. These virtual switches are connected to a common physical switch to enable network traffic flow between them. VMs are connected to the respective virtual switches.

The organization has to set up three functional groups:

- Sales group: Includes VM1, VM4, and VM5
- Finance group: Includes VM2 and VM5
- Marketing group: Includes VM3 and VM5

The requirement is fulfilled by including VMs in sales, finance, and marketing groups to VLAN 10, VLAN 20, and VLAN 30, respectively, as shown in the diagram on this slide. Since

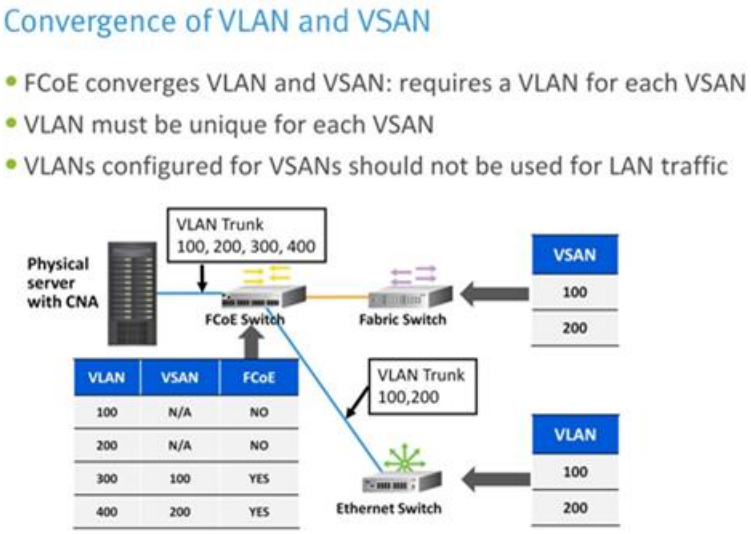
VM5 belongs to all the functional groups, it is included in all the VLANs. A trunk link is created between VM5 and the virtual switch to which VM5 is connected. The trunk link is used to transfer network traffic of VLAN 10, VLAN 20, and VLAN 30 simultaneously. To create the trunk link, the guest OS running on VM5 must support VLAN trunking. For same reason, links between physical and virtual switches are also configured as trunk links.



Virtual SAN or virtual fabric is a logical fabric, created on a physical FC SAN. Virtual SAN enables communication among a group of nodes (physical servers and storage systems) with a common set of requirements, regardless of their physical location in the fabric. VSAN conceptually functions in the same way as VLAN.

Each VSAN acts as an independent fabric and is managed independently. Each VSAN has its own fabric services (name server, zoning), configuration, and set of FC addresses. Fabric-related configurations in one VSAN do not affect the traffic in another VSAN. The events causing traffic disruptions in one VSAN are contained within that VSAN and are not propagated to other VSANs.

Similar to VLAN tagging, VSAN has its tagging mechanism. The purpose of VSAN tagging is similar to VLAN tagging in LAN. The diagram displayed on this slide shows the assignment of VSAN ID and frame-forwarding process.



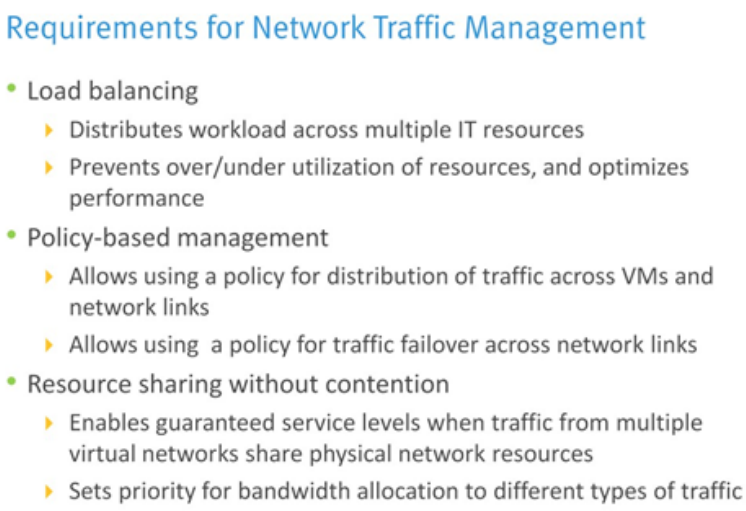
In Classic Data Center (CDC), FCoE facilitates the convergence of the LAN and FC SAN traffic over a single Ethernet infrastructure. In VDC, the same Ethernet infrastructure allows convergence of VLAN and VSAN traffic. In the converged environment, FC VSAN traffic transported over Ethernet must be assigned a unique VLAN on the FCoE switch. The VSAN to VLAN mapping is performed at the FCoE switch. VSANs cannot share a VLAN, and VLANs that

are used for LAN traffic should not be used for VSAN traffic.

In this example, the FCoE switch is configured with four VLANs – VLAN 100, VLAN 200, VLAN 300, and VLAN 400. The Ethernet switch is configured with two VLANs – VLAN 100 and VLAN

200. The fabric switch has VSAN 100 and VSAN 200 configured. To allow data transfer between physical server and fabric through the FCoE switch, VSAN 100 and VSAN 200 must be mapped to VLANs configured on the FCoE switch. Since VLAN 100 and VLAN 200 are already being used for LAN traffic, VSAN 100 should be mapped to VLAN 300 and VSAN 200 to VLAN 400.

This lesson covers requirements for network traffic management and the key network traffic management techniques that are used to control network traffic in a VDC.



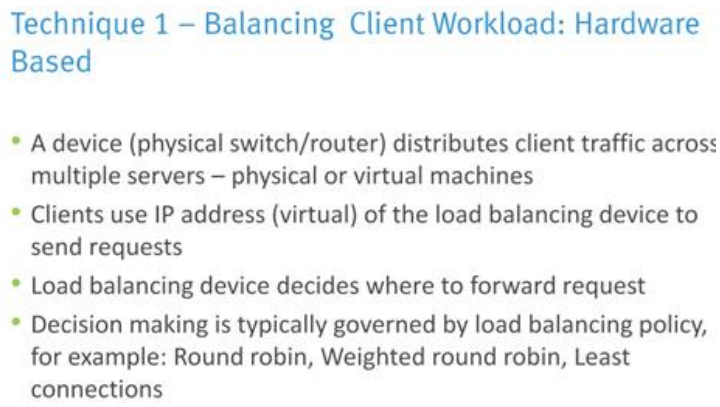
Similar to a Classic Data Center (CDC), in VDC, network traffic must be managed in order to optimize both performance and availability of networked resources. Although the network traffic management techniques described in this lesson are related to VDC, some of the techniques are similar to those used in Classic Data Center (CDC).

Load balancing is a key objective of managing network traffic. It is a technique to distribute workload across multiple physical or virtual machines and parallel network links to prevent overutilization or underutilization of these resources and to optimize performance. It is provided by a dedicated software or hardware.

In VDC, network administrators can apply a policy for distribution of network traffic across VMs and network links. Network traffic management techniques can also be used to set a policy to failover network traffic across network links. In the event of a network failure, the traffic from the failed link will failover to another available link based on a predefined policy. Network administrators have the flexibility to change a policy, when required.

When multiple VM traffics share bandwidth, network traffic management techniques ensure guaranteed service levels for traffic generated by each VM. Traffic management techniques allow an administrator to set priority for allocating bandwidth for different types of network traffic, such as VM, VM migration, IP storage, and management.

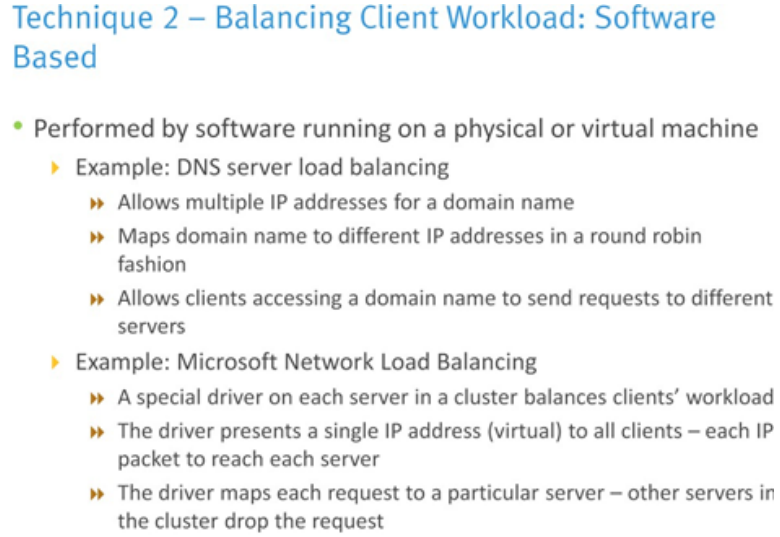
The slide provides a list of key network traffic management techniques. These techniques are described in subsequent slides.



The client load balancing service is usually provided by dedicated software or hardware such as a switch or router.

Hardware based load balancing uses a device, such as a physical switch or router, to split client traffic across multiple servers. The load balancing device resides between the server cluster and the Internet. This allows all client traffic to pass through the load balancing device. Clients use the IP address of the load balancing device to send requests. This IP address is called virtual IP address because it abstracts the real IP addresses of all servers in a cluster. The real IP addresses of the servers are known only to the load balancing device, which decides where to forward the request. Decision making is typically governed by a load balancing policy such as round robin, weighted round robin, and least connections policy.

- Round robin rotates client connections across servers.
- Weighted round robin allows an administrator to define a performance weight to each server. Servers with higher weight value receive a larger percentage of connections during the round robin process.
- Least connections maintains the same number of connections to all servers. Servers that are capable of processing connections faster receive more connections over time.

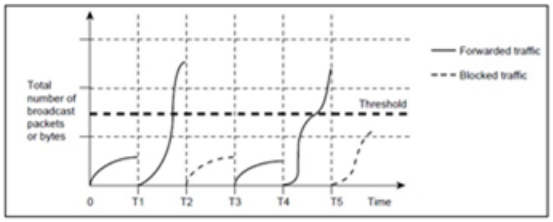


Software based client load balancing is performed by a software running on a physical or virtual machine. A common example is DNS server load balancing. In a DNS server, multiple IP addresses can be configured for a given domain name. This way, a group of servers can be mapped to a domain name. During translation of the domain name, the DNS server will map the domain name to a different server IP address in a round robin fashion. This allows clients accessing the same domain name to see different IP addresses, and thereby send request to different servers.

Microsoft Network Load Balancing is another key load-balancing technique available in Windows 2000 Advanced Server, Windows Server 2003, and Windows Server 2008. The load balancing is achieved by using a special driver on each server in a cluster, which balances clients' workload across clustered servers. The cluster presents a single IP address (virtual IP address) to clients, making the cluster appear as a single server. This allows each incoming IP packet to reach each server in the cluster. The load balancing driver software filters the incoming packets and maps each request to a particular server. The other servers in the cluster drop the request.

Technique 3 – Storm Control

- Prevents impact of storm on regular LAN/VLAN traffic
  - Storm: Flooding of frames on a LAN/VLAN creating excessive traffic and degrading network performance
- Counts frames of a specified type over 1-second and compares with the threshold
- Switch port blocks traffic if threshold is reached and drops the subsequent frames over the next time interval



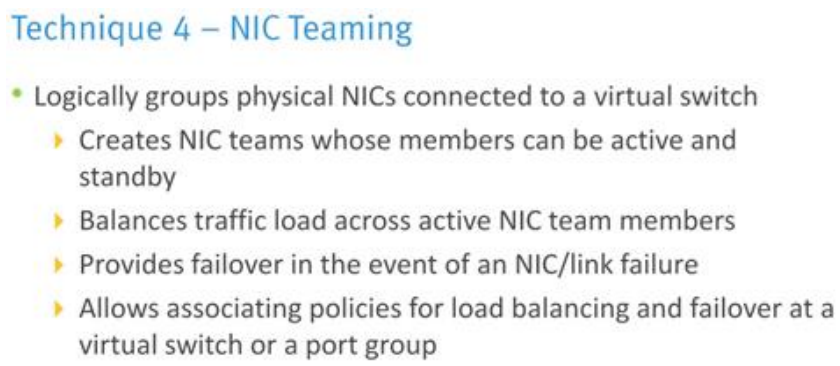
Source: "Catalyst 3550 Multilayer Switch Software Configuration Guide" – Cisco Systems, Inc.

A storm occurs due to flooding of frames on a VLAN or LAN segment, creating excessive traffic and resulting in degraded network performance. A storm could happen due to errors in the protocol-stack implementation, mistakes in the network configuration, or users issuing a denial-of-service attack.

Storm control is a technique to prevent regular network traffic on a LAN or VLAN from being disrupted by a storm and thereby improving network performance. If storm control is enabled on a supported LAN switch, it monitors all inbound frames to a switch port and determines if the frame is unicast, multicast, or broadcast. The switch calculates the total number of frames of a specified type arrived at a switch port over 1-second time interval.

The switch then compares the sum with a pre-configured storm control threshold. The switch port blocks the traffic when the threshold is reached and filters out subsequent frames over the next time interval. The port will be out of the blocking state if the traffic drops below the threshold. Storm control threshold may also be set as a percentage of port bandwidth.

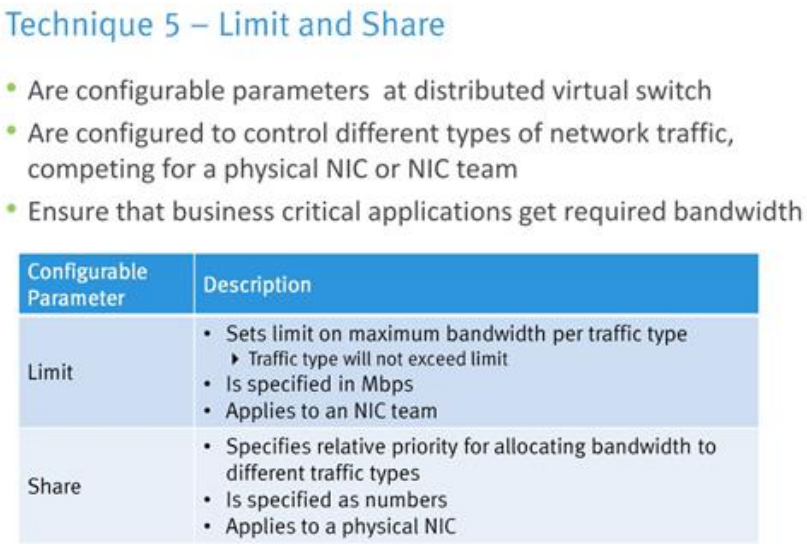
This slide shows an example of storm control where the broadcast traffic that arrived at a switch port exceeded the predefined threshold between the time intervals T1 and T2 and between T4 and T5. When the LAN switch monitored that the broadcast traffic had risen above the threshold, it dropped all broadcast traffic received by the switch port for the next time interval following T2 and T5. However, between T2 and T3, the broadcast traffic did not exceed the threshold. Hence, it was forwarded again at the next time interval following T3.



NIC teaming is a technique that logically groups (to create an NIC team) physical NICs connected to a virtual switch. The technique balances traffic load across all or some of physical NICs and provides failover in the event of an NIC failure or a network link outage.

NICs within a team can be configured as active and standby. Active NICs are used to send frames, whereas standby NICs remain idle. Load balancing allows distribution of all outbound network traffic across active physical NICs, giving higher throughput than a single NIC could provide. A standby NIC will not be used for forwarding traffic unless a failure occurs on one of the active NICs. In the event of NIC or link failure, traffic from the failed link will failover to another physical NIC.

Load balancing and failover across NIC team members are governed by policies set at the virtual switch. For example, a load balancing policy could be mapping a specific virtual NIC to a specific physical NIC. All outbound frames from the virtual NIC will be transferred through this physical NIC. In this process, network traffic from VMs will be forwarded through different physical NICs. Another load balancing technique is mapping between source and destination IP addresses of outbound traffic and physical NICs. Frames with specific IP addresses are forwarded through a specific physical NIC. NIC teaming policies may also be associated to a port group on a virtual switch.

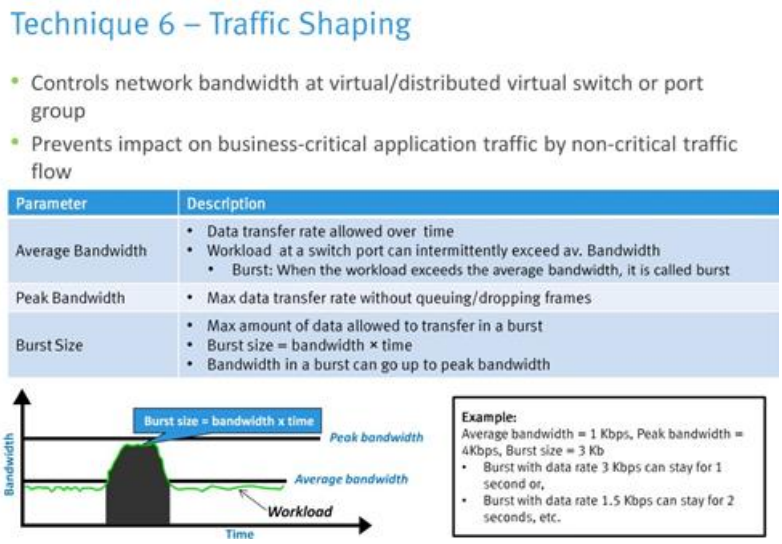


Limit and share are two network parameters, configured at the distributed virtual switch. These parameters are used to control different types of outbound network traffic such as VM, IP storage, VM migration, and management, when these traffic types compete for a physical NIC or NIC team. Limit and share improve service levels for critical applications. They prevent I/O intensive business critical application workload from being bogged down by less critical applications.

Limit, as the name suggests, sets a limit on the maximum bandwidth for a traffic type across an NIC team. The value is usually specified in Mbps. When set, that traffic type will not exceed the limit.

Shares specify the relative priority for allocating bandwidth to different traffic types when

different traffic types compete for a particular physical NIC. Share ensures that each outbound traffic type gets its share of physical NIC based on its priority. Shares are specified as numbers. For example, if iSCSI traffic has a share value of 1000, and VM migration traffic has a share value of 2000, then VM migration traffic will get twice the bandwidth of iSCSI traffic when they compete. If they were both set at same value, then both would get the same bandwidth.

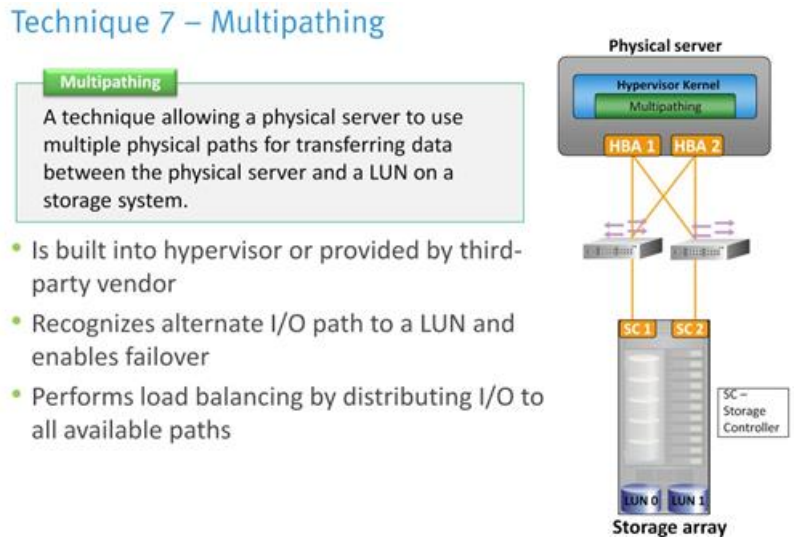


Traffic shaping controls network bandwidth so that business-critical applications have the required bandwidth to ensure service quality. Traffic shaping can be enabled and configured at the virtual switch/distributed virtual switch or at the port group level. Traffic shaping uses three parameters to throttle and shape the network traffic flow: average bandwidth, peak bandwidth, and burst size.

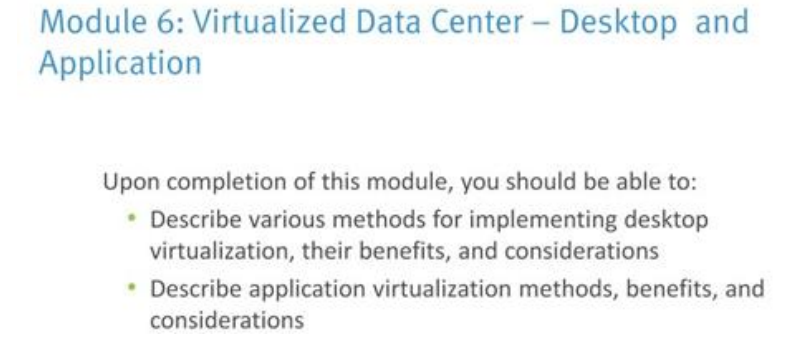
Average bandwidth is configured to set the allowed data transfer rate (bits per second) across a virtual switch/distributed virtual switch or a port group, over time. Since this is an averaged value over time, the workload at a virtual switch port can go beyond the average bandwidth for a small time interval.

The value provided for the peak bandwidth determines the maximum data transfer rate (bits per second) allowed across a virtual switch/distributed virtual switch or a port group without queuing or dropping frames. The value of peak bandwidth is always higher than the average bandwidth.

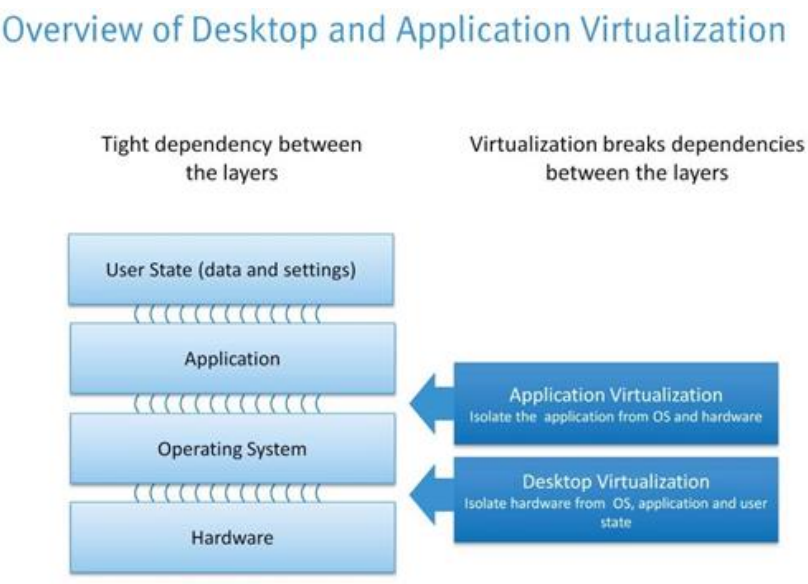
When the traffic rate at a virtual switch/distributed virtual switch port exceeds the average bandwidth, it is called burst. Burst is an intermittent event and typically exists for a small time interval. The burst size defines the maximum amount of data (bytes) allowed to transfer in a burst, provided it does not exceed the peak bandwidth. Burst size is a calculation of bandwidth multiplied by time interval during which the burst exists. Therefore, the higher the available bandwidth, lesser the time the burst can stay for a particular burst size. If a burst exceeds the configured burst size, the remaining frames will be queued for later transmission. If the queue is full, the frames will be dropped.



Multipathing is a technique which allows a physical server to use more than one physical path for transferring data between the physical server and a LUN on a storage system. Multipathing can be a built-in hypervisor function or can be provided by a third-party organization as a software module that can be added to the hypervisor. Typically, a single path from a physical server to a LUN consists of an HBA, switch ports, connecting cables, and the storage controller port. If any component of the path fails, the physical server selects another available path for I/O. The process of detecting a failed path and rerouting I/O to another is called path failover. Multipathing provides the capability to recognize an alternate I/O path to a LUN and enables failover. In addition to the failover, multipathing can perform load balancing by distributing I/O to all available paths.



This module focuses on the various aspects of desktop and application virtualization technologies.



With the traditional desktop, the OS, applications, and user data are all tied to a specific piece of hardware. With legacy desktops, business productivity is impacted greatly when an end-point device is broken or lost, because it affects the OS, applications, user data, and settings.

In a virtualized desktop, virtualization breaks the bonds between hardware and these elements, enabling the IT staff to change, update, and deploy these elements independently for greater business agility and improved response time. End users also benefit from virtualization because they get the same desktop, but with the added ability to access the computing environment from different kinds of devices and access points in the office, at home, or on the road.

Compute-based virtualization separates the OS from the hardware, but does not inherently address the broader and more important aspects of separating the user's data, settings, and critical applications, from the OSs themselves. These are very critical for maintaining a productive business environment. By separating the user's data, settings, and critical applications from OSs, organizations can begin to achieve a true IT–business aligned strategy. This is the fundamental concept underlying virtualized desktops.

Application virtualization breaks the dependency between the application and the underlying platform that includes the OS and hardware.

Module 6: Virtualized Data Center – Desktop and Application

Lesson 1: Desktop Virtualization

Topics covered in this lesson:

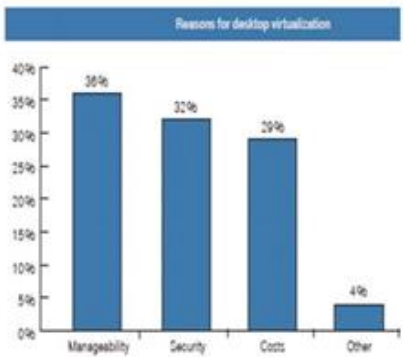
- Desktop virtualization drivers
- Benefits of desktop virtualization
- Desktop virtualization techniques
- User state virtualization

This lesson covers the drivers of desktop virtualization and benefits of desktop virtualization.

It also includes desktop virtualization techniques.

Desktop Virtualization - Drivers

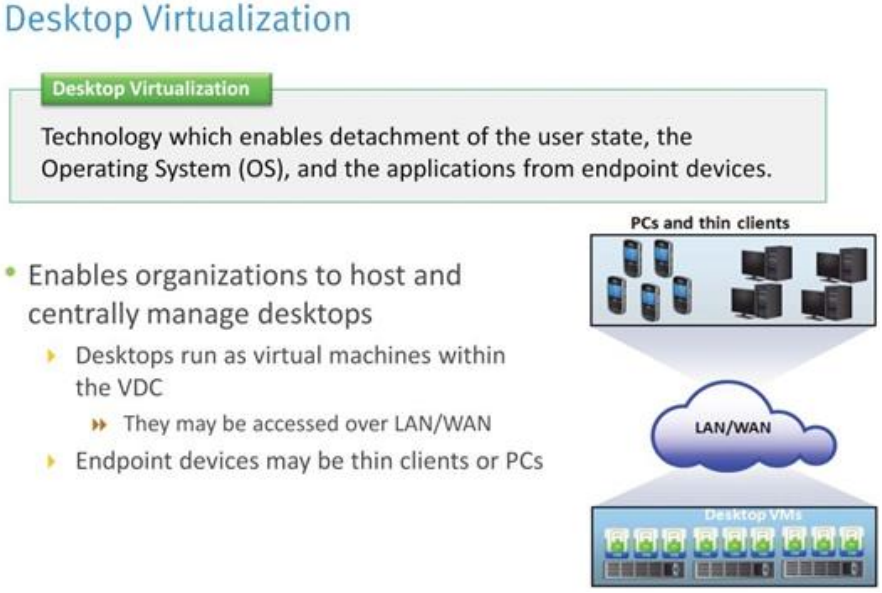
- Manageability concerns
  - Variety of hardware models, PC refresh cycles, and hardware incompatibilities
- Security concerns
  - Lost or stolen laptops/desktops
- Cost concerns



Manageability: One of the major concerns of an organization is management of desktop infrastructure. This is because the core component of a desktop infrastructure is the PC hardware, which continues to evolve rapidly. Organizations are forced to deal with a variety of hardware models, face complicated PC refresh cycles, and handle support problems related to PC hardware incompatibilities with applications. This makes the IT personnel spend most of their time supporting the existing environments instead of deploying new technology.

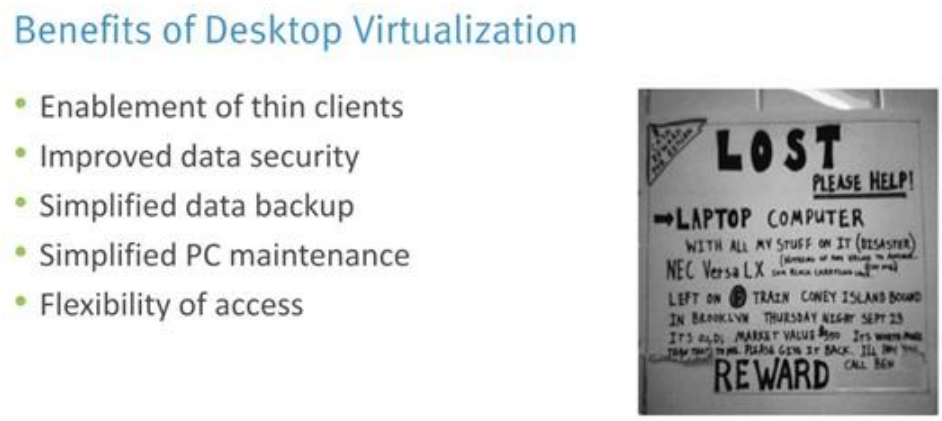
Security: Applications and data stored in traditional desktops pose severe security threats especially if the desktops/laptops are lost or stolen. Damage due to data loss can be immeasurable. Compliance regulations (e.g., Sarbanes-Oxley, HIPAA) protect privacy and corporate data by restricting the public disclosure of these data when issues arise. Failure to protect this data can lead to significant negative impact to the organization's reputation.

Cost: The cost of deploying and managing a traditional PC is high and ranges from $160 to $350.



The objective of desktop virtualization technology is to centralize the PC Operating System (OS) at the data center. This is to make security and management of desktops significantly easy. Desktops hosted at the data center run as Virtual Machines (VMs) within the VDC, while end users remotely access these desktops from a variety of endpoint devices.

Application execution and data storage do not happen at the endpoint devices; everything is done centrally at the data center.



The key benefits of desktop virtualization are as follows:

• **Enablement of thin clients:** Desktop virtualization enables the use of thin clients as endpoint devices. This creates an opportunity to significantly drive down the cost of endpoint hardware by replacing aging PCs with a thin-client device, which typically operates across a life span twice that of a standard PC. Thin clients consume very less power when compared to standard PCs. Hence, they support the "Go Green" strategy of organizations and reduce OPEX.

• **Improved data security:** Since desktops run as VMs within an organization's data center, it mitigates the risk of data leakage and theft and simplifies compliance procedures.

• **Simplified data backup:** Since centralized virtual desktops reside entirely within an organization's datacenter, it is easier to ensure full compliance with backup policies.

• **Simplified PC maintenance:** Virtual desktops are far easy to maintain than traditional PCs.

Because of the unique characteristics of VMs, it is simple to patch applications, provision/remove users, and migrate to new OSs.

• **Flexibility of access:** If corporate desktop environments are centralized using desktop virtualization techniques, access to them can be provided to users who do not have access to their corporate PCs. This is especially useful in situations where users need to work from home, away from their desks, or other remote worker scenarios.

Desktop Virtualization Techniques

- Technique 1: Remote Desktop Services(RDS)
- Technique 2: Virtual Desktop Infrastructure (VDI)
- Desktop virtualization techniques provide ability to centrally host and manage desktop environments
  - Deliver them remotely to the user's endpoint devices

There are two desktop virtualization techniques: RDS and VDI. Both of these solutions provide the ability to centrally host and manage desktop environments and deliver them remotely to the user's endpoint devices.

Remote Desktop Services

- RDS is traditionally known as terminal services
- A terminal service runs on top of a Windows installation
  - Provides individual sessions to client systems
  - Clients receive visuals of the desktop
  - Resource consumption takes place on the server

Remote Desktop Services (RDS) is traditionally known as Terminal Services. In RDS, a Terminal Service runs on top of a Windows installation and provides individual sessions to client systems. These sessions can provide a thorough desktop experience while remotely accessing via a terminal services client. The workstation receives the visual feedback of the session, while resource consumption takes place on the server.

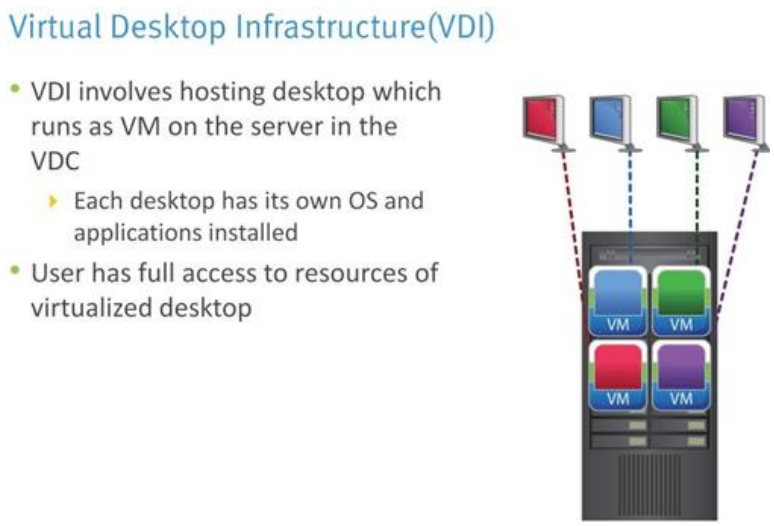
Benefits of Remote Desktop Services

- Rapid application delivery
  - Applications are installed on the server and accessed from there
- Improved security
  - Applications and data are stored in the server
- Centralized management
- Low-cost technology when compared to VDI

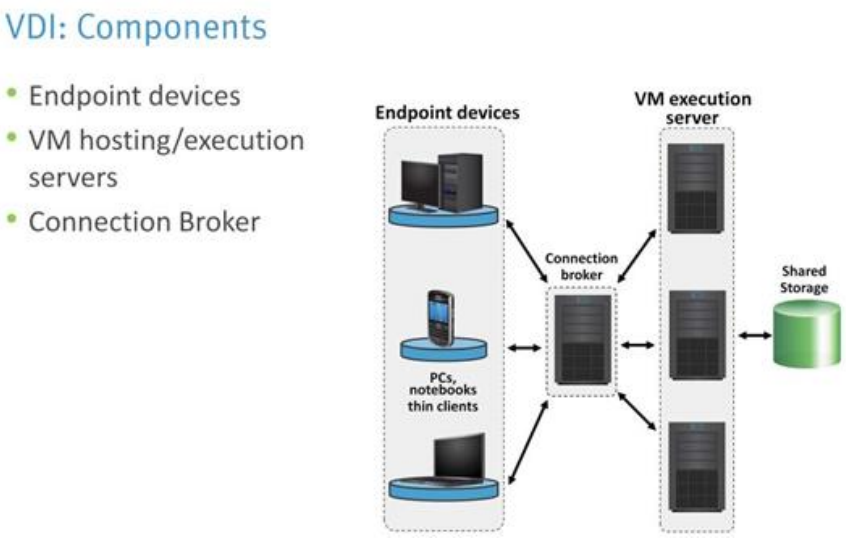
Application delivery is more agile and rapid because applications are installed once on the server and can be accessed by the user without having to install the application locally. Security of

the user environment is improved because the user's applications and data are stored, secured, executed, and accessed centrally instead of being located on the user's local desktop device. Management of the user environment is easier than a distributed desktop environment because of the centralization of the OS environment, applications, and data settings and its configuration on servers in the datacenter. For example, adding a new user environment is completed without having to install an OS or applications on the local endpoint device. Software updates, patches, and upgrades are also completed on servers instead of on the user's endpoint device. RDS is a mature technology, easy to implement, and scales well, compared to a VDI solution. An RDS solution can support more users per server than a VDI solution; for example, a typical RDS solution can support 250 or more users per server, compared to a typical VDI solution supporting 30–45 users per server. RDS is also less expensive to implement in terms of initial acquisition cost and supporting infrastructure.

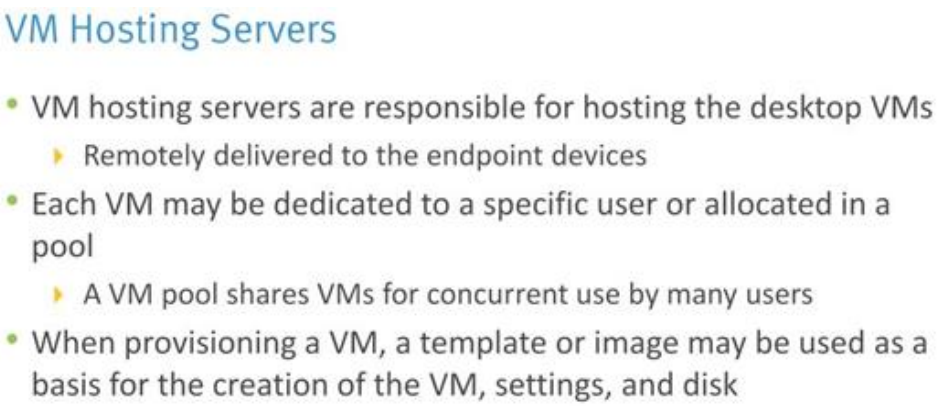
The application developed for desktop operating systems may not necessarily be compatible with server operating systems; for example: Microsoft Windows XP versus Windows Server 2007. This limits the number of desktop applications that could be centralized using RDS.



Virtual Desktop Infrastructure (VDI) refers to the hosting of a desktop OS running in a VM on a server in the VDC. A user has full access to the resources of the virtualized desktop. The server-hosted desktop virtualization solution approach is sometimes called as Virtual Desktop Environment (VDE). VDI allows a user to access a remote desktop environment from an endpoint device via a remote desktop delivery protocol. The hosted remote OS and associated applications are shown on the user's endpoint device display and controlled via the endpoint device's keyboard and mouse. For the user, the experience is very similar to using the RDS solution, except that the desktop OS is running in a VM hosted on a server, instead of on a remote user session on a single Server OS.



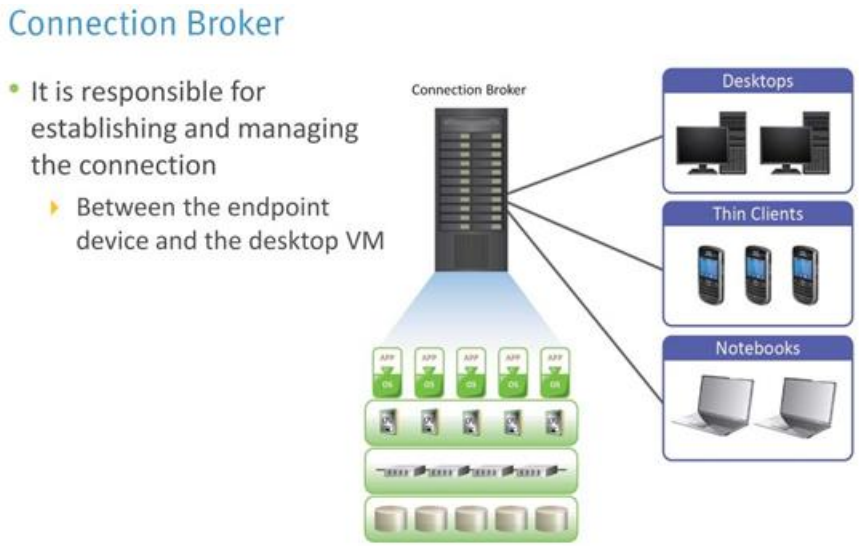
The VDI architecture consists of several components that work together to provide an end-to-end solution. The main components are endpoint devices, a connection broker, and VM hosting servers.



VM hosting servers are responsible for hosting the desktop VMs that are remotely delivered to the endpoint devices. Type 1 hypervisors are most suitable for this type of solution because of their performance and scalability benefits. Additionally, they support a greater number or density of VMs than type 2 hypervisors. The number of VMs that a single server can support depends on several factors including hardware, software, configurations, and user workloads of the desktop VMs.

The desktop VMs may be running any desktop OS, supported by the hypervisor. In VDI, each VM may be dedicated to a specific user(persistent desktop) or allocated in a pool(non-persistent desktop). A VM pool shares VMs for concurrent use by many users. VM pools have the potential to save VM disk storage requirements. For example, if an organization has 100 users, but only 50 are concurrent at any one time, then they could pre-provision 50 VMs in a pool and allocate them as per request. This deployment option saves 50 percent of the VM storage required for individual OSs and applications. Each VM can be preprovisioned for later use or dynamically provisioned at the time of end-user request.

When provisioning a VM, a template or image can be used as a basis for VM creation, settings, and disk. Advanced provisioning technology allows a VM to be provisioned using a single image. When there is a request, instead of allocating a complete disk image for every user VM, only a single image is used to provision the VM dynamically. This provisioning solution option has the potential to save a significant amount of VM disk storage. The savings result in provisioning a VM from a single image, thereby not requiring a full VM disk image for every VM.



The connection broker is responsible for establishing and managing the connection between the endpoint device and the desktop VM that is running on the VM-hosting server. The connection broker provides manageability, security, reliability, and scalability benefits. A connection broker is generally needed as part of the VDI solution for a larger or more complex environment. One example where a connection broker would be needed is with VM pools. In this case, the connection broker would connect the user to an available VM in the pool. There are some cases where a connection broker may not be needed for a VDI solution. One example would be when an organization dedicates VMs to each user. In this case, the user would be configured to connect directly to the same VM by a computer name or IP address. It is important to determine whether the connection broker is compatible and supports the hypervisor that is used by the VM hosting server.

Improved Deployment and Management: VDI has the potential to improve desktop OS deployment agility and management efficiency. Deployment agility improvements come from centralizing the desktop OS images on servers in the datacenter and creating virtual images of the OSs. A VDI environment can enable rapid desktop OS environment provisioning and deployment. This is enabled by virtualizing the desktop environment by decoupling the OS environment from the physical hardware of the desktop device. Desktop OS environments are provisioned from preconfigured VM templates or images. They are then deployed by distributing the VM files to the appropriate hosting server, from where they will run and be accessed by the user. A VDI environment can be more efficient to manage than a distributed, nonvirtualized desktop environment. Increased management efficiency is achieved because the desktop OS environment is centralized on hosting servers in the datacenter. By this approach, desktop environments can be more effectively managed.

Improved Security: VDI also provides security benefits for the desktop environment. The benefits come from storing and running the desktop OS environment on centralized servers located in the datacenter, instead of on distributed desktops or endpoint devices. If the desktop or laptop is lost or stolen, the user's desktop environment is still protected.

Centralization of the desktop environment on servers in the datacenter, good management practices, and software can make implementation of security software updates more efficient.

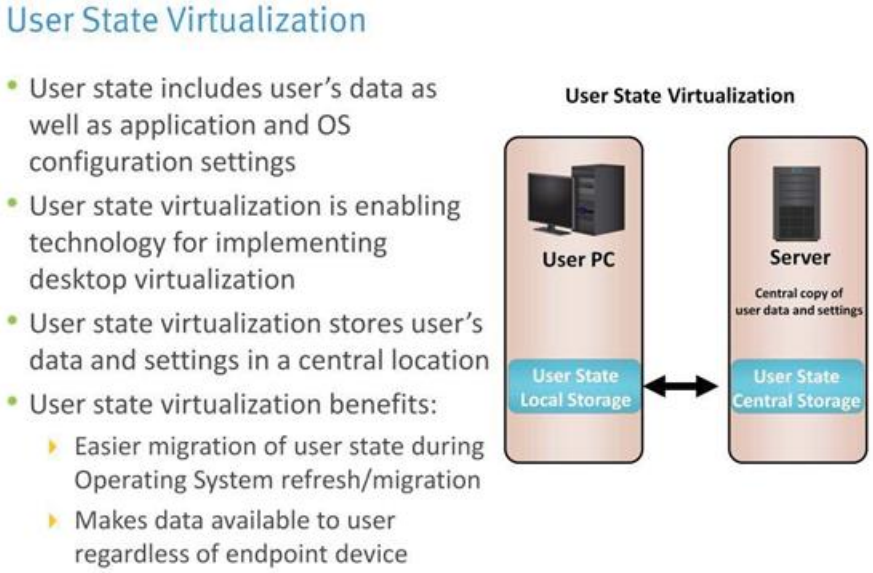
Better Business Continuity and Disaster Recovery: Desktop business continuity and disaster recovery can be improved by implementing VDI. Business continuity improvements are enabled by centralizing the desktop OS, applications, and data in the datacenter, where administrators can easily perform backup and recovery operations.

The Limitations of VDI are as follows:

• **Reliance on Network Connection:** A VDI solution relies on the network connection to work. If the user's endpoint device cannot connect to the hosting server, the user will be unable to access their desktop, applications, and data. In addition to reliability of connection, the network should have suitable bandwidth and low latency to provide a good user experience.

• **Unsuitable for High-End Graphic Applications:** Typically, a VDI solution is not suitable for users who use high-end graphic applications. This is mainly due to the inability of the remote desktop delivery protocols in a VDI solution to provide the required performance levels for this type of application.

• **Additional Infrastructure Required:** A VDI solution requires additional servers, storage, and networking infrastructure. The amount of additional infrastructure required depends on the scale, performance, and service level requirements of an organization. A VDI solution may potentially introduce a significant number of new servers and added storage into the datacenter; for example, if an organization is planning to support 5,000 dedicated users with their VDI solution, hosting 50 users per server, they would need 100 additional servers just to run the VMs. Additional shared storage would also be needed for this solution and the amount depends on the VM provisioning methods used.



User state virtualization is a key requirement for implementing desktop virtualization.

The user state includes user's data as well as application and OS configuration settings. Traditionally, user's PCs contain the authoritative copy of user's data and settings. There are three main challenges with managing the user state.

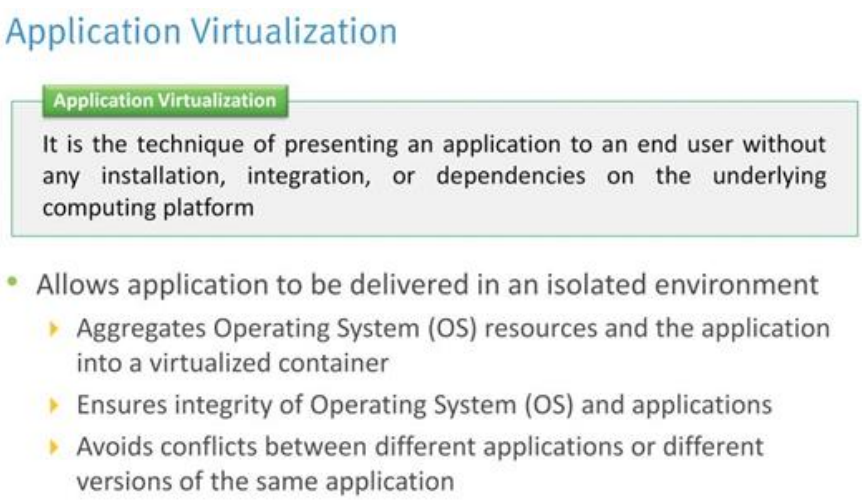
• The first challenge is how to back up user data and settings that are scattered from PC to PC and restore the user's productivity after a laptop is lost or stolen.

• The second challenge is how to migrate the user state during OS refresh/migrations.

• The final challenge is how to make the data available to the user, regardless of the PC (endpoint device) used. In any case, user state virtualization provides a solution.

In any of the above situation, 'user state' virtualization provides a solution. With user state virtualization, organizations store user's data and settings in a central location. The result is that users are free to roam, and their data and settings follow them. User state virtualization can also mitigate productivity loss of PC replacement. The central copy of the data is on the network. So, it is easily restored in case of a lost or stolen PC, and the user's settings can be re-applied automatically.

This lesson covers the application virtualization deployment methods and benefits of application virtualization.



Application virtualization software is increasingly leveraged by organizations to address challenges associated with the application life cycle, particularly those processes associated with application testing and deployment. Fundamentally, application virtualization software aggregates OS resources and isolates them within a virtualized container along with the application that accesses them. This technology provides the ability to deploy applications without modifying or making any change to the underlying OS, file system, or registry of the computing platform in which they are deployed. Because virtualized applications run in an isolated environment, the underlying OS and other applications are protected from potential corruptions which may be caused due to installation modifications. There are many scenarios where conflicts may arise if multiple applications/multiple versions of the same application are installed on the same computing platform. One such example is where Microsoft Access databases have been employed by an organization using a different version of Microsoft Access than the company standard. Because two versions of Microsoft Access cannot be installed on the same computing platform at the same time, one (or both) of them may be virtualized to overcome this problem and used simultaneously.

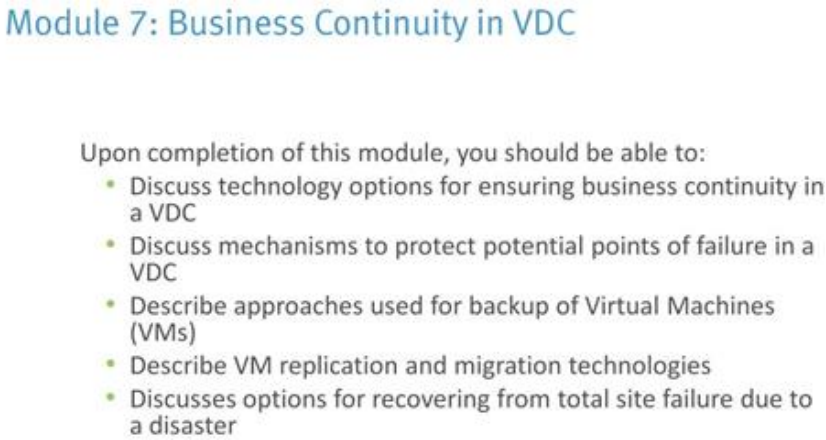
The two methods which are commonly used for deploying application virtualization are application streaming and application encapsulation.

Application encapsulation packages the application in a self-contained executable package that does not rely on a software installation or an underlying OS for any dependencies. This package is accessible from the network on a USB key or via local storage. Because these applications have the capability to function like standalone executables, they do not require any agent to be installed locally in the client machine where they run (built-in agents are present within the package).

Application Streaming involves transporting application specific data/resources to the client device when the application is executed. Only a minimum amount of data (commonly between 10 to 30 percent of the total application) is delivered to a client before the application is launched. Hence, the first time launch of the application happens very quickly and the load on the network is also reduced. Additional features of the application are delivered on demand or in the background without user intervention. Application packages are stored on a (centralized) server. Application streaming is suitable in a well-networked environment (Example: VDI, RDS, and so on). Streaming involves the use of a locally installed agent on the client machine. This agent has the functionality to setup and maintain the Virtual Environment for each application. The agent takes care of management tasks (such as Shortcut creation) and is a key component in the streaming behavior.

Benefits of Application Virtualization are as follows:

- **Simplified application deployment/retirement:** Applications are never installed on to an OS; hence, the deployment of the applications is greatly simplified. Furthermore, complete removal of all application bits from a PC during retirement is assured.
- **Simplified Operating System image management:** Because applications are completely separated from the OS, managing OS images is simpler, especially during OS patches and upgrades. It helps to create a more dynamic desktop environment, in which the desktop is an aggregation of separately-managed components.
- **Elimination of resource conflicts:** Because each application has its own virtual OS resources, resource contention and application conflict issues are eliminated.



This module focuses on the concepts and techniques employed for ensuring business continuity in a Virtualized Data Center (VDC) environment. It discusses the mechanisms to protect single point of failure in a VDC. Next, it describes the various technology options for backup, replication, and migration of Virtual Machines (VM) and their data in a VDC environment. Finally, it discusses the various options for recovering from total site failure due to a disaster.

Module 7: Business Continuity in VDC

Lesson 1: Fault Tolerance Mechanisms in VDC

Topics covered in this lesson:

- An overview of BC in a VDC
- Single point of failure in a VDC environment
- Mechanisms to protect compute, storage, and networking components in a VDC
- Disaster recovery mechanisms for a VDC site failure

This lesson covers an overview of Business Continuity (BC) in a VDC environment, identification of single point of failure, mechanisms to protect compute, storage, and networking components in a VDC, and disaster recovery mechanisms during a VDC site failure.

Business Continuity in VDC: An Overview

- BC planning should include end-to-end protection of both physical and virtual resources at
  - Compute, storage, and network layers
- Ensuring BC mainly involves redundancy of components at each layer
- BC technology for data protection includes
  - Backup and replication of data (similar to CDC environment)
- Besides traditional approaches, VDC environment has additional BC solutions typically implemented at the compute layer

In a VDC environment, technology solutions that enable the Business Continuity (BC) process need to protect both physical and virtualized resources. This protection should include resources at all the layers including compute, storage, and network. Ensuring BC in a VDC environment involves consistent copies of data and redundant infrastructure components. BC solutions for physical network and storage are the same as those for the Classic Data Center (CDC) environment. This module focuses on the BC solutions for virtual resources built upon compute infrastructure including VMs and their data.

Advantages of Compute Virtualization in BC

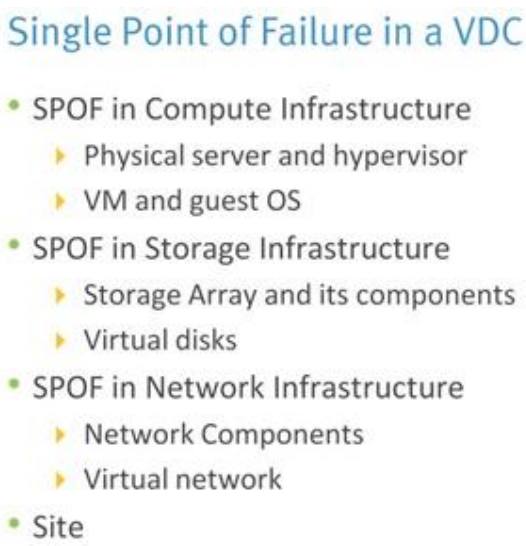
- Hardware independence
- Cross platform compatibility
- Mutual isolation
  - Different BC policies may be applied to different VMs, even if they are running on the same physical server
- Encapsulation of complete computing environment
- Relatively robust BC processes
  - Comparatively easier to maintain VM copies in diverse geographic locations
- Higher data availability

Virtualization considerably simplifies the implementation of a BC solution in a VDC

environment. For example, VMs have inherent properties that facilitate the planning and implementation of a BC strategy:

- VMs are compatible with standard x86 architectures and run independent of the underlying x86 hardware configurations. Therefore, BC solutions for VMs need not consider the details of the x86 hardware available at the failover site. In this module, the term compute hardware term is used to represent x86 hardware.
- VMs are isolated from each other, as if physically separated, even if they run on a single physical server. This isolation prevents failure of one VM from spreading over to other VMs. Different DR policies may be applied to different VMs, even if they are running on the same physical server.
- VMs encapsulate a complete computing environment. Therefore, they can be moved and restarted easily, compared to recovery of physical servers in a CDC.

It is also comparatively easier to maintain VM copies in diverse geographic locations, which makes the BC process robust. Similarly, data maintained within a VDC has higher availability. Restoring of data after an outage in a VDC is faster and more reliable, compared to CDC.



To plan for an effective BC solution in a VDC, administrators need to identify potential single point of failure (SPOF). In a VDC, the SPOF includes the following:

SPOF in Compute Infrastructure:

•**Physical Server**: Failure of a physical server would result in the failure of complete virtualized infrastructure running on it. Similarly, failure of a hypervisor can cause all the running VMs and virtual network, which are hosted on it, to halt.

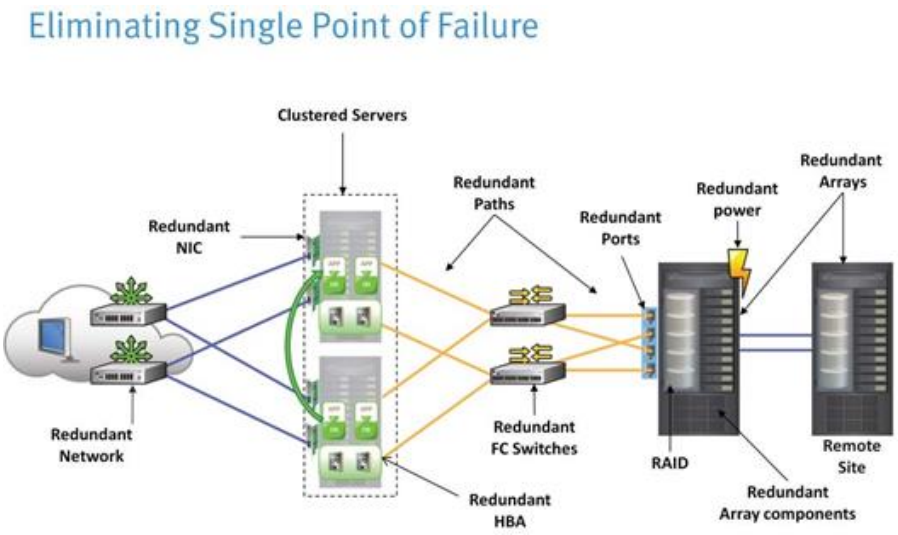
•**VM**: Failure of a VM might result in the failure of the critical applications running on it. If there are (distributed) applications running across multiple VMs (for example an electronic banking application), failure of a VM might cause disruptions in the execution of distributed or dependent applications on other VMs. Similarly, the failure of a Guest OS might cause critical applications to stop executing, which would consequently cause disruptions in the client services.

SPOF in Storage Infrastructure: Failure of a storage array and its components might cause disruptions in the operations of hypervisor, VMs, and applications accessing data from that array. Apart from physical storage, virtual disks are also subject to potential failure. However failures in virtual disks are often due to errors in VMFS, which would be either due to incorrect operation of an hypervisor or a security attack on the file system itself. Failure in VMFS or virtual disk would impact all the applications in the associated VMs.

SPOF in Network Infrastructure: Based upon the underlying network infrastructure – SAN, NAS, or converged network type (for example, FCoE, IP SAN, etc.) – the supporting

components (NICs, communication paths, interconnect device, etc.) might fail and thus disrupt communication between the devices. Apart from physical network component, virtual network is also subject to potential failures. These failures in a virtual network could be due to incorrect design, the configuration of the applications realizing these virtual components, or due to security attacks on these applications. Failure in a virtual network component results in the disruption of communication between the VMs using that component for their communication.

Site: Failure of a VDC site due to a disaster could have significant impact on business operations and thus need special consideration in terms of detailed disaster recovery planning and testing.



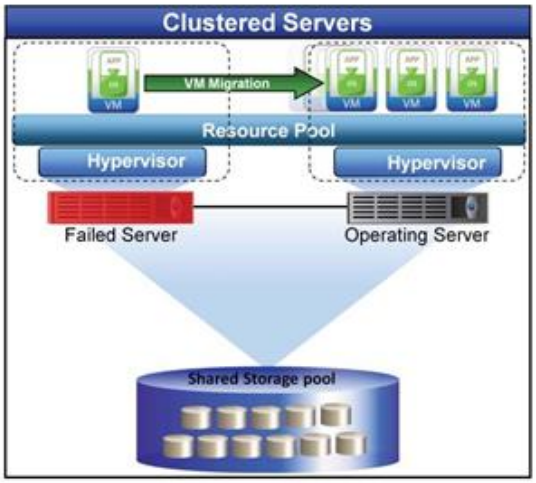
To mitigate SPOFs identified before, VDC components are often designed with redundancy so that the system fails only if all the components, including the redundant components, fail simultaneously. Such a failure is very unlikely to happen except during catastrophic disasters. Redundancy ensures that the failure of a single component does not affect the availability of the corresponding operations.

Because of the large number of components involved in the overall operation of a VDC and their underlying complexity and interactions with each other, a careful analysis need to be performed to eliminate every SPOF. In the diagram, critical enhancements in the infrastructure of a VDC to mitigate SPOF is illustrated. Note that many of the employed mechanisms are similar to a CDC environment:

- Configuration of multiple HBAs to mitigate single HBA failure.
- Configuration of multiple fabrics to account for a switch failure.
- Configuration of multiple storage array ports to enhance the storage array's accessibility.
- RAID configuration to ensure continuous operation in the event of disk failure.
- Implementation of a storage array at a remote site to replicate data for mitigating the effect of local site failure.
- Configuration of multiple copies of virtual disks and VM configuration files.
- Implementation of server clustering.
- A fault-tolerance mechanism whereby two VMs in a cluster access the same set of storage volumes. If one of the VMs fails, the other takes up the complete workload.

Protecting Compute: Clustering

- Clustered Servers
  - Groups of physical servers are combined and managed as an aggregated compute resource pool
    - All servers have access to shared memory
  - Provides protection from server and hypervisor failures
  - Uses clustered file system, such as VMFS, to enable failover



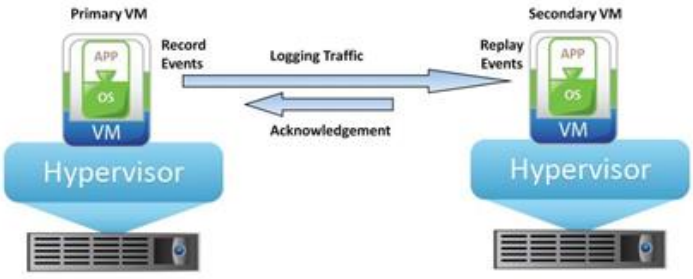
One of the important techniques to protect compute infrastructure, for providing high availability of VMs and their data in a VDC environment, is clustering.

Clustered Servers : Virtual Infrastructure provides the technology to combine groups of physical servers and manage them as an aggregated compute resource pool. These resource pools are an ideal way to abstract the underlying physical servers and present only the logical capacity to the user. Additionally, clustered servers provide an effective way to ensure compute resource availability in the event of a physical server or hypervisor failure. For example, as depicted in the diagram on this slide, if a server fails, all the VMs running on it are failed over (moved) to other servers, which are operational (active) at that time. Clustered servers use a clustered file system, such as VMFS, to enable failover.

VMFS provides multiple VMs with shared access to a consolidated pool of clustered storage. A VM sees the (virtual) disks in a VMFS as local targets, where as are actually just files on the VMFS volume. Additionally, a VMFS encapsulates the entire VM state in a single directory, making the tasks of backup, replication, and migration of VMs easy. VMFS is implemented over CS and allows each hypervisor in the cluster to store its VM files in a specific subdirectory on the VMFS. When a VM is operating, VMFS locks those files so that other hypervisors cannot update them. VMFS, thus ensures that a VM cannot be opened by more than one hypervisor at the same time. VMFS also provides each VM a separate isolated directory structure for storing its files so that files belonging to one VM cannot be accessed by other VMs.

Protecting Compute: VM Fault Tolerance

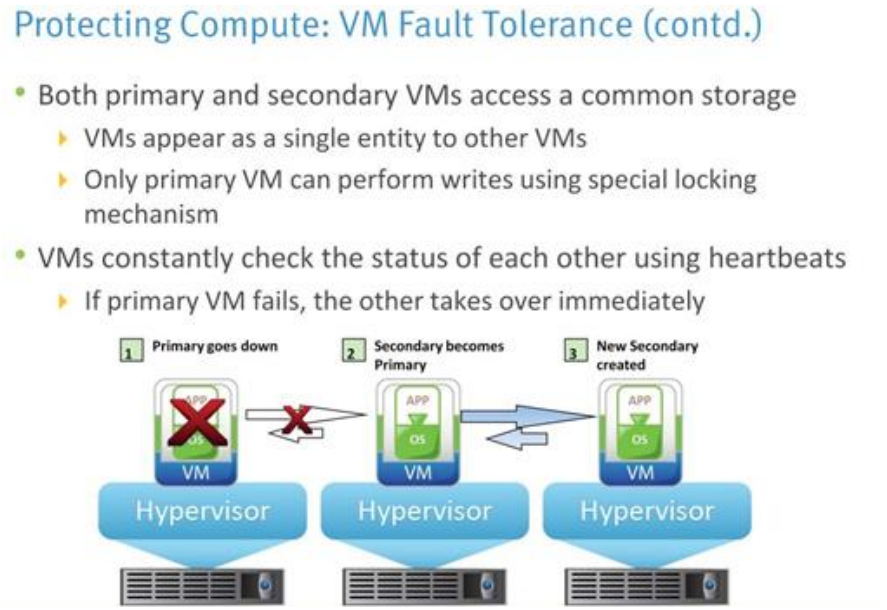
- Uses a secondary VM running on another physical machine as a live copy of the primary VM
- The two VMs remain in synchronization
  - Event logs of the primary are transmitted to the secondary
  - Received event logs are replayed by the secondary



The VM Fault Tolerance (VMFT) technique ensures that in the event of a VM failure,

there is no disruption in business operations. The VMFT technique ensures that individual VMs are functioning well and are responding to failures without interruption in service. VMFT creates hidden duplicate copies of each VM so that when it detects that a VM has failed due to hardware failure, the duplicate VM can be used for failover.

VMFT creates a live instance of the primary VM that runs on another physical machine. The two VMs are kept in synchronization with each other. The logs of the event executions by the primary VM are transmitted over a high speed network to the secondary VM. Secondary VM replays them to bring its own state same as of the primary VM. The hypervisor running on the primary server captures the sequence of events for the primary VM, including instructions from the virtual I/O devices, virtual NICs, user inputs, etc. And transfers them to the secondary server. The hypervisor running on the secondary server receives these event sequences and sends them to the secondary VM for execution. The primary and the secondary VMs share the same virtual disk using VMFS, but all output (for example, write) operations are performed only by the primary VM. A special locking mechanism ensures that the secondary VM does not perform write operations on the shared virtual disks. Any other output instructions from the secondary VM are also not allowed. The hypervisor posts all events to the secondary VM at the same execution point as they occurred on the primary VM. This way, the two VMs play exactly the same set of events and their states are synchronized with each other.

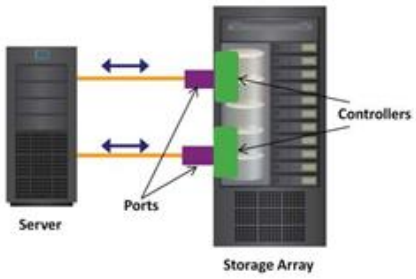


The two VMs essentially access a common storage and appear as a single entity with a single IP address and a single MAC address to other VMs. However, only the primary VM is allowed to perform writes. These two VMs constantly check the status of each other so that if the primary VM fails, the secondary VM immediately assumes the role of primary and starts executing writes on the shared storage. Another VM is also created to take the role of a new secondary VM, with the old secondary VM working as a new primary VM.

The process of checking the status of a VM requires an exchange of periodic signals between the primary and the secondary VMs over a reliable network. These periodic signals are known as heartbeats, and if there are specific communication channels to transfer these heartbeats, they form a heartbeat network.

Protecting Storage

- RAID
  - Provides data protection against drive failures
- Hot spare
  - Is a standby disk drive in a RAID array
  - Temporarily replaces a failed disk drive without disrupting the compute access
- Redundant components
  - Array controllers
  - Ports in a storage array
  - Storage arrays



The key techniques for protecting storage in a VDC are similar to the ones used in CDC, including:

- **RAID**: RAID is an enabling technology that leverages multiple drives as part of a set to provide data protection against drive failures. For example, RAID uses mirroring and parity techniques to rebuild the data after a disk drive fails.
- **Hot spare**: A hot spare refers to a spare disk drive in a RAID array that temporarily replaces a failed disk drive of a RAID set. A hot spare takes the identity of the failed disk drive in the array. When the failed disk drive is replaced with a new disk drive, either the hot spare replaces the new disk drive permanently, or the data from the hot spare is copied to it. The hot spare returns to its idle state and is used to replace the next failed drive. Multiple hot spares can be used to provide higher data availability.

In addition to RAID architecture and hot spares, high availability design for storage infrastructure is achieved by using:

- Redundant array controllers to address primary array controller failures
- Redundant ports in a storage array if one of the currently active port fails
- Redundant storage array when the whole array goes down

Protecting Network

- Protecting physical network by using redundancy
  - Interconnect devices with redundant hot swappable components
  - Redundant links and multipathing
  - Redundant NICs and NIC teaming
- Protecting virtual network
  - Failure in a virtual network on a hypervisor may be result of
    - Software error or security attack
    - Physical server failures
  - Virtual network failure is not common

The following are the techniques to protect external physical network:

- **Interconnect devices with redundant hot swappable components**: Allows the user to add or remove components while the interconnect device is operational.
- **Redundant links and multipathing**: A technique to enable multiple paths for accessing the same storage device for I/O operations so that in case of a failure of one path, dynamic failover to an alternate path is possible.

- **NIC teaming**: Grouping of two or more physical NICs into a single logical device.

Apart from the protection of external physical network, another important aspect is to protect the virtual network (running on a single server or hypervisor) from failure. A virtual network failure may be caused either by an incorrect operation of the software components (for example, virtual NIC, virtual Switch) or because of the failures in the compute infrastructure, for example, physical server going down, or hypervisor, or VM crashes. In general, failures due to software errors or security attacks require software patches from hypervisor vendors. The second reason is more often a cause for the failure in a virtual network. Therefore, the methods to protect compute infrastructure, equally apply for protecting such virtual network as well.

Protecting Network: Multipathing to Storage

- A technique to enable multiple paths for accessing the same storage device
  - Dynamic failover to an alternate path if current active path fails
- Multipathing requirement
  - Either a server has two or more HBAs available
  - Or, one HBA port is set up with multiple storage controllers
- Execution options
  - At hypervisor (in-built multipathing capability)
  - Third-party software module added to the hypervisor

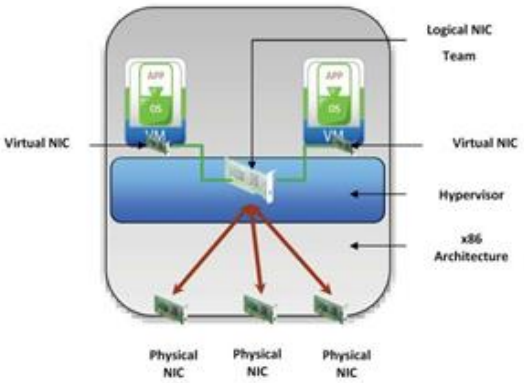
Loss of access to a storage device can cause serious disruption in service delivery. Multipathing is the technique to address such access failures to storage devices by enabling multiple alternative paths to the same storage device so that in case of failure of one path (for example, due to failures on switch, HBA port, storage processor, or cable) another alternative path can be activated and used for data transfer.

For multipathing, a server either has two or more HBAs available to reach the storage device using one or more switches, or the setup may have one HBA and multiple storage controllers so that the HBA can use different paths to reach the storage.

Multipathing is enabled either by using the hypervisor's built-in capability or by running a third-party software module, added to the hypervisor. In case of a path failure, I/Os are redirected to the surviving paths based upon a preconfigured path selection policy. In the case of SAN storage arrays, multipathing may also require adequate synchronization between the host software or hypervisor with the array controller. This ensures that all the paths are consistently visible across the LUNs on the storage array.

Protecting Network: NIC Teaming

- Enables failover in case of physical NIC failures/link outages
  - Supports the IEEE 802.1AX-2008 link aggregation standard
- VMs unaware of the underlying physical NICs
  - Packets sent to the logical NIC team are dispatched to one of the physical NICs
  - Packets arriving at any of the physical NICs are automatically directed to the appropriate virtual NIC

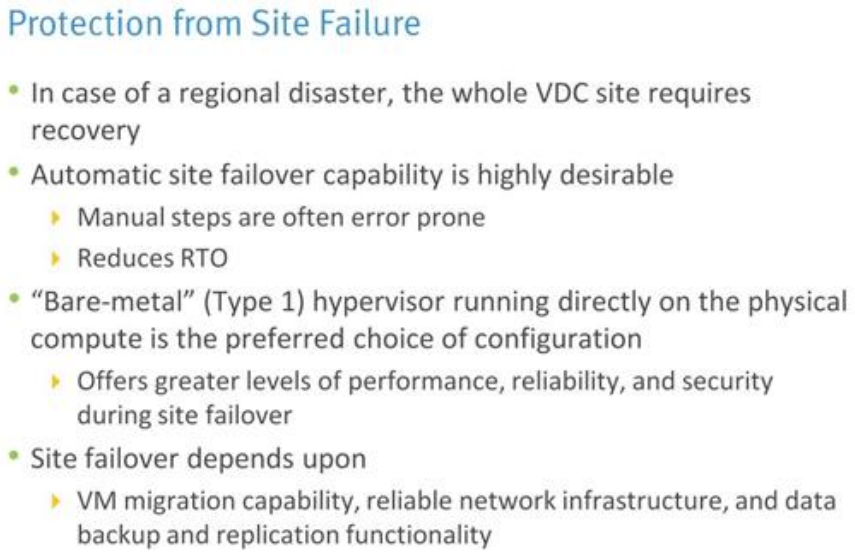


NIC teaming allows grouping of two or more physical NICs—which may be of different

types—into a single logical device called the NIC team. After an NIC team is configured, the VM will not be aware of the underlying physical NICs. Packets sent to the NIC team are dispatched to one of the physical NICs in the group. Packets arriving at any of the physical NICs are automatically directed to the NIC team, which consequently redirects them to the designated virtual NIC.

Apart from load balancing, which was discussed earlier, NIC teaming enables fault tolerance such that if one of the underlying physical NICs fails or its cable is unplugged, the traffic is redirected to another physical NIC in the team. Thus, NIC teaming eliminates the SPOF associated with a single physical NIC. Such failover remains totally transparent to the VMs and applications may experience performance degradation during this process.

In a VDC environment, with the hypervisor's support for NIC teaming, a guest OS or a VM need not install separate drivers for NICs or configure NIC teaming. NIC teaming implementations in most of the hypervisors support the IEEE 802.1AX-2008 link aggregation standard.



In case of a regional disaster, the whole VDC site requires recovery. To ensure error-free execution of BC solutions during a disaster, a non disruptive testing of Disaster Recovery (DR) plan is often required to be carried out beforehand.

To ensure a robust and consistent failover in case of a site failure or during testing, automatic site failover capabilities are highly desirable. This is because manual steps are often error prone. RTO with automated failover is significantly less compared to the manual process. However, the DR product that automates setup, failover, and testing of DR plans must be compatible with the guest OS.

A hypervisor provides robust, reliable, and secure virtualization platform that isolates applications and OSs from their underlying hardware. This considerably reduces the complexity of implementing and testing DR strategies. Among the different types of hypervisors for BC, the "bare-metal" (Type 1) hypervisor provides a robust DR. During the DR process, the bare-metal approach offers greater levels of performance, reliability, and security; and is better equipped to leverage the power of x86 server architectures found in the datacenters.

A VDC site failover process also depends upon other capabilities, including VM replication and migration capabilities, reliable network infrastructure between primary (production) site and the secondary (recovery or DR) site, and data backup capabilities.

Module 7: Business Continuity in VDC

Lesson 2: Backup in VDC

Topics covered in this lesson:

- Traditional approaches to VM backup
- Image based VM backup
- Challenges for VM backup in a VDC environment
- Deduplication as a mechanism for backup optimization
- Restoring a VM using backup copy

This lesson covers the challenges and the mechanisms for VM backup in a VDC environment. The lesson includes traditional and image based methods for VM backup, deduplication as a mechanism for backup optimization, and the process of restoring a VM using a backup copy.

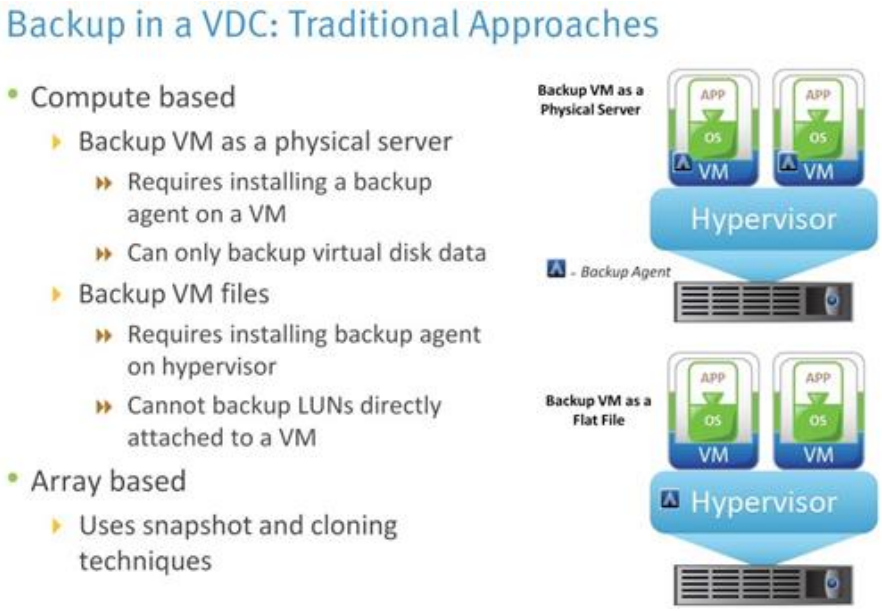
Backup in a VDC: An Overview

- VM backup includes
  - Virtual disks containing system and application data
  - Configuration data including network and power state
- Backup options
  - File based
  - Image based
- Backup optimization
  - Deduplication

A backup operation in a VDC environment often requires backing up the VM state. The VM state includes the state of its virtual disks, memory (i.e. RAM), network configuration, as well as the power state (on, off, or suspended). A virtual disk includes all the information typically backed up (OS, applications, and data.) As with physical machines, a VM backup needs to be periodically updated with respect to its source, in order to recover from human or technical errors.

A VM backup may be performed either as a "set of files" retaining the directory structure or as a complete image. File-level backup provides a way to access individual files, whereas image level backup does not. Another important difference between file based and image based backup is that an image based backup is independent of the guest OS running on the VM, whereas file based backup may have guest OS dependency, owing to file system structure. In the case of an image based backup, recovery of individual files requires additional operations to be performed**.**

Owing to the increased capacity requirements in a VDC environment, backup optimization methods are necessary. Deduplication, which aims to reduce duplication of backup data, is one of the important techniques towards achieving optimum backup.



Traditional approaches for backup in a VDC environment use the same technologies as are used in a CDC, with minor modifications. These approaches often treat a VM as a set of files while creating its backup. The approaches are as following:

- **First approach**: A VM is treated as if it were a physical server. A backup agent is installed on the VM that streams the VM data to the backup server. Management of the backup operation is similar to that of CDC, and hence, creating application consistent backups is easy. It is also possible to restore specific files to the guest OS. However, this solution requires the management of agents on all VMs. Note that the solution does not capture the VM files (i.e., Virtual BIOS file, VM swap file, Virtual disk file, Log file, and Configuration file). So, in the event of a restore, a user should manually recreate the VM and then restore data into it.
- **Second approach**: Since a VM's virtual disks are nothing but a collection of files, it is possible to simply back them up by performing a file system backup from a hypervisor. This is a relatively simple method because it requires only an agent on the hypervisor, and that backs up all the VM files. However, LUNs assigned directly to a VM (using RDM) cannot be backed up using this approach.
- **Third approach**: It uses array based technologies to backup a VM and utilizes the storage subsystem for backup operations. These technologies often use snapshot and cloning techniques to capture and backup a VM.

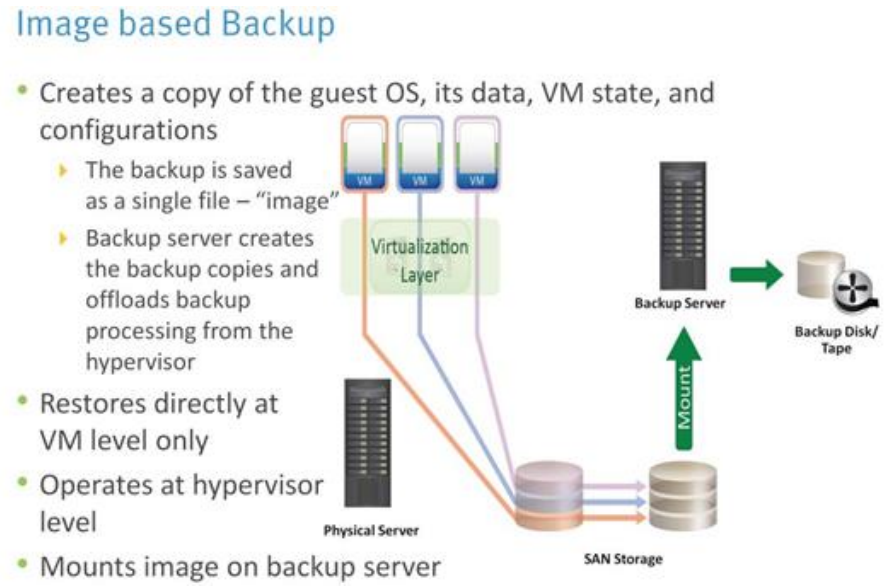


Image based backup creates a copy of the guest OS and all the data associated with it,

including the VM state (snapshot of VM disk files) and application configurations. The backup is saved as a single file called an "image". Image based backup operates at the hypervisor level and essentially takes a snapshot of a VM and then mounts the files on the backup server. The backup server consequently creates the backup copies to disks or tapes. This effectively offloads the backup processing from the hypervisor and transfers the load on the backup server, thereby reducing the impact to VMs and applications running on the hypervisor.

The advantage of an image based backup is that all information can be collected in a single pass, providing a Bare Metal Recovery (BMR) capability. Image based backup can even be done online. It can also be restored to dissimilar hardware resources and can recover servers remotely. Image based backup allows for fast and complete restoration of a VM. However, if the user has to recover few corrupted files, then the full VM image needs to be restored first, followed by restoration of individual files using O/S features.

Backup Considerations in a VDC

- Reduced computing resources
  - Existence of multiple VMs running on the same physical machine leaves fewer resources available for backup process
- Complex VM configurations
  - A backup agent running on VM has no access to VM configuration files
  - Not possible for a backup agent running on hypervisor level to access storage directly attached to a VM using RDM

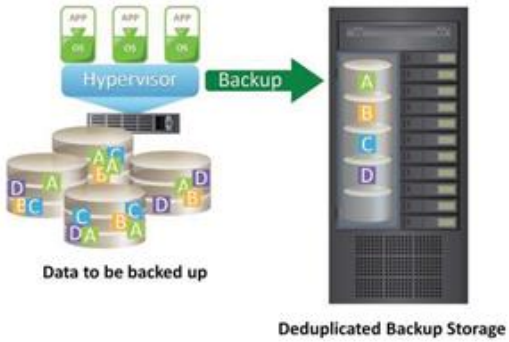
In a VDC environment, a single physical machine can run multiple VMs, that increases server utilization. This results into reduced compute resources available for backup and recovery tasks.

Apart from the increased capacity requirements and reduced computing resources, backing up a VM is more complex than backing up a physical machine. With a physical machine, the OS manages all the files on the machine. In a virtual environment, the guest OS is encapsulated in a VM that is managed by the hypervisor. Because of this architecture, there are files related to a VM; however, the VM does not have any direct access to these files; for example, the VM configuration files. In the case of RDM, where storage is directly attached to a VM, an application running on the hypervisor level will have difficulty while accessing that LUN.

These factors render backup processing more challenging in a VDC as compared to CDC.

Backup Optimization: Deduplication

- Backup images of VM disk files are candidates for deduplication in a VDC environment
- Deduplication types and methods are same as those employed in CDC

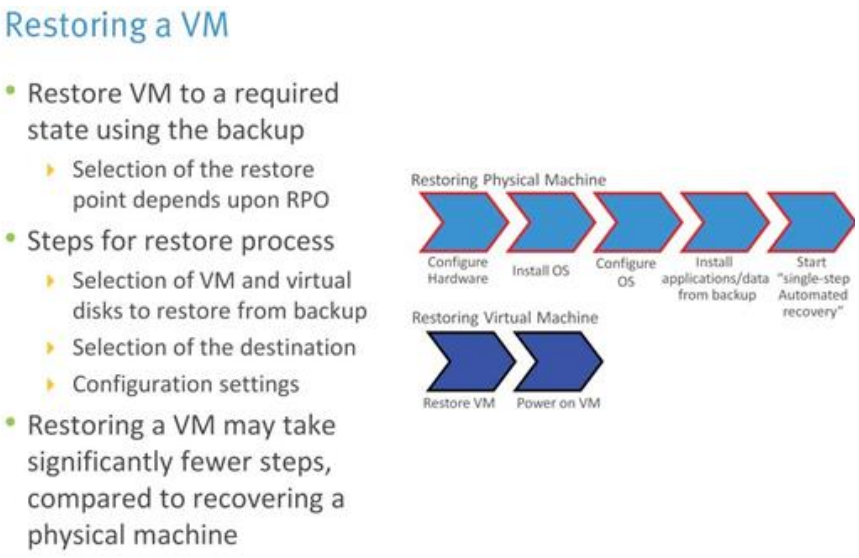


Virtualization benefits in a data center, for example, server consolidation ,often come at the cost of additional requirements for storage and backup resources. Therefore, to fully realize

benefits of virtualization, storage and network-optimized backup methods are necessary. Data deduplication, which aims to reduce duplication of backup data is one of the important techniques towards achieving the optimized backup.

Deduplication significantly reduces the amount of data to be backed up in a VDC environment. The potential sources for duplication are present when there are many similar VMs deployed in a data center. They include similar images of VM disk files, including configuration files and application data.

Deduplication types and methods to be applied in a VDC environment remain same as in CDC practice.

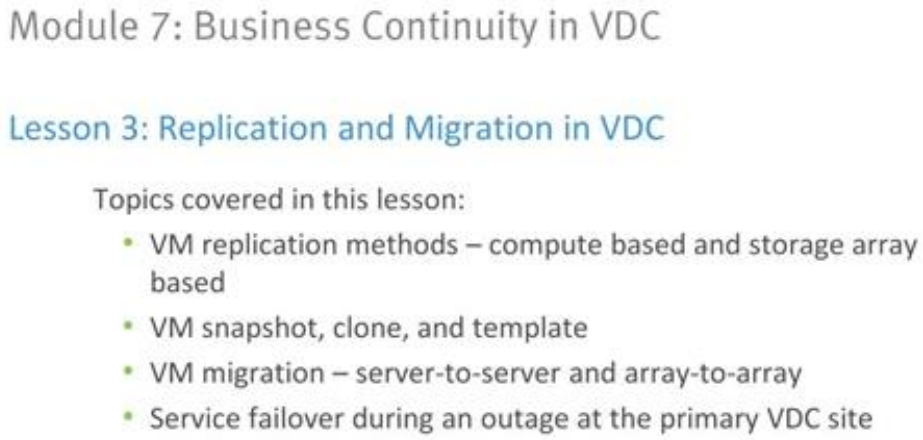


A restore process returns a VM to a selected previous state using the backup copies. Selection of the restore point actually depends upon the choice of the RPO.

The key steps involved in the restore process include:

1. Selection of the VM and virtual disks to be restored from the backup.
2. Selection of the destination location for the restoration: In case of an actual VM failure, the original physical machine itself is selected as a destination, whereas for restore rehearsal purposes, some alternate physical machine is selected.
3. Configuration settings: This includes deciding VM configuration settings either by using the existing backup or by applying new settings.

Restoring a VM may take significantly fewer steps, compared to the recovery of physical machines. Therefore, recovery time requirements for a VM and consequently for the whole VDC are relatively shorter, compared to the CDC environment.



This lesson covers challenges and techniques for VM replication and migration in a VDC environment. It includes compute based and storage array based VM replication methods

and discusses VM snapshot, clone, and template. It further describes migration of a VM across servers and across storage arrays. Finally, it covers service failover during an outage at the primary VDC site.

VM Replication in VDC

- VM replication is a critical requirement for successful BC
- VM replication is performed at the hypervisor level
  - Relies upon replication software to propagate changes made to a VM's virtual disk
- To ensure data integrity, *quiescing* of VM is necessary before replication process starts
  - Quiescing pauses currently running applications within a VM and forcibly flushes all data in the memory to the disk
  - To achieve application-level consistency, is performed at the application level
    - Applications complete any pending transactions and write the pending data to the disk

VM replication is a critical requirement for successful BC and DR processes in a VDC. This is because, in a VDC environment, to restart operations, a user needs to primarily restart VMs on the secondary (backup) site. VM replication makes this possible by making the copies of these VMs available at the backup site. VM replication is performed at the hypervisor level and relies on replication software that can copy all the changes made to a VM disk to another server. VM replication requires a secondary site which is up and a network connectivity linking these sites.

A virtual disk snapshot taken at the hypervisor level temporarily redirects incoming writes to a separate delta file. The delta file maintains these I/Os to be written to the virtual disk after the snapshot is taken. The virtual disk is then mounted by the replication software and any updates since the last replication cycle are copied to another identical virtual disk on a VM at the secondary site.

To ensure data integrity, *quiescing* of VM is necessary before the replication process starts. Quiescing pauses currently running applications within a VM and forcibly flushes all data in the memory to the disk. To ensure application-level consistency, applications (for example, Microsoft Exchange or SQL Server) are instructed to complete any pending transactions and write the pending data to the disk.

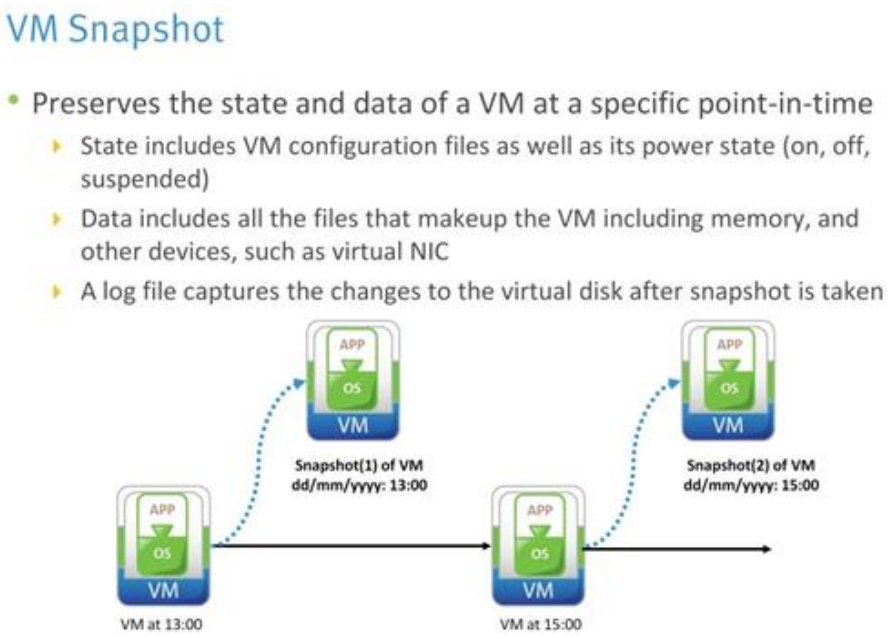
VM Replication Methods

- Compute based replication
  - Enables replicating VMs between dissimilar storage; for example, from a local disk drive to a storage array in a different location
  - Creates VM snapshot, VM clone, or VM template
- Storage array based replication
  - Is performed at the storage array level
  - Is similar to traditional array based replication in CDC
  - Is performed either at local (primary) site or to a remote (secondary) site

There are two main methods to replicate VMs. The first approach uses compute based replication. In this method, replication happens between similar/dissimilar storage devices. For example, using compute based replication, a VM is replicated from a local disk drive to a storage array in a different location. Compute based replication creates either a VM snapshot,

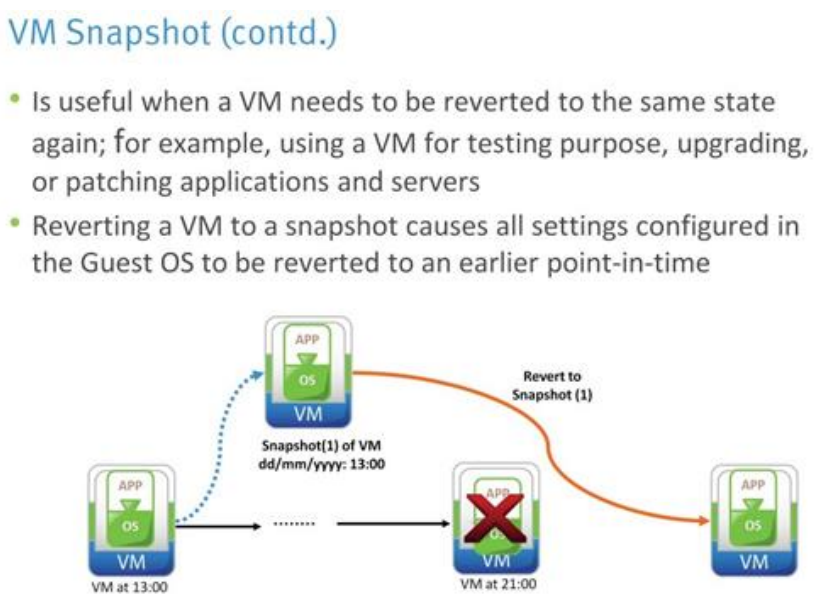
or a VM clone, or a VM template.

Another method is to use storage array to replicate the VM either to an array at the primary site itself or to a remote array (for example, iSCSI, or Fibre channel, or NAS storage array) at the secondary site. This method is similar to the traditional array based replication method used in a CDC. It works by copying LUNs of the source VM to the target array. Replication to a remote array may be either synchronous or asynchronous.



A snapshot preserves both the state and data of a VM at a specific point in time. State includes VM files such as BIOS, network configuration, and its power state (powered-on, powered-off, or suspended). Data includes all the files that makeup the VM, including virtual disks, memory, and other devices. In this method, instead of making a copy of the entire virtual disk of the VM, which may be very large in size, a snapshot is created. A snapshot includes a separate log file which captures the changes to the virtual disk of the VM since the snapshot was taken.

Referring to the diagram on this slide, the first snapshot of the VM is taken at time-point 13:00 (on date dd/mm/yyyy) and the second snapshot is taken after two hours at time- point 15:00. These snapshots are saved with unique names and details of the points-in- time, when they are created.



Snapshots are useful when a VM needs to be reverted to the same state repeatedly; for example, when using a VM for testing purposes or while upgrading or patching applications

and servers.

Reverting a VM to a snapshot causes all settings configured in the guest OS to be reverted to that time-point in past when that snapshot was created. The configuration which is reverted includes previous IP addresses and guest OS patch versions. The log file of the snapshot is used to revert the virtual disk to its original state.

Referring to the diagram on this slide, Snapshot(1) of the VM is created at time-point 13:00. Later, after a couple of hours, at time-point 21:00, the VM crashes and needs to be reverted. Snapshot(1) is selected for that purpose and VM is reverted to Snapshot(1) and starts running in the same way as it was running at time-point 13:00, when Snapshot(1) was created.

VM Clone

- An identical copy of an existing VM
  - Created when the VM is required for a different use; for example, identical VM need to be deployed for testing purpose
  - The existing VM is called the parent
  - Changes made to a clone VM do not affect the parent VM
  - Changes made to the parent VM do not appear in a clone
  - A clone may share virtual disks with the parent VM
- Clone VM is assigned separate network identity
  - Clone has its own separate MAC address

A clone is an identical copy of an existing VM. The existing VM is called parent of the clone VM. When the cloning operation is complete, the clone becomes a separate VM with its own MAC address, although it may share virtual disks with the parent VM in case of a Linked clone (discussed next). However, changes made to a clone VM do not affect the parent VM. Changes made to the parent VM do not appear in a clone. To other VMs in the network, a clone VM appears different from the parent VM.

Installing a guest OS and applications on multiple VMs is a time consuming task. With clones, a user can make many copies of a VM from a single installation and configuration process. Clones are useful when required to deploy many identical VMs. For example, deployment of standardized operating and/or application environment across a group or an organization.

VM Clone – Types

- Full Clone
  - An independent copy of a VM that does not share virtual disks with the parent
    - Cloning process may take relatively longer time
- Linked Clone
  - Shares virtual disk with the parent VM
    - Conserves disk space
    - Virtual disk of the parent VM is read-only for the linked clone
    - Writes of the linked clone are captured in a separate delta disk of smaller size
  - Is made from a snapshot of the parent VM
    - Snapshot is given a separate network identity and assigned to the hypervisor to run as an independent VM
    - Cloning process takes relatively less time

Two types of clones can be created from a VM:

- **Full Clone**: A full clone is an independent copy of a VM that shares nothing with the

parent VM. Full clone has no access or an ongoing connection with the parent VM. Because a full clone needs to have its own independent copy of the virtual disks, cloning process may take a relatively longer time. In relation to a snapshot, full clone duplicates the state of the parent VM only at the instant of the cloning operation and does not have access to any of the snapshots of the parent VM.

- **Linked Clone**: A linked clone is made from a snapshot of the parent VM. A snapshot is given a separate network identity and assigned to the hypervisor to run as an independent VM. However, all files available on the parent at the moment of the snapshot creation continue to remain available to the linked clone VM in a read-only mode. This conserves disk space and allows multiple VMs to use the same software installation.

Ongoing changes (writes) to the virtual disk of the parent do not affect the linked clone and changes to the virtual disk of the linked clone do not affect the parent. All the writes by the linked clone are captured in a separate file called delta disk, which is often of a much smaller size, compared to a full virtual disk.

Even though a linked clone is made by using a snapshot, it differs from a snapshot in the sense that a linked clone is a running VM that would change its state over time. However, a snapshot is only a state of the VM at a specific point-in-time, which cannot change on its own.

Linked clones can be very advantageous for collaborative network environments. For example, a support team can reproduce a bug in a VM running an application and an engineer can quickly make a linked clone of that VM to work on the bug simultaneously.

A VM template is a reusable image created from a VM. The VM template works like a master copy of a VM to be used for creating and provisioning new copies of the VM. The template typically includes virtual hardware components, an installed guest OS (with any applicable patches), and software applications.

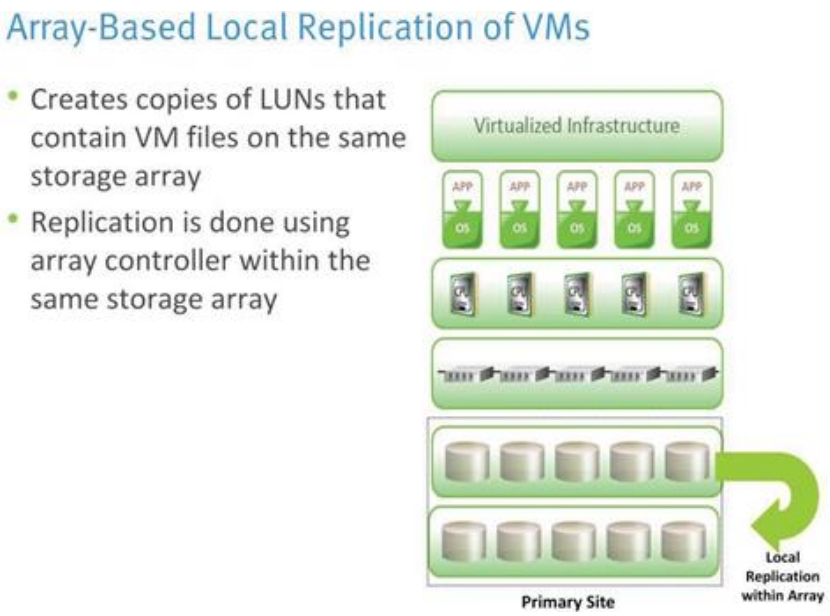
Templates can be created in two ways, either by converting a powered-off VM into a template or by cloning a VM to a template. The advantage of converting a VM to a template is that the conversion is instantaneous. However, the template will not be used immediately. Cloning a VM to a template leaves the original VM intact, but will require waiting for the entire capacity of the VM to be duplicated into the template's files.

A template differs from a clone because new VMs are deployed from templates. From a clone, only additional cloning can be done to create new VMs. However the new VMs which are created out of a Clone will reflect the changed state of the clone as compared to a VM template which preserves the original state.

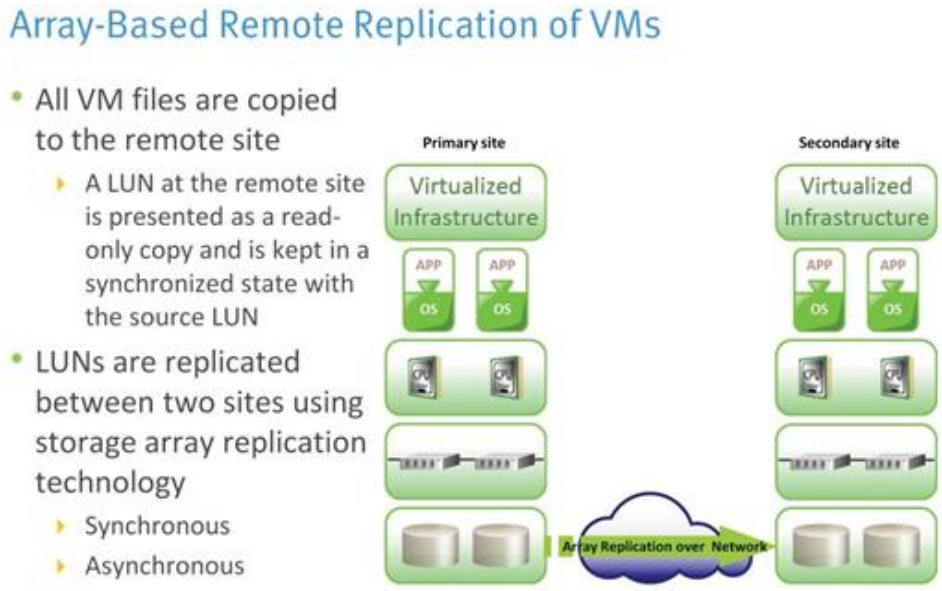
Key advantages of the VM template include the following:

- VM templates increase efficiency; for example, with templates, many repetitive installation and configuration tasks can be avoided.
- VM templates are also used to enforce consistency and standards. Deploying from

templates helps to enforce standards such as including antivirus and management software in any machine connected to the network.



VMs typically reside on LUNs that reside on the same storage arrays. In single-site replication, these LUNs are replicated using the array controller within the same storage array.



In a remote replication, all VM data residing on storage array (LUNs) at the primary (production) site is copied to the remote secondary (failover) site where it is kept ready to be used by the VMs, when required. This copying process remains transparent to the VMs, and they remain unaware of the underlying details. The LUNs are replicated between the two-sites using the storage array replication technology. This replication process can be either synchronous (limited distance, near zero RPO) or asynchronous (extended distance, non zero RPO). During business-as-usual periods, a LUN at the recovery site is presented as a read-only copy to any VMs connected to it at that site. The LUN is read-only at this stage as it is kept in a synchronized state with its source LUN.

VM Migration

- Moving a VM from one hypervisor to another hypervisor
  - For example: hypervisor failure or scheduled hypervisor or storage array maintenance
- Types: Server-to-Server and Array-to-array
- Migration process
  - Involves movement of entire active state of a VM
  - In case of server-to-server migration, virtual disks are not moved within clustered servers
    - For remote migration, virtual disks are also moved
    - In case of array-to-array migration, virtual disks are always moved from source array to target array
  - VM in the source hypervisor needs to be deleted after migration is complete
    - Virtual disks are also deleted if they were actually moved

There are many situations where moving a VM from one hypervisor to another is necessary. The primary purpose is to achieve BC in case of hypervisor failure. Another situation might involve load balancing when one hypervisor is running many VMs but another hypervisor is relatively less occupied. Yet another reason might involve facilitating scheduled hypervisor or storage array maintenance.

The process of moving VMs from one (source) hypervisor to another (target) is known as VM Migration. There are primarily two different types of VM migrations:

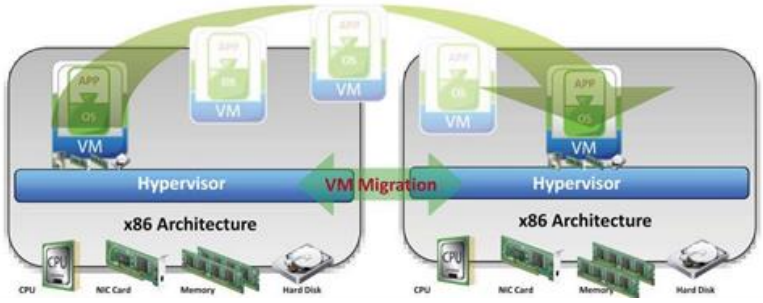
- **Server-to-server migration**: This type of VM migration involves moving a VM from one hypervisor to another using a client (migration agent software) running on these hypervisors. The receiving hypervisor may be remotely located in a different data center.
- **Array-to-array migration**: This type of VM migration involves moving VMs or virtual disks across multiple physical servers using storage array based methods.

VM migration involves moving the entire active state of a VM from the source hypervisor to the target. The state information includes memory contents and all other information which identifies the VM. VM identification information includes data, which maps the VM hardware elements such as BIOS, Devices, CPU, and MAC address for Ethernet cards.

When a VM is migrated to another hypervisor on a clustered server, virtual disks of the VM are not migrated because these can be accessed from another hypervisor as well. However, in situations where a VM needs to be migrated to a hypervisor on a remote server, virtual disks are also moved. In case of array-to-array migration, virtual disks are always moved from the source array to the target array. After migration, VM in the source hypervisor is deleted. Virtual disks are also deleted if they were actually moved. For example, in the case of server-to-server remote migration and array-to-array based migration, virtual disks are deleted after their copies are saved at the target.

VM Migration: Server-to-Server

- **Hot-On** – Migrate a VM that is powered on
  - VM needs to be quiesced before migration
- **Cold** – Migrate a VM that is powered off
- **Hot-Suspended** – Migrate a VM that is suspended
- **Concurrent** – Migrate multiple VMs simultaneously

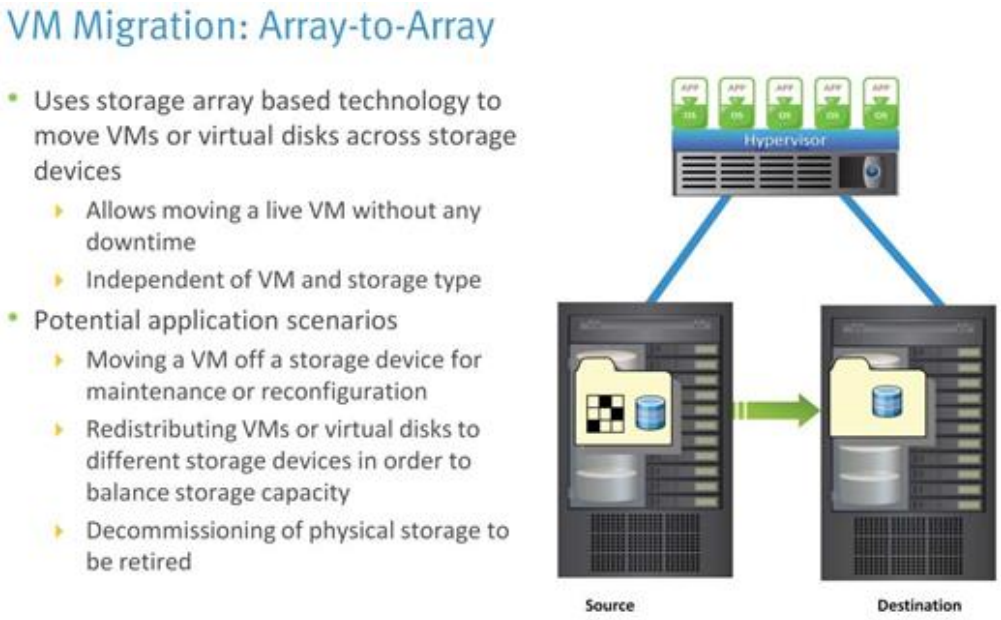


There are primarily four major modes of VM migration between servers:

- **Hot-On migration:** VM remains in a powered-on state (currently running). Shared storage and CPU compatibility might be necessary between the source and destination servers. VM also needs to be quiesced during migration so that the state of the VM after migration remains consistent. After migration, client accesses are directed to the VM on the target hypervisor. Hot on migration is useful in scenarios where a server or hypervisor is overloaded or requires maintenance and/or repair soon because it might be underperforming.
- **Cold migration:** VM is in a power off state. Cold migration is typically used when a VM

needs to be moved to another remote location or VDC.

- **Hot-Suspended migration:** VM is in a suspended state. In a suspended state, all virtual machine activities are paused until the resume command is issued. This kind of migration could be useful when a VM needs to be migrated from a failing server to another operational server.
- **Concurrent migration:** Multiple VMs are migrated simultaneously to one or more hypervisors. The maximum number of concurrent sessions is generally dependent upon the available network bandwidth. Concurrent migration can be performed when VM is on a power on/off/suspended state. Concurrent migrations are useful for continuously optimizing VM placement across the entire IT environment.



Array-to-array based VM migration uses storage array based technology to move VMs or virtual disks across storage devices. This allows moving a VM off a storage device for maintenance or reconfiguration without VM downtime. By using this approach, it is possible for storage administrators to move VMs across dissimilar storage types. For example, VM migration can be performed across storage types such as Fibre Channel, iSCSI, and local SCSI.

Potential applications of array-to-array VM migration include:

- Redistributing VMs or virtual disks to different storage devices in order to balance capacity and to improve performance.
- Decommissioning physical storage that is soon to be retired, for example, storage arrays whose maintenance and release cycles are coming to an end.

Service Failover

- Steps performed during an outage at the primary VDC site
  - Replicated VMs are activated in the secondary (failover) site on the target servers
  - The LUN at primary site is made read-only and LUN at the secondary site is made write-enabled
- Secondary site can have different x86 hardware configuration than those at the primary site
- Failover challenges
  - Constraints on VM placement, for example, if selected VMs have to be placed on the same server/clustered servers

At the simplest level, VMs run Guest OS, which consequently runs applications that provide IT services to the businesses. During an outage at the primary site, it is this service that must be protected from a BC point of view.

VMs are isolated from one another and are not dependent on the underlying x86 hardware configurations available at the failover location. The in-built properties of VMs allow mapping of resources from one virtual environment to another and at the same time, enable failover of an IT service. In addition, storage is replicated to the failover site, and then presented to the virtual infrastructure at the failover location where the VMs may be activated. When there is an outage at the production site, the source LUN is made read-only and the target is made write-enabled. Applications resume operations at the secondary site.

There exist certain challenges, which might prevent complete and/or correct failover of VMs in case of an outage at the production site. For example, there may exist VM placement rules specifying that selected VMs should be placed on the same server/clustered servers. A scenario in which such constraints might be present, is when a group of VMs are communicating with one another heavily, because of which they required to be placed on the same server or within clustered servers. These constraints might prevent a VM to failover completely and/or correctly to a different server at the failover site if all of the specifies constraints are not satisfied at the failover site.

UNIT-5

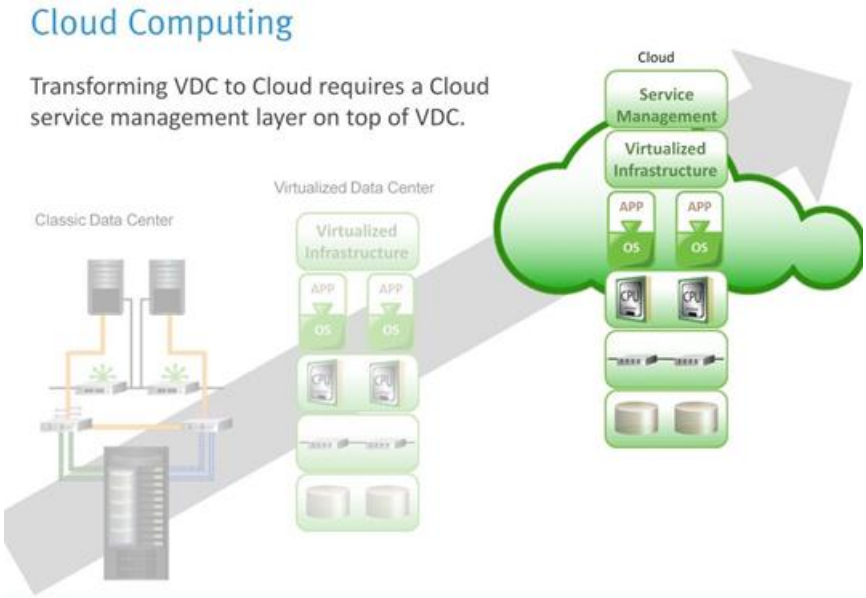
Cloud Computing and Infrastructure

Module 8: Cloud Computing Primer

Upon completion of this module, you should be able to:

- Explain the essential characteristics of Cloud Computing
- Describe the different Cloud services models
- Describe the different Cloud deployment models
- Explain the economics of Cloud
- Discuss the benefits and challenges of Cloud

This module focuses on the essential characteristics of Cloud Computing, the different Cloud services and deployment models, and the economics of Cloud.



Virtual Data Center (VDC) provides flexibility, improved resource utilization, and ease of management over CDC. The continuous cost pressure on IT, growth of information, on-demand data processing, and agility requirements of a business gave way to the emergence of a new IT model – "Cloud Computing". Building a Cloud infrastructure needs a layer of Cloud service management on the top of the virtualized infrastructure (VDC). Before discussing Cloud infrastructure and service management, it is important to understand Cloud, its characteristics, benefits, services, and deployment models. This and the following modules cover the fundamentals, service management, migration strategy, and security aspects of Cloud Computing.

Lesson 1: Overview of Cloud Computing

Topics covered in this lesson:

- Technological foundations of Cloud Computing
- Essential characteristics of Cloud Computing
- Benefits of Cloud Computing

This lesson covers various business drivers, essential characteristics, and key benefits of Cloud Computing.

Cloud Computing: Technological Foundations

Technologies	Description
Grid Computing	• Form of distributed computing which applies the resources of numerous computers in a network to work on a single complex task at the same time
Utility Computing	• Service provisioning model that offers computing resources as a metered service
Virtualization	• Provides improved utilization of resources • Enables optimization of resources by over subscription
Service Oriented Architecture (SOA)	• An architectural approach in which applications make use of services available in the network • Each service provides a specific function, for example, business function (Payroll Tax calculation)

Historically, Cloud Computing has evolved through grid computing, utility computing, virtualization, and service oriented architecture.

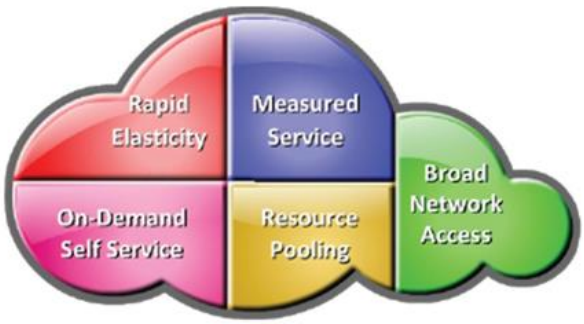
- **Grid computing**: It is a form of distributed computing which enables the resources of numerous heterogeneous computers in a network to work on a single complex task at the same time. Grid computing enables parallel computing, although its utility is best for large workloads.

- **Utility computing:** It is a service-provisioning model in which a service provider makes computing resources available to the customer, as required, and charges them for specific usage rather than charge a flat rate. The word 'utility' is used to make an analogy to other services, such as water, electrical power, etc., that seek to meet fluctuating customer needs, and charge for the resources based on usage, rather than on a flat-rate basis.

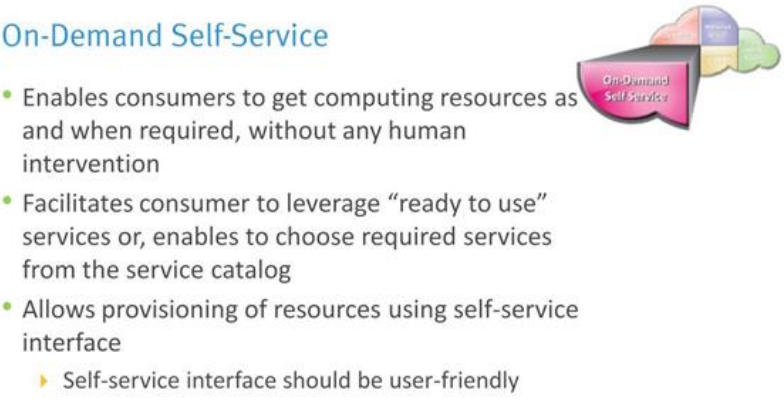
- **Virtualization:** The conversion of traditional computing environments to what is called a virtualized environment has also accelerated the movement to Cloud Computing. Virtualizing a computing environment means that the various hardware and the software resources are viewed and managed as a pool, which provides improved utilization of resources. The objectives of virtualization are to centralize management, optimize resources by over subscription, and use the available computing capacity as efficiently as possible among the users and applications.

- **Service Oriented Architecture (SOA):** An architectural approach in which applications make use of services available in the network. Each service provides a specific function, for example, a business function, such as payroll tax calculation or processing purchase order. A deployed SOA- based architecture provides a set of services that can be used in multiple business domains.

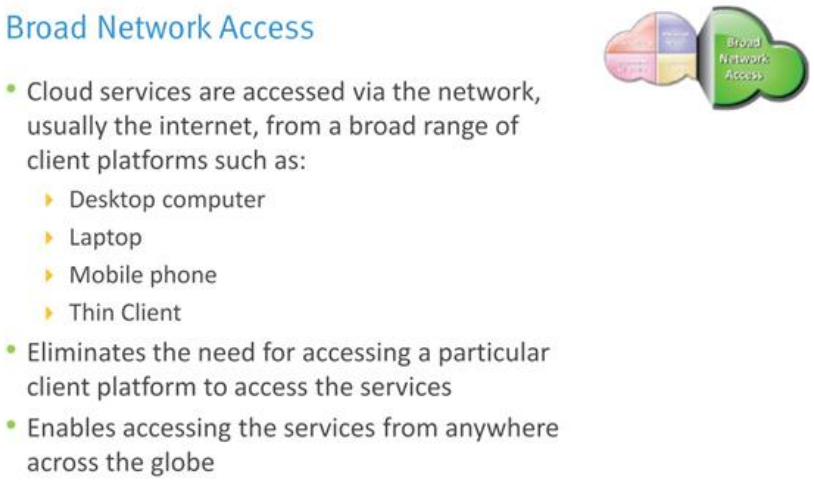
Cloud Computing: Essential Characteristics



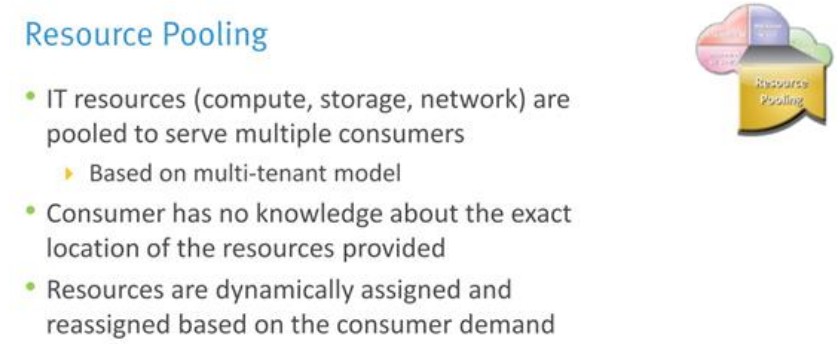
The NIST definition of Cloud Computing states that the Cloud infrastructure should essentially have five essential characteristics, which will be discussed in the following slides.



The on-demand and self-service aspects of Cloud Computing mean that a consumer can use Cloud services as required, without any human intervention with the Cloud service provider. By using the self-service interface, consumers can adopt Cloud services by requesting for the necessary IT resources from the service catalog. In order to be effective and acceptable to the consumer, the self-service interface must be user-friendly.



Cloud services are accessed via the network, usually the internet, from a broad range of client platforms, such as desktop computer, laptop, mobile phone, and thin client. Traditionally, software, such as Microsoft Word or Microsoft PowerPoint, have been offered as client-based software. Users have to install the software on their computers in order to use this software application. It is not possible to access this software if the user is away from the computer where the software is installed. Today, much of the software used can be accessed over the internet. For example, Google Docs, a Web-based document creator and editor allows users to access and edit documents from any device with an internet connection, eliminating the need to have access to a particular client platform to edit documents.



A Cloud must have a large and flexible resource pool to meet the consumer's needs, to provide the economies of scale, and to meet service-level requirements. The resources (compute, storage, and network) from the pool are dynamically assigned to multiple consumers based on a multi-tenant model. Multitenancy refers to an architecture and design by which multiple independent clients (tenants) are serviced using a single set of resources. In a Cloud, a client (tenant) could be a user, a user group, or an organization/company. Multitenancy enables compute,

storage, and network resources to be shared among multiple clients. Virtualization provides ways for enabling multitenancy in Cloud. For example, multiple VMs from different clients can run simultaneously on the same server with hypervisor support.

There is a sense of location independence, in that the consumer generally has no knowledge about the exact location of the resources provided .

Rapid Elasticity

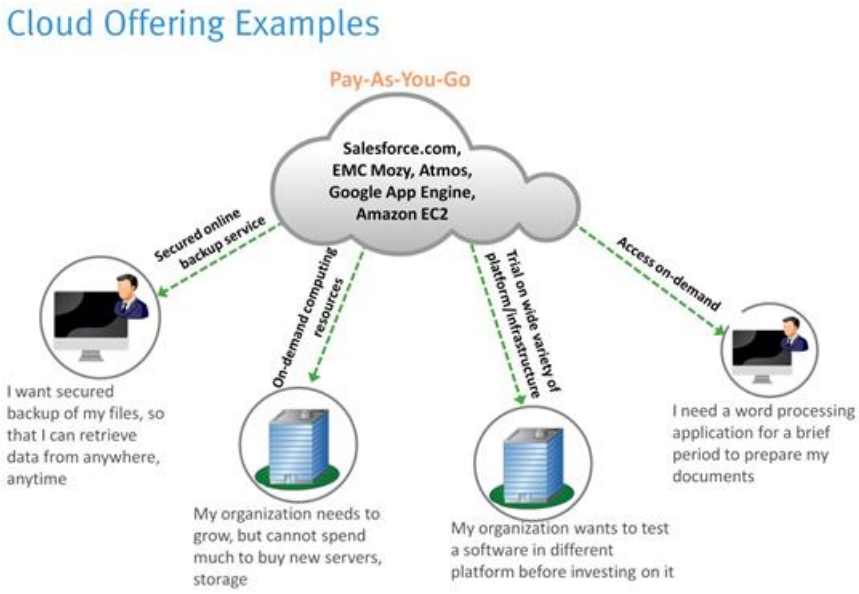
- Ability to scale IT resources rapidly, as required, to fulfill the changing needs without interruption of service
  - Resources can be both scaled up and scaled down dynamically
- To the consumer, the Cloud appears to be infinite
  - Consumers can start with minimal computing power and can expand their environment to any size

Rapid elasticity refers to the ability of the Cloud to expand or reduce allocated IT resources quickly and efficiently. This allocation might be done automatically without any service interruption. Consumers will take advantage of the Cloud when they have large fluctuation in their IT resource usage. For example, the organization may be required to double the number of Web and application servers for the entire duration of a specific task. They would not want to pay the capital expense of having dormant (idle) servers on the floor most of the time and also would want to release these server resources after the task is completed. The Cloud enables to grow and shrink these resources dynamically and allows the organizations to pay on a usage basis.

Metered Service

- Consumers are billed based on the metered usage of Cloud resources
  - Cost incurred on a pay-per-use basis
  - Pricing/billing model is tied up with the required service levels
- Resource usage is monitored and reported, which provides transparency for chargeback to both Cloud service provider and consumer about the utilized service

Metered service provides billing and chargeback information for the Cloud resource used by the consumer. The metered services continuously monitor resource usage (CPU time, bandwidth, storage capacity) and report the same to the consumer. Metered services enable transforming capital expenditure (CAPEX) into ‘pay as you use’ operational cost.



Organizations might need to rapidly expand their businesses, which may enforce them to enlarge the IT infrastructure by adding new servers, storage devices, network bandwidth, etc. Critical business data must be protected and should be available to the intended user, which, in turn, requires data security and disaster recovery infrastructure. As the capital expenditure rises to fulfill the requirements, the risk associated with the investment too increases. For small and medium size businesses, this may be a big challenge, which eventually restricts their business to grow. As an individual, it may not be sensible or affordable every time to purchase new applications if they are required only for a brief period. Instead of purchasing new resources, Cloud resources are hired based on pay-per-use without involving any capital expenditures.

Cloud service providers offer on-demand network access to configurable computing resources, such as servers, storage, network, and applications. Consumers (organization or individual) can scale up or scale down the demand of computing resources with minimal management effort or service provider interaction. Consumers can leverage Cloud service provider's expertise to store, protect, backup, and replicate data empowered by the most advanced technology, which otherwise would cost more.

Cloud Computing Benefits

Benefit	Description
Reduced IT Cost	• Avoids the up-front capital expenditure
Business agility support	• Provides the ability to add new resources quickly
Flexible scaling	• Scales up and down easily and instantly, based on demand
High availability	• Ensures application availability at varying levels, depending on policy and priority of the application
Less energy consumption	• Enables organizations to reduce power consumption and space usage

Reduced IT cost: Cloud services can be hired. Therefore, consumers can save money because there is no capital expenditure or CAPEX required. Consumers can leverage the Cloud service provider's infrastructure. Hence, there are no ongoing expenses for running a datacenter, such as the cost of power, cooling, and management. Additionally, the real estate cost can be minimized.

Business agility support: The speed at which a new computing capacity can be provisioned is a vital element of Cloud Computing. Cloud can reduce the time required to provision and deploy new applications and services from months to minutes. Cloud allows organizations to react more

quickly to market conditions and enables to scale up and scale down the resources, as required.

Flexible scaling: A Cloud can be easily and instantly scaled up and scaled down based on demand. It appears to the consumers that the Cloud resources are expandable to infinite limit. Cloud service users can independently and automatically scale their computing capabilities without any interaction with the Cloud service providers.

High Availability: Cloud computing has the ability to ensure application availability at varying levels, depending on customer policy and priority of the application. Redundant server, network resources, and storage equipment along with clustered software enable fault tolerance for Cloud infrastructure. These techniques encompass multiple datacenters in different geographic regions that have identical resource configuration and application instances. Hence, data unavailability due to regional failures is prevented.

Less Energy Consumption: ‘’Going Green” is an important focus for many organizations. Cloud enables organizations to reduce power consumption and space usage.

Module 8: Cloud Computing Primer

Lesson 2: Cloud Services and Deployment Models

Topics covered in this lesson:

- Cloud service models – SaaS, PaaS, and IaaS
- Cloud deployment models – Private, Public, Hybrid, and Community
- Economics of Cloud
- Challenges of Cloud

This lesson covers different Cloud services, deployment models, and economics of Cloud. It also covers the challenges of Cloud from the consumer and provider’s perspectives.

Cloud Service Models

Cloud Service can be classified into three categories:

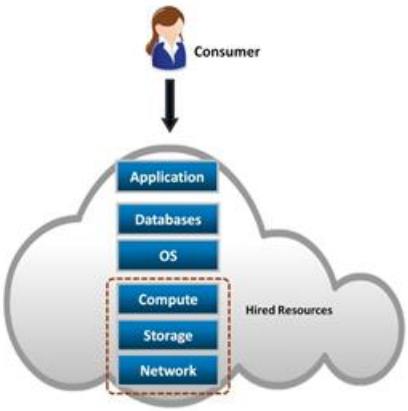
- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

Cloud service models can be classified into three categories:

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

Infrastructure-as-a-Service

- Provides capability to the consumer to hire infrastructure components such as servers, storage, and network
- Enables consumers to deploy and run software, including OS and applications
- Pays for infrastructure components usage, for example, Storage capacity, CPU usage, etc.



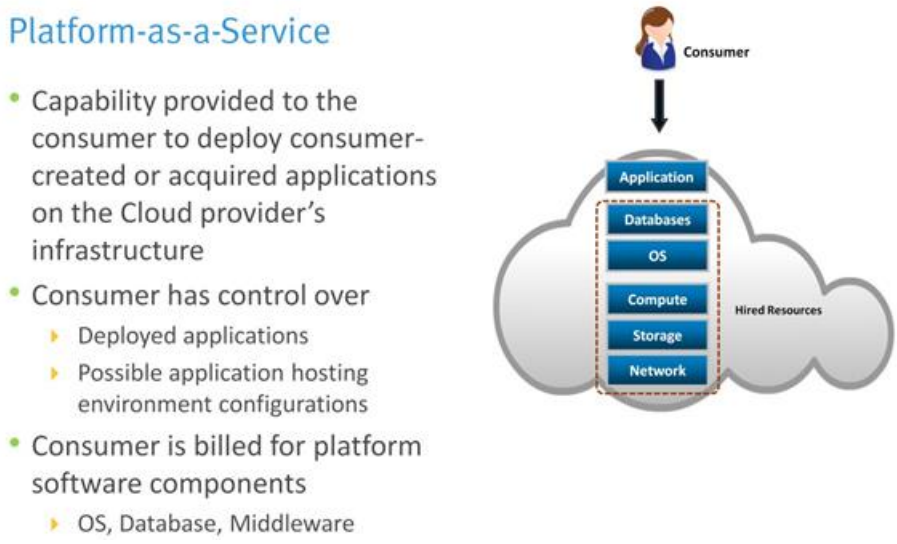
Infrastructure-as-a-Service (IaaS) is the base layer of the Cloud stack. It serves as the foundation for the other two layers (SaaS, PaaS) for their execution. The Cloud infrastructure such as servers, routers, storage, and other networking components are provided by the IaaS provider. The consumer hires these resources as a service based on needs and pays only for the usage.

The consumer is able to deploy and run any software, which may include Operating Systems (OSs) and applications. The consumer does not manage or control the underlying Cloud infrastructure, but has control over the Oss and deployed applications. Here, the consumer needs to know the resource requirements for the specific application to exploit IaaS well.

Scaling and elasticity are the responsibilities of the consumer, not the provider. In fact, IaaS is a mini do-it-yourself data center that you would need to configure the resources (server, storage) and to get the job done.

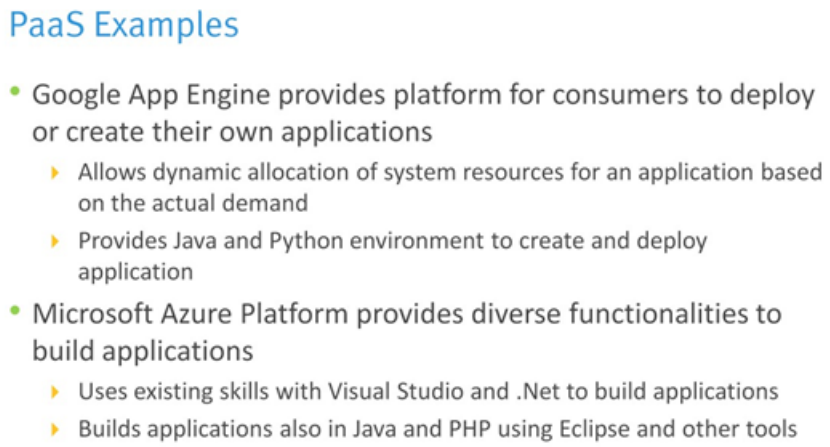
Until now, small consumers did not have the capital to acquire massive compute resources and to ensure that they had the capacity they needed to handle unexpected spikes in load. Amazon Elastic Compute Cloud (Amazon EC2) is an Infrastructure-as-a-Service model that provides scalable compute capacity, on demand, in the Cloud. It enables consumers to leverage Amazon's massive infrastructure with no up-front capital investment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing consumers to quickly scale capacity—both up and down—as their computing requirements change.

EMC Atmos is a globally accessible Cloud storage that provides solutions to meet the needs of enterprises and service providers. The 'Atmos Cloud Delivery Platform' provides IT departments and service providers a way to build, manage, and deliver self-service and robust metering for chargeback. Atmos, with its Web service access mechanism, is the ideal infrastructure for storage as a service.



Platform-as-a-Service is the capability provided to the consumer to deploy consumer-created or acquired applications on the Cloud infrastructure. PaaS can broadly be defined as application development environments offered as a 'service' by the Cloud provider. The consumer uses these platforms that typically have Integrated Development Environment (IDE), which includes editor, compiler, build, and deploy capabilities to develop their applications. They then deploy the applications on the infrastructure offered by the Cloud provider. When consumers write their applications to run over the PaaS provider's software platform, elasticity and scalability is guaranteed transparently by the PaaS platform. Here, the consumer does not manage or control the underlying Cloud infrastructure, such as network, servers, OSs, and storage, but controls the

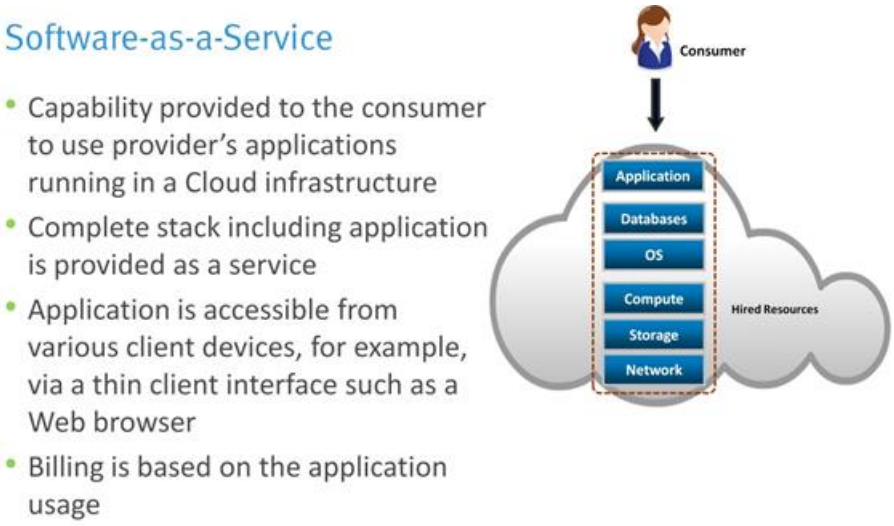
deployed applications and possibly the application-hosting environment configurations. For PaaS, consumers pay only for the platform software components such as databases, OS instances, and middleware, which includes its associated infrastructure cost.



Google App Engine is a Platform-as-a-Service that allows consumers to build Web applications using a set of APIs and to run those applications on Google's infrastructure. With App Engine, there are no servers to maintain. You just need to upload the application, and it is ready to serve. Google App Engine makes it easy to build an application that runs reliably, even under heavy load and with large amounts of data. Consumer's applications can run in Java or Python environments. Each environment provides standard protocols and common technologies for Web application development. The App engine Software Development Kits (SDKs) for Java and Python include a Web server application that emulates all of the App Engine services on the consumer's local computer. Each SDK includes all of the APIs and libraries available on App Engine. Each SDK also includes a tool to upload the consumer's application to App Engine. After the consumer has created the application's code and configuration files, the consumer can run the tool to upload the application.

Azure Platform is a Microsoft PaaS offering. Microsoft's Azure Platform supplies a broad range of functionalities to build, host, and scale applications in Microsoft data centers. Developers can use familiar tools, such as visual studio and .NET Framework to develop their application.

Windows Azure is a Cloud-based OS that enables and provides development, hosting, and service management environments for the Azure platform.



SaaS is the top most layer of the Cloud Computing stack, which is directly consumed by the end user. It is the capability, provided to the consumer, to use the service provider's applications running on a Cloud infrastructure. It is accessible from various client devices through a thin client interface such as a Web browser. On-premise applications are quite expensive and requires high upfront CAPEX (Capital Expenditure). They also incur significant administration costs. In a SaaS model, the applications such as Customer Relationship Management (CRM), Email, and Instant Messaging (IM) are offered as a 'service' by the Cloud provider. Here, the consumers will use

only the applications they really want and pay a subscription fee for the usage. The Cloud provider will host and manage the required infrastructure and applications to support these services.

SaaS offers the following advantages:

- Reduces the need for infrastructure because storage and compute powers can be provided remotely.
- Reduces the need for manual updates because SaaS providers can perform those tasks automatically.

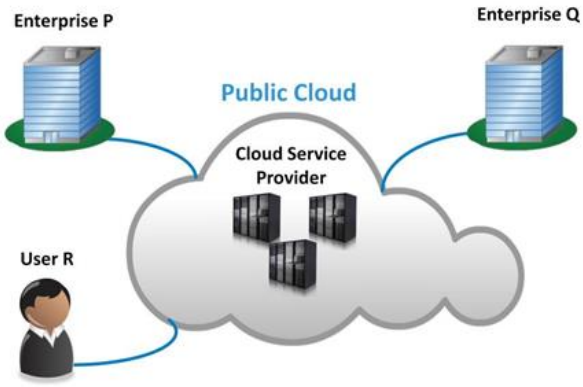
SaaS Examples

- EMC Mozy is a Software-as-a-Service solution for on-line backup
  - Consumers can leverage the Mozy console to perform automatic, secured, online backup and recovery of their data with ease
- Salesforce.com is a Software-as-a-Service solution for CRM application
  - Consumers can access CRM applications from anywhere, any time

EMC Mozy is a Software-as-a-Service solution, built on a highly scalable and available back-end storage architecture. Consumers can leverage the Mozy console to perform automatic, secured, online backup and recovery of their data with ease. EMC Mozy has two main products – MozyHome and MozyPro. MozyHome is for the individual consumer who is looking for a cost-effective way to backup all of their data, such as photos, music, and documents. MozyPro is dedicated to organizations looking for a cost-effective way to backup the end user's data. This low-cost software service is available at a monthly subscription fee. EMC Mozy does not require consumers to purchase any new hardware and requires minimal IT resources to manage.

Salesforce.com is a provider of SaaS-based CRM products. Organizations can use CRM applications to gain fast, easy access to the tools and services, required to build closer relationships with customers. The CRM applications run in the Cloud. They enable the consumer to access the application from anywhere through an Internet-enabled compute system. So, organizations need not buy the CRM applications and manage them on their own infrastructure. They subscribe for CRM applications from Salesforce.com.

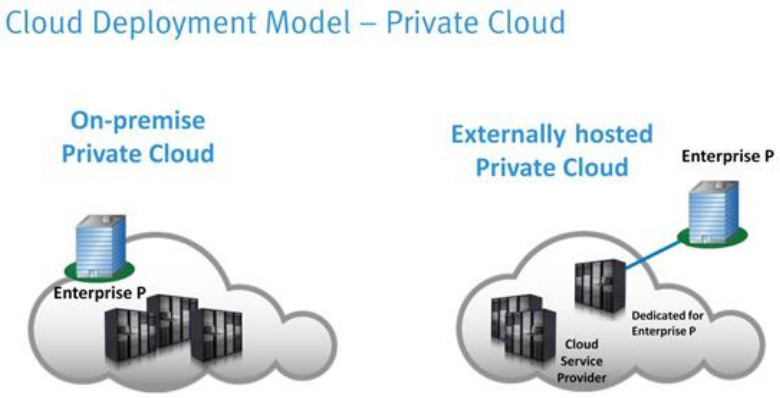
Cloud Deployment Model – Public Cloud



Cloud Computing can be classified into three deployment models: private, public, and hybrid. These models provide a basis for how Cloud infrastructures are constructed and consumed.

In a Public Cloud, IT resources are made available to the general public or organizations and are owned by the Cloud service provider. The Cloud services are accessible to everyone via standard Internet connections. In a public Cloud, a service provider makes IT resources, such as applications, storage capacity, or server compute cycles, available to any consumer. This model can be thought of as an "on-demand" and as a "pay-as-you-go" environment, where there are no on-site infrastructure or management requirements. However, for organizations, these benefits come with certain risks: no control over the resources in the cloud, the security of confidential data, network performance issues, and interoperability. Popular examples of public clouds include

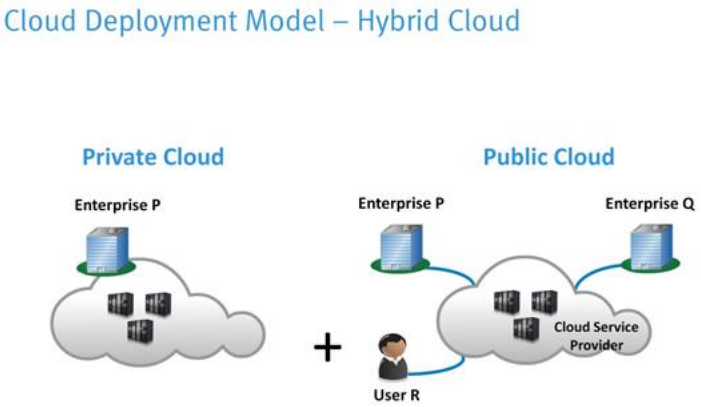
Amazon's Elastic Compute Cloud (EC2), Google Apps, and Salesforce.com.



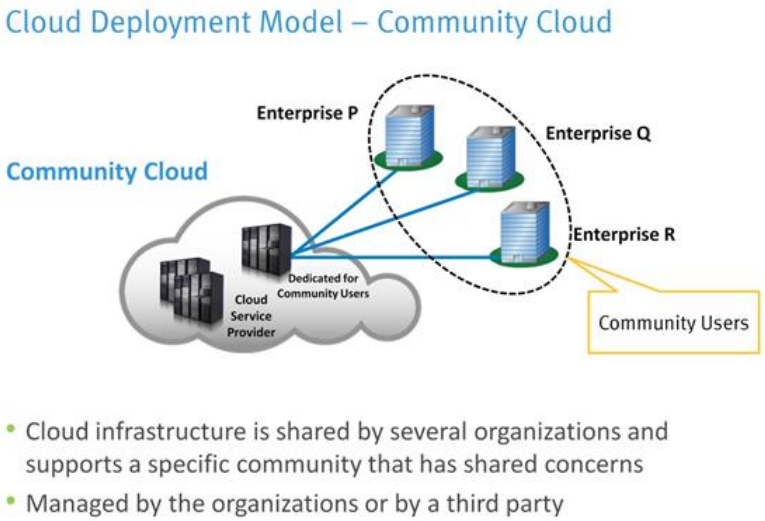
In a private Cloud, the Cloud infrastructure is operated solely for one organization and is not shared with other organizations. This Cloud model offers the greatest level of security and control. There are two variations to a private Cloud:

- **On-premise Private Cloud:** On-premise private Clouds, also known as internal Clouds, are hosted by an organization within their own data centers. This model provides a more standardized process and protection, but is limited in terms of size and scalability. Organizations would also need to incur the capital and operational costs for the physical resources. This is best suited for applications which require complete control and configurability of the infrastructure and security.
- **Externally-hosted Private Cloud:** This type of private Cloud is hosted externally with a Cloud provider, where the provider facilitates an exclusive Cloud environment for a specific organization with full guarantee of privacy or confidentiality. This is best suited for organizations that do not prefer a public Cloud due to data privacy/security concerns.

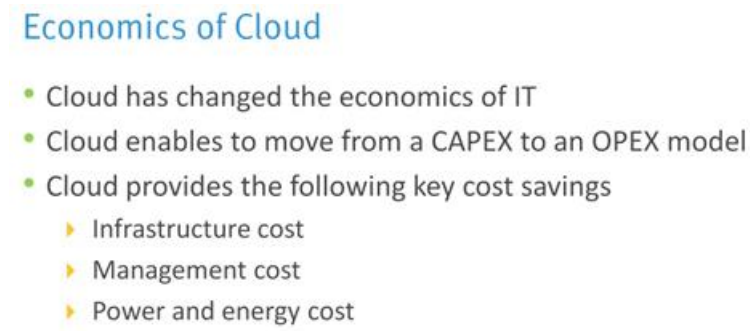
Like a public Cloud, a private Cloud also enables provisioning an automated service request rather than a manual task processed by IT. In on-premise private Cloud, organizations will have to run their own hardware, storage, networking, hypervisor, and Cloud software. Many enterprises, including EMC, Cisco, IBM, Microsoft, Oracle, and VMware, now offer Cloud platforms and services to build and manage a private Cloud.



In hybrid Cloud environment, the organization consumes resources from both private and public Clouds. The ability to augment a private Cloud with the resources of a public Cloud can be utilized to maintain service levels in the face of rapid workload fluctuations. Organizations use their computing resources on a private Cloud for normal usage, but access the public Cloud for high/peak load requirements. This ensures that a sudden increase in computing requirement is handled gracefully. For example, an organization might use a public Cloud service, such as Amazon Simple Storage Service (Amazon S3), for archiving data, but continues to maintain in-house storage for operational customer data. Ideally, the hybrid approach allows a business to take advantage of the scalability and cost-effectiveness that a public Cloud Computing environment offers without exposing mission-critical applications and data to third-party vulnerabilities.



The Cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). An example where a community Cloud could be useful is in a state government setting. If various agencies within the state government operate under similar guidelines, they could all share the same infrastructure and spread the cost among themselves. A community Cloud may be managed by the organizations or by a third party. With the costs spread over to fewer users than a public cloud, this option is more expensive but may offer a higher level of privacy, security, and/or policy compliance. The community Cloud offers organizations access to a vast pool of resources than that in the private Cloud.

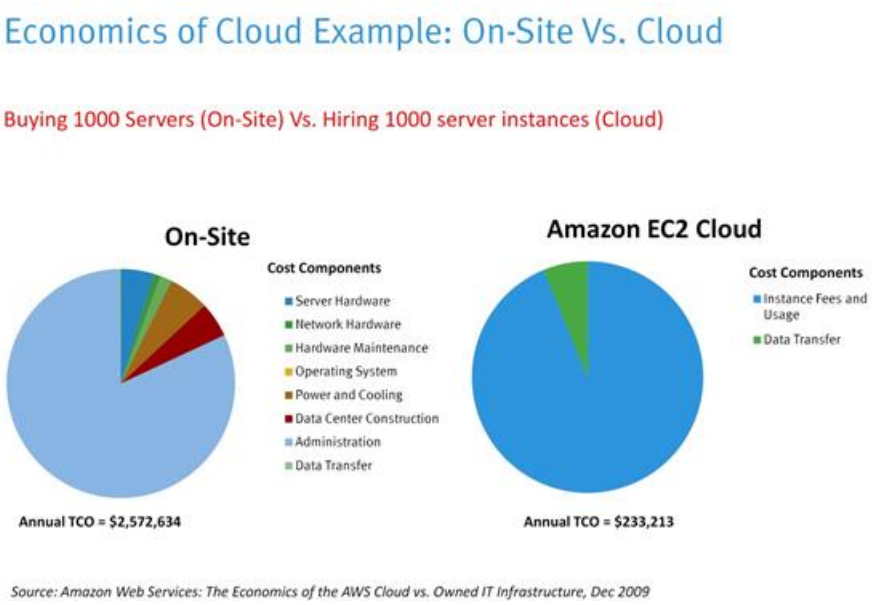


Cloud computing has changed the economics of IT. Capital expenditure (CAPEX) is required to build IT infrastructure. Because organizations hire and use resources from Cloud service providers, they will see more of Operational Expenditure (OPEX). The Cloud provides various cost saving options:

- **Infrastructure cost:** If an organization needs to build a large-scale system, they may need to invest in buying hardware (servers, storage, routers), software licensing, etc., which involves high upfront cost (CAPEX). With Cloud, IT infrastructure investment is minimized.
- **Management cost:** Lack of in-house IT infrastructure minimizes the people cost associated with the management of those infrastructures.
- **Power and Energy cost:** Power consumption has become a concern for most organizations because energy costs continue to rise. The organizations that use Cloud applications and services save on power and energy use. An increase in energy efficiency translates into smaller carbon footprints for organizations, making Cloud a greener solution than traditional on-premise models.

Note:

- ***CAPEX:*** *A capital expenditure which is incurred to buy fixed assets, for example, servers, storage, etc.*
- ***OPEX:*** *An ongoing day-to-day expense to run business, for example, management cost, power and cooling cost, etc.*



Consider a scenario where an organization wants to run its business-critical applications using 1000 servers to meet the desired service levels. They have two options to consider: the first option is to set up an on-site infrastructure for running 1000 servers and the second one is to hire 1000 instances of servers on an Amazon EC2 Cloud.

Let us consider various cost components involved in both these options:

In the first option, to set up an on-site infrastructure, the organization would require capital investment for purchasing server, storage, and network hardware, together with additional expenses for hardware maintenance, licensing OSs, power and cooling options, building data center infrastructure, administrative costs, and data transfer.

In contrast to that, the second option involves only two cost components: the major cost on instance usage and a minor cost on data transfer.

The diagram displayed on this slide—sourced from Amazon.com—shows that the first option incurs 10 times more TCO, compared to the second option. This clearly illustrates the economic benefit of Cloud, compared to an on-site infrastructure.

Both the Cloud consumers and providers have their own challenges. The following are the challenges of the consumers:

- **Security and Regulations:** Consumers may have business-critical data, which calls for protection and continuous monitoring of its access. With the Cloud, the consumer may lose control

of the sensitive data – for example, the consumer may not know in which country the data is being stored – and may violate some national data protection statutes (EU Data Protection Directive and U.S. Safe Harbor program). Many regulations impose restrictions to distribute data outside of the organization's territory.

- **Network latency:** Consumers may access Cloud services from anywhere in the world. Although Cloud resources are distributed, the resources may not be close to the consumer location, resulting in high network latency. A high network latency results in application timeout, thereby disabling end users from accessing the application.
- **Supportability:** Cloud may not support all applications. For example, a consumer may want to leverage the Cloud platform service for their proprietary applications, but the Cloud provider may not have a compatible Operating System (OS). Also, legacy applications may not be supported on Cloud.
- **Interoperability:** Lack of interoperability between the APIs of different Cloud service providers create complexity and high migration costs for consumers when it comes to moving from one service provider to another.

The following are the challenges for the Cloud service providers:

- **Service warranty and service cost:** Cloud service providers usually publish a Service Level Agreement (SLA), so that their consumers are aware of the availability of service, quality of service, downtime compensation, and legal and regulatory clauses. Alternatively, customer-specific SLAs may be signed between a Cloud service provider and a consumer. Cloud providers must ensure that they have adequate resources to provide the required level of services. SLAs typically mention penalty amount, if the Cloud service providers fail to provide services. Because the Cloud resources are distributed and continuously scaled to meet variable demands, it is a challenge to the Cloud providers to manage physical resources and estimate the actual cost of providing the service.
- **Huge numbers of software to manage:** Cloud providers, especially SaaS and PaaS providers, manage a number of applications, different Operating Systems (OSs), and middleware software to meet the needs of a wide range of consumers. This requires service providers to possess enough licenses of various software products, which, in turn, results in unpredictable ROI.
- **No standard Cloud access interface:** Cloud service providers usually offer proprietary applications to access their Cloud. However, consumers might want open APIs or standard APIs to become tenants of multiple Clouds. This is a challenge for Cloud providers because this requires an agreement among Cloud providers and an upgrade of their proprietary applications to meet the standard.

Module 9: Cloud Infrastructure and Management

Upon completion of this module, you should be able to:

- Explain the Cloud infrastructure components
- Describe the Cloud service creation processes
- Describe the Cloud service management processes

This module focuses on the Cloud infrastructure components and Cloud service creation processes. It also includes the Cloud service management processes that ensure that the delivery of Cloud services is aligned with business objectives and expectations of Cloud service consumers.

Module 9: Cloud Infrastructure and Management

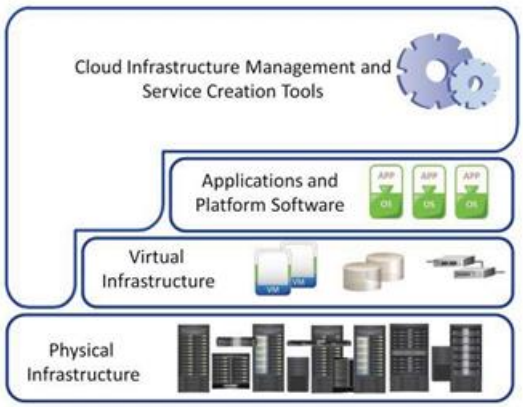
Lesson 1: Cloud Infrastructure and Service Creation

Topics covered in this lesson:

- Cloud infrastructure framework and components
- IT resources included in each Cloud component
- Processes to create Cloud services

This lesson covers the Cloud infrastructure framework, framework components, and IT resources in each of the component. It also includes Cloud service creation processes.

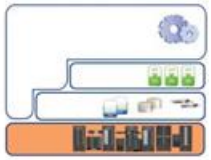
Cloud Infrastructure Framework



The Cloud infrastructure framework consists of the following components:

- Physical infrastructure
- Virtual infrastructure
- Applications and platform software
- Cloud infrastructure management and service creation tools

The resources of the above components are aggregated to provide Cloud services.



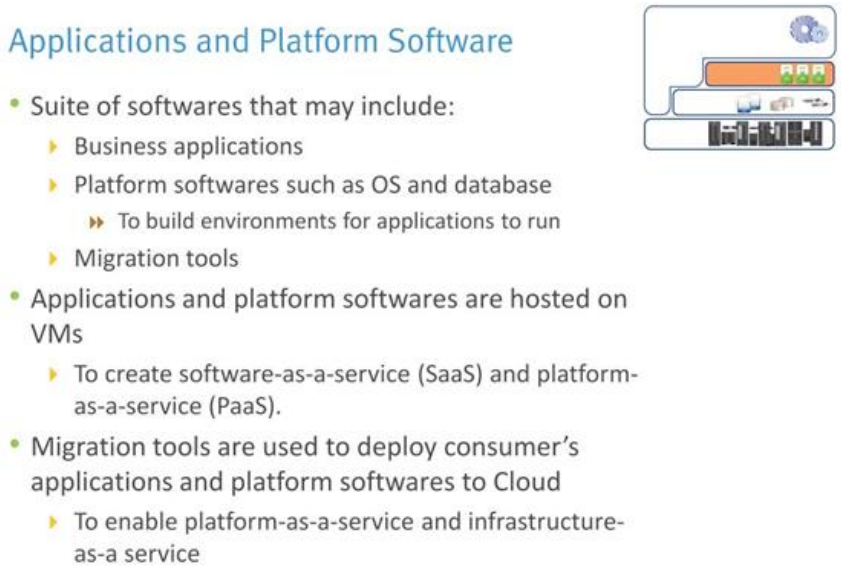
The physical infrastructure consists of physical IT resources that include physical servers, storage systems, and physical network components, such as physical adapters, switches, and routers. Physical servers are connected to each other, to the storage systems, and to the clients via physical networks such as IP network, FC SAN, IP SAN, or FCoE network.

Cloud service providers may use physical IT resources from one or more data centers to provide services. If the physical IT resources are distributed across multiple data centers, connectivity must be established among them. The connectivity enables data centers in different locations to work as single large data center. This enables both migration of Cloud services across data centers and provisioning Cloud services using resources from multiple data centers.

Virtual infrastructure consists of the following resources:

- Resource pools such as CPU pools, memory pools, network bandwidth pools, and storage pools
- Identity pools such as VLAN ID pools, VSAN ID pools, and MAC address pools
- Virtual IT resources consist of:
- VMs, virtual volumes, and virtual networks
- VM network components such as virtual switches and virtual NICs

Virtual IT resources gain capacities such as CPU cycles, memory, network bandwidth, and storage space from the resource pools. Virtual networks are defined using network identifiers such as VLAN IDs and VSAN IDs from the respective identity pools. MAC addresses are assigned to virtual NICs from the MAC address pool.

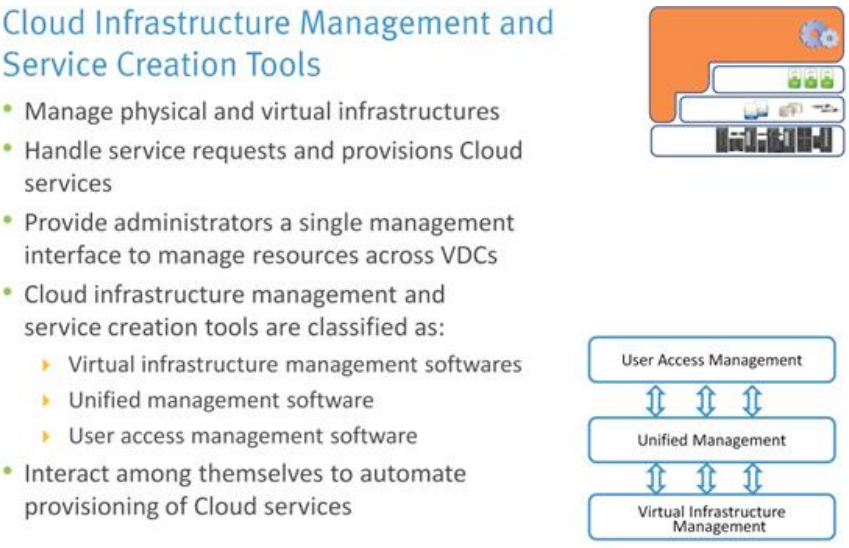


Applications and platform software layers include a suite of softwares such as:

- Business applications
- Operating systems and database. These softwares are required to build environments for running applications.
- Migration tools

Applications and platform software are hosted on VMs to create software-as-a-service (SaaS) and platform-as-a-service (PaaS).

For SaaS, applications and platform software are provided by the Cloud service providers. For PaaS, only the platform software is provided by the Cloud service providers; consumers export their applications to Cloud. In infrastructure as a service (IaaS), consumers upload both applications and platform software to Cloud. Cloud service providers supply migration tools to consumers, enabling deployment of their applications and platform software to Cloud.

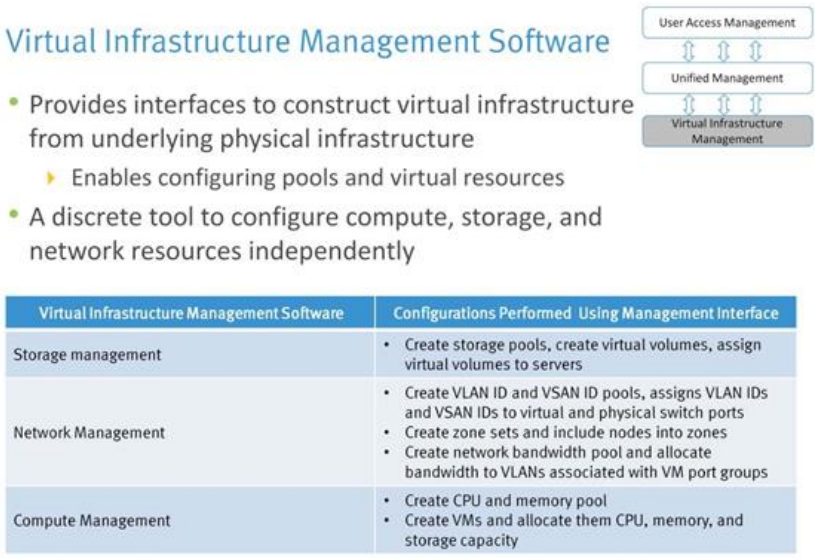


Cloud infrastructure management and service creation tools are responsible for managing physical and virtual infrastructures. They enable consumers to request for Cloud services; they provide Cloud services based on consumer requests and allow consumers to use the services. Cloud infrastructure management and service creation tools automate consumer requests processing and creation of Cloud services. They also provide administrators a single management interface to manage resources distributed in multiple virtualized data centers (VDCs).

Cloud management tools are classified as:

- Virtual infrastructure management software: Enables management of physical and virtual infrastructure resources.

- Unified management software: Responsible for creating Cloud services.
- User access management software: Enables consumers to request for Cloud services. These software interact with each other to automate provisioning of Cloud services.

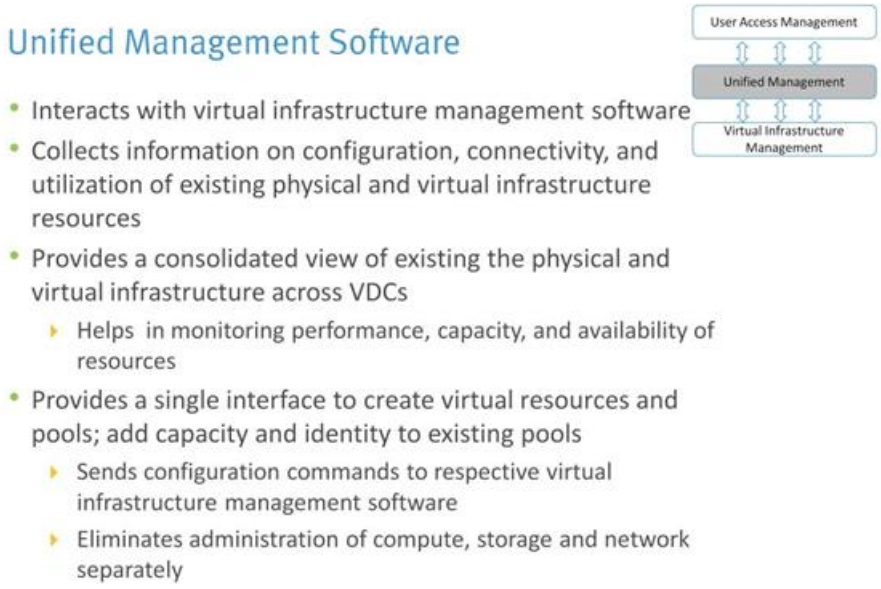


Virtual Infrastructure Management Software	Configurations Performed Using Management Interface
Storage management	• Create storage pools, create virtual volumes, assign virtual volumes to servers
Network Management	• Create VLAN ID and VSAN ID pools, assigns VLAN IDs and VSAN IDs to virtual and physical switch ports • Create zone sets and include nodes into zones • Create network bandwidth pool and allocate bandwidth to VLANs associated with VM port groups
Compute Management	• Create CPU and memory pool • Create VMs and allocate them CPU, memory, and storage capacity

Virtual infrastructure management software provides interfaces to construct virtual infrastructure from underlying physical infrastructure. It enables communication with tools, such as hypervisors and physical switch operating systems, and also configuration of pools and virtual resources with the assistance of these tools.

In a VDC, typically, compute, storage, and network resources (within physical and virtual infrastructure) are configured independently using a different virtual infrastructure management software; for example, a storage array has its own management software. Similarly, network and physical servers are managed independently using network and compute management software respectively.

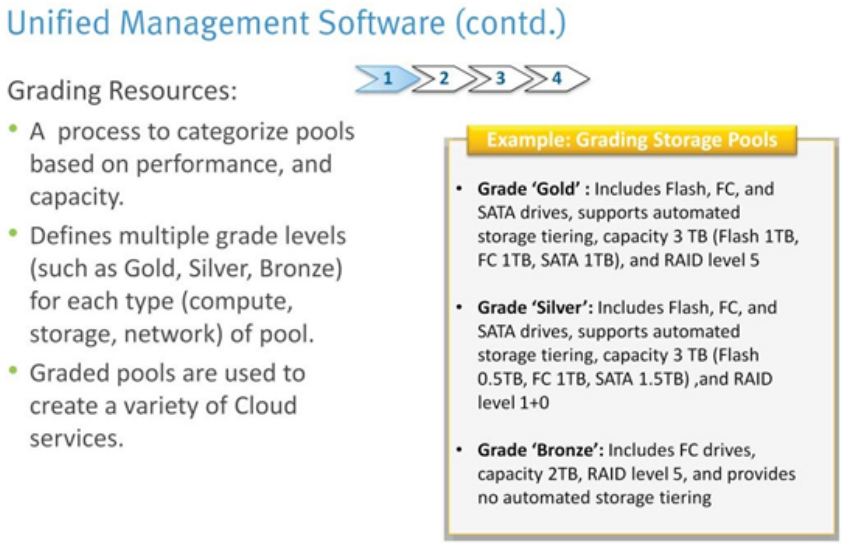
This slide provides a list of common configurations performed using different virtual infrastructure management software.



Unified management software interacts with all standalone virtual infrastructure management software and collects information on the existing physical and virtual infrastructure configurations, connectivity, and utilization. Unified management software compiles this information, and provides a consolidated view of IT resources scattered across VDCs. This allows an administrator to monitor performance, capacity, and availability of physical and virtual resources centrally.

In addition to providing a consolidated view, unified management software provides a single management interface to create virtual resources and pools. It also enables an administrator to add capacity and identity to the existing pools. It passes configuration commands to the respective VDC management software, which executes the instructions. This eliminates the administration of compute, storage, and network resources separately using native management software.

The key function of a unified management software is to create Cloud services. It performs a series of processes to create Cloud services. This slide provides a list of these processes and the subsequent slides describe these processes.



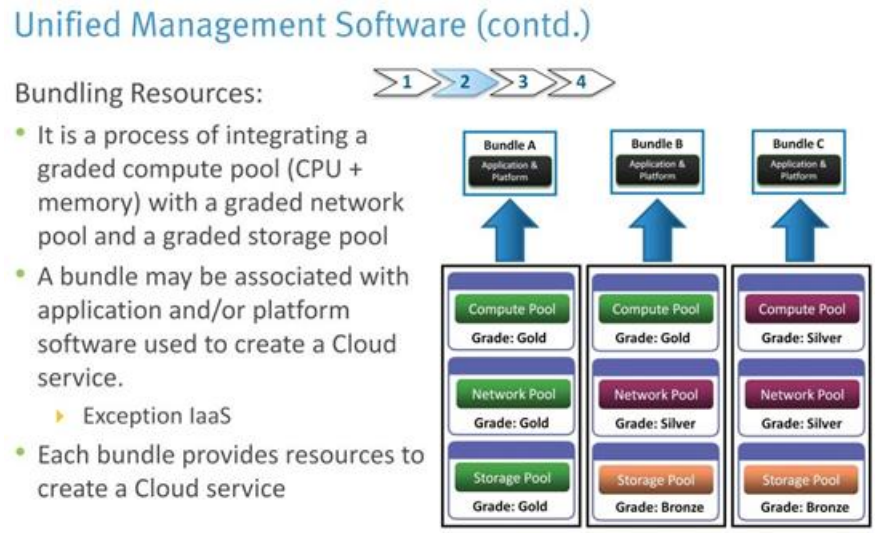
Unified management software allows an administrator to grade pools. Resource grading is a process to categorize pools based on their capabilities, such as performance and capacity.

Multiple grade levels may be defined for each type (such as compute, storage, and network) of pool. Each grade level is marked with a grade value such as 'Gold', 'Silver', and 'Bronze'.

The number of grade levels for a type of pool depends on business requirements.

This slide provides an example of grading storage pools. Three grade values are used to mark three different grade levels.

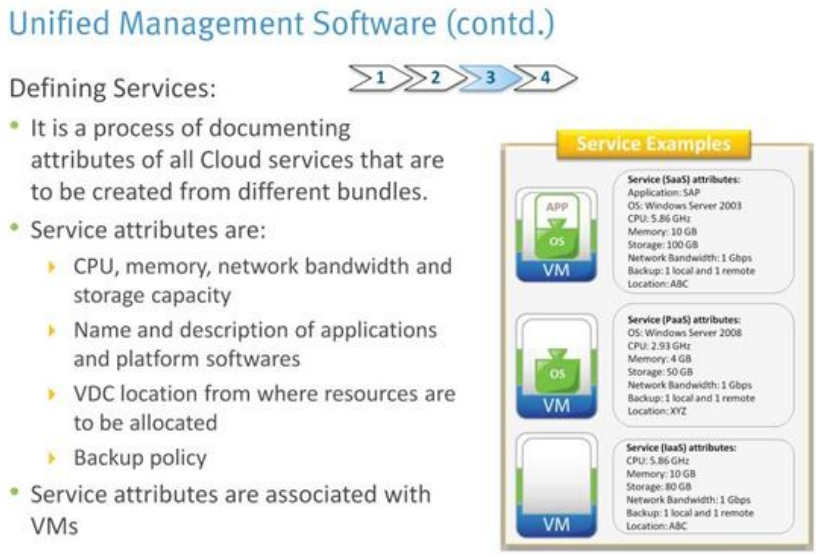
Resource grading standardizes pools based on their capabilities. Pools of different grades are used to create a variety of Cloud services, providing choices to the Cloud service consumers.



Resource bundling is a process of integrating a graded compute pool with a graded network pool and a graded storage pool. A compute pool implies a CPU pool plus a memory pool. The integration allows a group of physical servers, from which the compute pool is created, to use the storage and network pools for storing and transferring data respectively. The integrated pools are given a bundle name and are treated as a single entity.

A bundle may be associated with a platform software and/or an application to enable software- and platform- as a service. An exception is infrastructure-as-a-service, where only compute, network, and storage resources are required. The association enables specific application and platform software to use specific bundle resources.

A Cloud service provider typically maintains multiple resource bundles with different combinations of grade values for compute, network, and storage pools (shown in the figure). Each bundle is used to create a Cloud service, which inherits capabilities of the pools in the bundle.

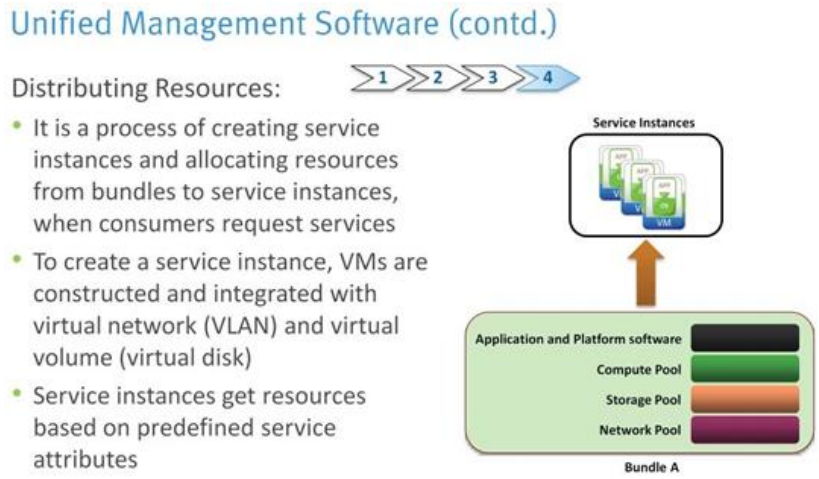


A unified management software helps in defining Cloud services. It allows an administrator to list all the services along with their service attributes. The service attributes are:

- CPU, memory, network bandwidth, and storage capacity that are to be allocated to services from different bundles
- Name and description of applications and platform software
- VDC location (Bundle location) from where resources are to be allocated
- Backup policy, such as the number of backup copies of a service instance and the location of the backup data. For example, a backup policy could be creating two copies of VM files and maintaining the copies in the same VDC or transferring to another VDC.

A unified management software allows creating a variety of Cloud services with diverse service attributes. It defines service attributes based on capabilities of the bundle resources associated with the service.

Service attributes are associated with VMs used to provide a service. Depending on the type of service (such as SaaS, PaaS, IaaS), VMs, as shown in the example, may or may not host applications and platform software.



Resource distribution involves creating service instances and allocating resources from bundles to service instances when consumers request for services. At the time of creating service instances, VMs are constructed and integrated with virtual networks (VLANs) and virtual volumes (virtual disks). Application and platform software may be installed on the VMs. The service instances obtain compute, network, and storage capacity from appropriate bundles. The allocation of capacity and software installation follows attributes defined for the service.

A user access management software provides a web based user interface to consumers. Consumers may use the interface to request for Cloud services. User access management software interacts with unified management software and forwards all service requests. The unified management software provisions these services, which are made available to the consumers via user access management software.

User access management software allows an administrator to create and publish service catalogue. A service catalog is a structured document with information about all Cloud services available to consumers. It includes information about service attributes, prices, and request

processes.

A user access management software authenticates consumers before forwarding their requests to a unified management software. It maintains a database for all consumer accounts.

A user access management software monitors allocation or usage of resources associated to a Cloud service instance. Based on allocation or usage, it generates a chargeback report. The chargeback report is visible to consumers, providing transparency between a consumer and a provider.

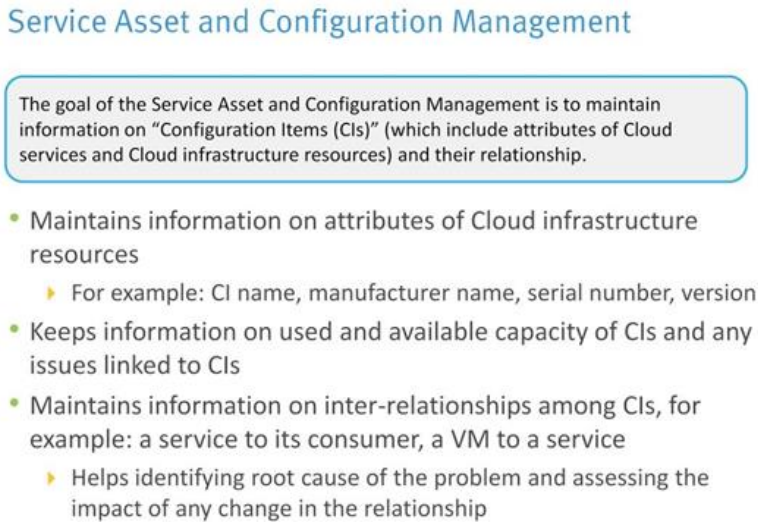
This lesson covers an overview of the processes in Cloud service management. It also includes automation of service management processes using service management tools.

Cloud service management involves a set of organizational processes which align the delivery of Cloud services with the business objectives and to the expectation of Cloud service consumers. Creating and delivering services involve giving consumers what they want.

However, it is Cloud service management processes that work at the background to ensure all services perform as committed.

An organization with the best service creation tools, but poor service management processes, often fails to deliver services of required quality and meet business objectives. For example, suppose a service instance cannot not obtain the required capacity because the provider has a shortage of resources, or a consumer is unable to use a service for a significant time period due to an unresolved error in the provider's infrastructure. Here, Cloud service providers must employ proper service management processes to provide Cloud services. Hence, there is a need to understand the objectives and activities in each service management process.

This slide provides a list of service management processes. In general, the activities associated with each of the processes are automated using tools. Each of these processes are described in the subsequent slides.



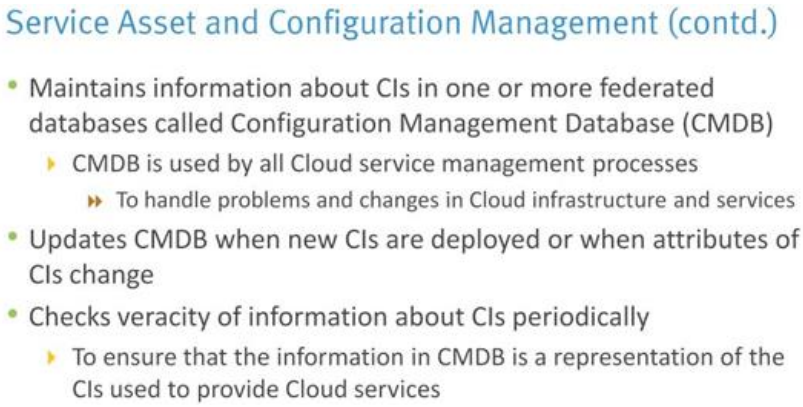
A service asset and configuration management process maintains information about:

- Attributes of Configuration Items (CIs) such as Cloud services and Cloud infrastructure resources. CIs are considered as IT assets.
- Relationship among the CIs

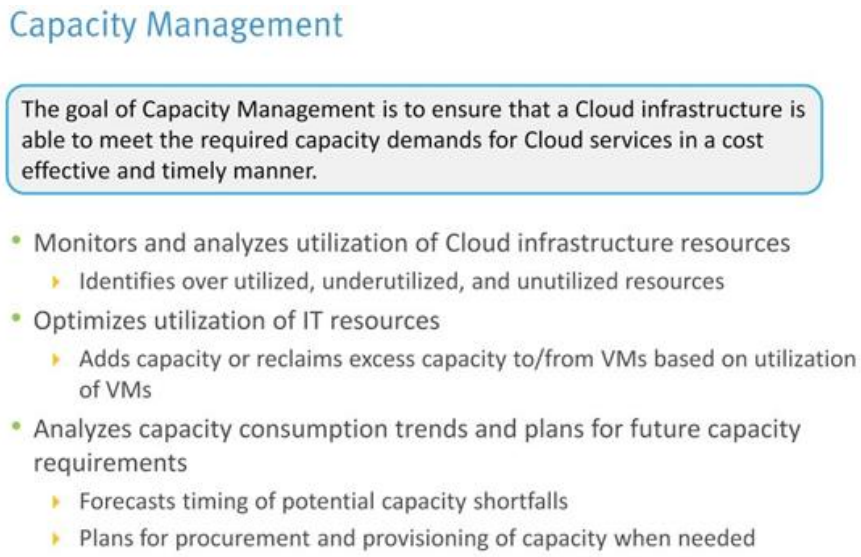
Service asset and configuration management maintains information about attributes of Cloud infrastructure resources, such as physical servers, storage arrays, and spare components. The information includes CI's name, manufacturer name, serial number, license status, version, description of modification, location, and inventory status (Example: on order, available, allocated, or retired).

It keeps information on used and available capacities of CIs and any issues linked to the CIs.

It also maintains information on the inter-relationships among CIs such as, a service to its consumer, a VM to a service, a physical server to a VM hosted on the server, a physical server to a switch sending data to the server, and a VDC to its location. This ensures that configuration items are viewed as integrated components. Consequently, it helps identifying the root cause of the problem and assessing the impact of any change in the relationship. For example, when an administrator finds that a switch has failed, he/she will be able to determine what all are affected by that outage. Alternatively, when an administrator decides to upgrade the CPU and memory of a physical server, he/she will be able to identify the items affected by the change.



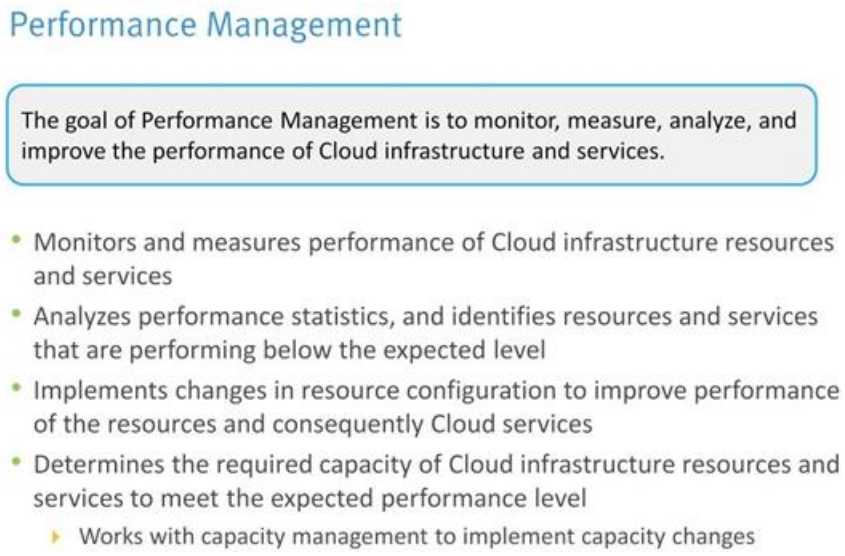
Service asset and configuration management maintains all information about configuration items in one or more federated databases that are called Configuration Management Database (CMDB). CMDB is used by all Cloud service management processes to deal with problems or to include changes into Cloud infrastructure and services. Service asset and configuration management updates CMDB as and when new configuration items are deployed or when attributes of configuration items change. It periodically checks the veracity of information on configuration items to ensure that the information it maintains is an exact representation of the configuration items used to provide Cloud services.



Capacity management ensures that a Cloud infrastructure is able to meet the required capacity demands for Cloud services in a cost effective and timely manner.

Capacity management monitors the utilization of IT infrastructure resources. It identifies over utilized and underutilized/unutilized resources. It optimizes the utilization of IT resources by adding capacity or reclaiming the excess capacity to/from VMs based on the utilization of VMs.

Capacity management is responsible for planning future IT infrastructure requirements for Cloud services. It gathers information on the present and past utilization of resources and establishes trends on capacity consumption. Based on the trend, it forecasts growth in consumer demand for capacity in future. It identifies the timing of potential capacity shortfalls. It plans for procurement and provisioning of capacity as and when needed.

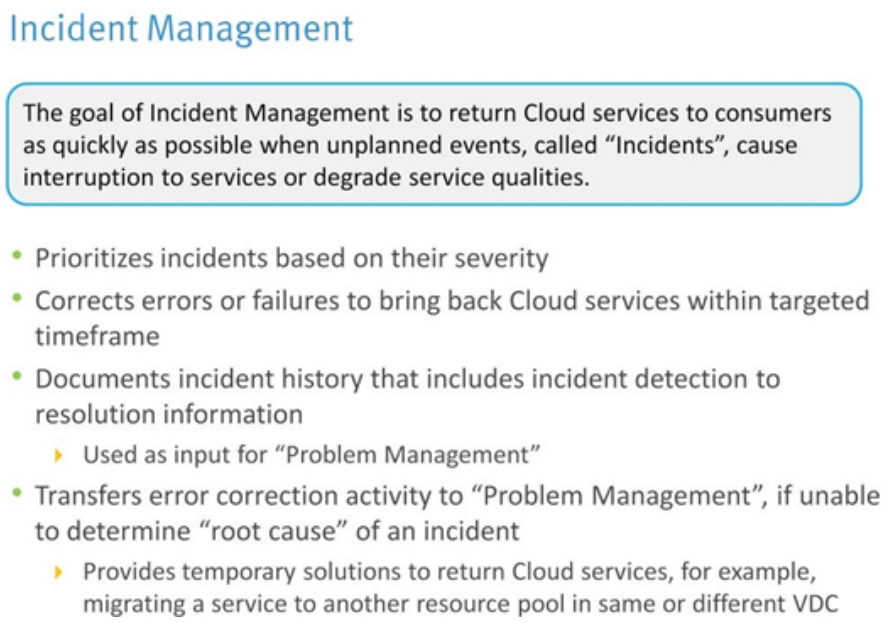


Performance management involves monitoring, measuring, analyzing, and improving the performance of Cloud infrastructure and services.

Performance management monitors and measures performance, such as response time and data transfer rate, of the Cloud infrastructure resources and services. It analyzes the performance statistics and identifies resources and services that are performing below the expected level.

Performance management ensures that all infrastructure resources responsible for providing Cloud services are meeting or exceeding required the performance level. It implements changes in resource configuration to improve performance of the resources. For example, changing CPU cycles and memory distribution setting at hypervisor, adding virtual CPUs on a VM to fulfill needs of a multi-threaded application, or selecting a different RAID level for storing VM files.

Overutilization of resources is a key reason for performance degradation. Performance management determines the required capacity of Cloud infrastructure resources and services to meet the expected performance level. It works together with capacity management and implements changes related to resource capacity and performance.



An incident is an unplanned event that causes or may cause interruption to a Cloud service or reduces the quality of a service, such as a degraded performance. An incident may not always cause service failure; for example, failure of a disk from a mirror set of a RAID 1 protected storage. However, if not attended, recurring incidents may cause service interruption in future. Incident management involves returning Cloud services with the required qualities to consumers as quickly

as possible when incidents cause service interruptions or degradations.

Incident management prioritizes incidents based on their severity and provides solutions to bring back Cloud services within an agreed timeframe. It tries to recover Cloud services as soon as possible by correcting the error or failure that caused the incident.

Incident management documents the incident history with details of the incident symptoms, affected services and consumers, Cloud infrastructure resources that form the Cloud service, time to resolve the incident, severity of the incident, description of the error, and the incident resolution data. The incident history is used as an input for problem management.

If incident management is unable to determine and correct the 'root cause' of an incident, an error-correction activity is transferred to problem management. In this case, incident management provides a temporary solution to the incident; for example, migration of a service to different resource pools in the same VDC or a different VDC.

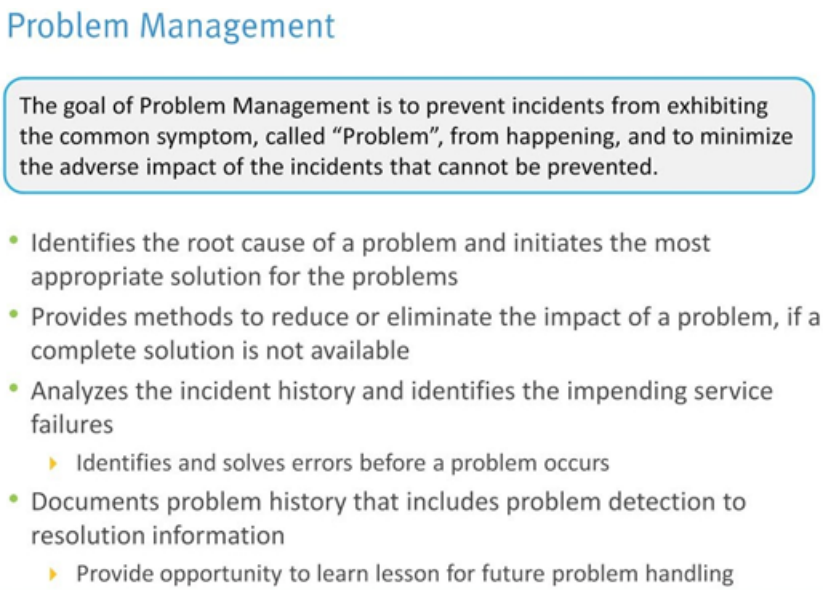
Incident Management (contd.)

- Involves multiple support groups to solve incidents

Support Group	Description
First level support group: Service Desk	• Single point of contact between Cloud service provider and consumers. • Registers incidents and sets priority to incidents. • Undertakes corrective measures to restore a failed service. • Transfers incidents to Technical Support Group, if unable to solve the incidents. • Keeps consumers informed about incident status.
Second level support group: Technical Support Group	• Provides solution to incidents which can not be solved by first level support group. • May request for incident resolution from hardware and software manufacturers
Third level support group: Hardware and Software Manufacturers	• Provide solutions to incidents, if requested by second level support group.

Incident management may involve multiple support groups to provide solution to incidents; for example:

- First level support group: Service desk is the first level support group. It is the single point of contact between the Cloud service provider and consumers. It registers received incidents and sets priority to incidents. Incidents may be reported by consumers or populated by a monitoring tool, such as a unified management software, by means of event alerts. Following the recording of the incidents, the group undertakes corrective measures to restore a failed service. If no solution can be achieved, the first level support group transfers the incident to the technical support group (second level support group). The first level support group keeps the consumers informed about the incident status.
- Second level support group: It consists of technical experts who provide solution to the incidents that cannot be solved by the first level support group. If necessary, it may request for resolution from hardware and software manufacturers (third level support group)
- Third level support group: It provides solutions to incidents, if requested by the second level support group.

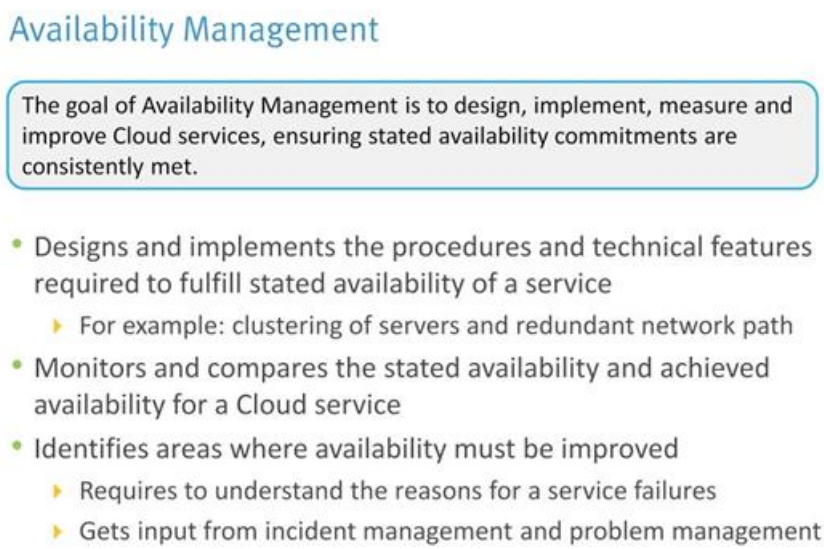


A problem occurs as a result of multiple incidents exhibiting a common symptom. Problems can also be identified from a single significant incident that is indicative of a single error, for which the cause is unknown, but for which the impact is high. Problem management prevents incidents from repeating, and minimizes the adverse impact of incidents that cannot be prevented.

Problem management prioritizes problems based on their impact to business and takes corrective actions. It identifies the root cause of a problem and initiates the most appropriate solution for the problem. If a complete resolution is not available, problem management provides methods to reduce or eliminate the impact of a problem.

Problem management proactively analyses the incident history and identifies the impending service failures. It identifies and solves errors before a problem occurs.

Problem management also documents the problem history with details on the symptoms, the affected services and consumers, the Cloud infrastructure resources that form the Cloud service, the time to resolve the problem, the severity of the problem, the error description, and the resolution data. The problem history provides an opportunity to learn and handle future problems.



Availability management ensures that the availability requirements of a Cloud service is constantly met. It designs and implements the procedure and technical features required to fulfill the stated availability of a service. An example is clustering of physical servers, where a server failure results in the failover of services from the failed server to another available server in the cluster. Other examples are implementing redundant network paths and remote replication of VM files.

Availability management ensures that the Cloud infrastructure, business continuity processes, and tools are appropriate to meet the required availability level. It continuously

monitors and compares the stated availability and the achieved availability of Cloud services and identifies the areas where the availability must be improved. Availability management requires understanding of reasons of service failures. This allows to identify areas for availability improvement. The incident and problem management provide the key input to availability management regarding the causes of service failures and the time required to resume services.

Availability management ensures improved consumer service and cost effectiveness for the Cloud service provider. It reduces the time for consumers to engage with the service desk. This improves consumer satisfaction and consequently lifts the Cloud service provider's reputation. As the service availability level is raised, less staff is required to handle service desk and problem management. This significantly reduces administration cost.

Service Catalogue Management

The goal of Service Catalogue Management is to ensure that a "Service Catalogue" is created and maintained with accurate information on all the available Cloud services.

- Ensures that the information in the service catalogue is accurate and up-to-date
- Ensures clarity, completeness, and usefulness when describing service offerings in the service catalogue
- Evaluates and upgrades service catalogue continually to include new services and improvements in service offerings

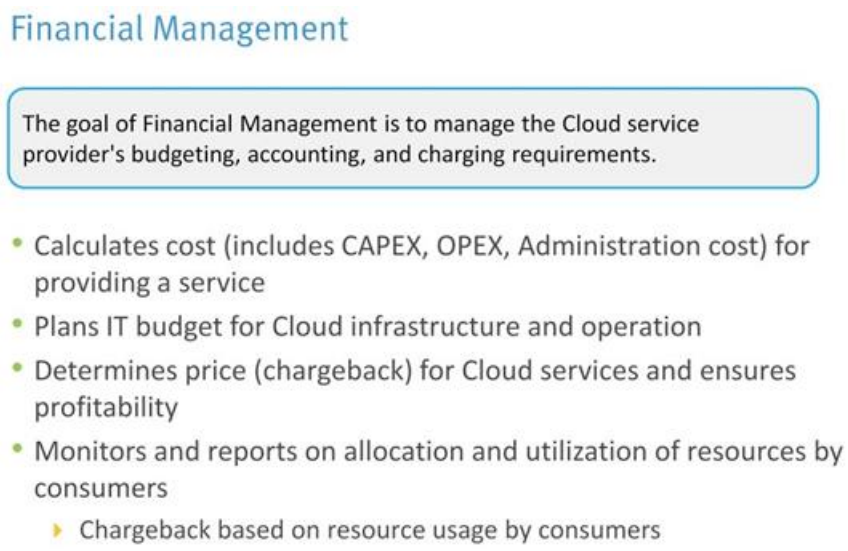
Service catalog management involves creating and maintaining a service catalogue. It ensures that the information in the service catalogue is accurate and up-to-date. Service management brings in clarity, completeness, and usefulness when describing service offerings in the service catalogue. It ensures that the service description is unambiguous and valuable to consumers.

Business drivers and technologies are ever-changing, and consequently, Cloud services are bound to change. Service catalogue management continually evaluates service offerings in a service catalogue and upgrades the service catalogue to include new services and changes in the service offerings.

Common Attributes of a Service in Service Catalogue

Attribute	Description
Service name	• Enables consumers to easily understand what a service offering is, by its name.
Description	• Tells consumers what the service is and what value the service provides. • It should be clear, concise, and complete, and use simple and appropriate language.
Features and Options	• Provides a list of options for selection, such as a list of operating systems for virtual desktop service. • Provides technical description of each option, such as the software version.
Service and Support Expectations	• Describes quality of service such as performance, availability and security. • Describes technical support provided with a service; for example, a technical support may be from Monday through Friday during normal business hours or continuous (24x7) technical support.
Price	• Provides price of service, including price list for different features and options.
Provisioning Timeframe	• Provides a timeline for Cloud service provider to deliver service.

An important activity in service catalogue management is to represent Cloud services in a manner that clearly indicates value of the services. Typically, each service offering in a service catalog is presented using a list of attributes. This slide provides a list of common service attributes that may be included for each service offering.

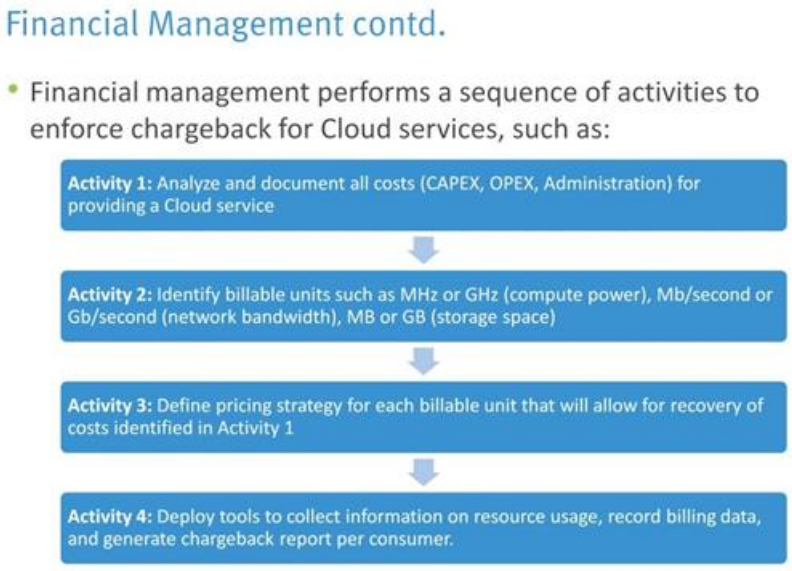


Financial management involves calculating the cost of providing a service. The cost includes capital expenditure (CAPEX), such as procurement and deployment costs of Cloud infrastructure, on-going operational expenditures (OPEX), such as power, cooling, facility cost, and administration cost such as cost of labor.

Financial management plans for investments to provide Cloud services and determines the IT budget for Cloud infrastructure and operation for a fixed time period.

Financial management determines the price (chargeback) a consumer is expected to pay for a service and ensures profitability. It allows the Cloud service provider to recover the cost of providing Cloud services from consumers.

It also monitors and reports on allocation and utilization of resources by consumers. This ensures that the consumers pay for what they actually use.

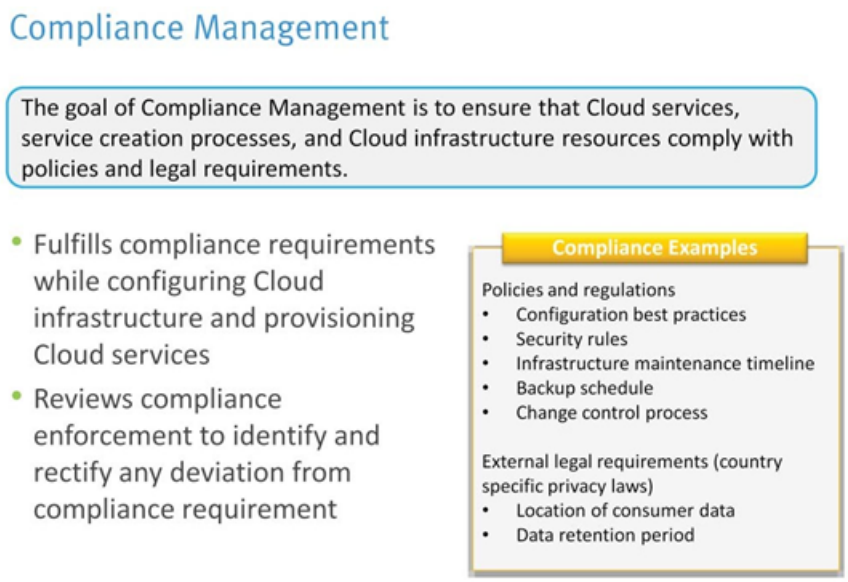


Financial management typically performs a sequence of activities to enforce chargeback for providing Cloud services. These activities are listed below:

- **Activity 1**: Analyze and document all relevant costs, including all capital, operational, and administration costs.
- **Activity 2**: Identify billable units for Cloud services. A unit could be MHz or GHz (for compute power), Mb/second or Gb/second (for network bandwidth), and MB or GB (for storage space).
- **Activity 3**: For each billable unit, define a pricing strategy by choosing pricing options that

will allow for recovery of costs identified in Activity 1.

- **Activity 4**: Deploy the tools necessary to collect information on resource usage, record the billing data, and generate the chargeback report per consumer. These tools are integrated with Cloud infrastructure management and service creation tools.



Compliance management ensures that the Cloud services, service creation processes, and Cloud infrastructure resources comply with policies and legal requirements.

Compliance management ensures that the compliance requirements are fulfilled while configuring Cloud infrastructure and provisioning Cloud services. Compliance requirements primarily include:

- Policies and regulations
- External legal requirements

Adhering to policies and regulations applies to both Cloud service provider and consumers. Policies and regulations could be on configuration best practices, security rules such as administrator roles and responsibilities, physical infrastructure maintenance timeline, information backup schedule, and change control processes such as changes to Cloud infrastructure resources and services.

External legal requirements are data privacy laws imposed by different countries. These laws may specify geographical locations to store consumer data and disallow modification or deletion of data during its retention period; for example, many countries do not allow financial data to cross country borders. In this case, compliance management must ensure that the Cloud infrastructure that provides the required service is located within the periphery of these countries and data access is restricted to the citizens of those countries.

Compliance management periodically reviews compliance enforcement in infrastructure resources and services. If it identifies any deviation from compliance requirement, it initiates corrective actions.

The Cloud service management processes may be automated using service management tools. These tools may be integrated with the Cloud infrastructure management and service creation tools to automate several activities in various service management processes.

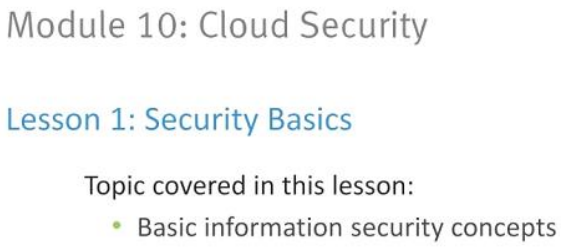
Manual approach to Cloud service management is subject to errors, difficult to audit, and requires considerable time and effort to enforce. It increases the administration cost and consequently increases the cost of providing Cloud services. It also raises risk of deviating from increasingly stringent compliance requirements and service qualities. Service management tools accomplish many repetitive manual activities provided below and mitigate risk of human error:

- CMDB population: Gathers information on configuration items and their relationships and populates CMDB with that information accurately and on time.
- Incident and problem creation: Continuously monitors availability and performance of Cloud infrastructure and deployed services. Automatically registers incidents when components fail. Periodically analyzes past history of incidents and automatically creates problem cases when recurring incidents are detected.
- Analysis and forecasting: Identifies unused and over-allocated capacity to VMs and reclaims the capacity. Analyzes capacity usage trend and forecasts capacity requirements based on the trend analysis.
- Chargeback: Measures consumed capacity in the form of billable units. Calculates price (chargeback) against resource usage and generates chargeback report.
- Compliance enforcement: Ensures that Cloud infrastructure configuration and service creation processes adhere to the required internal and external policies and guidelines.

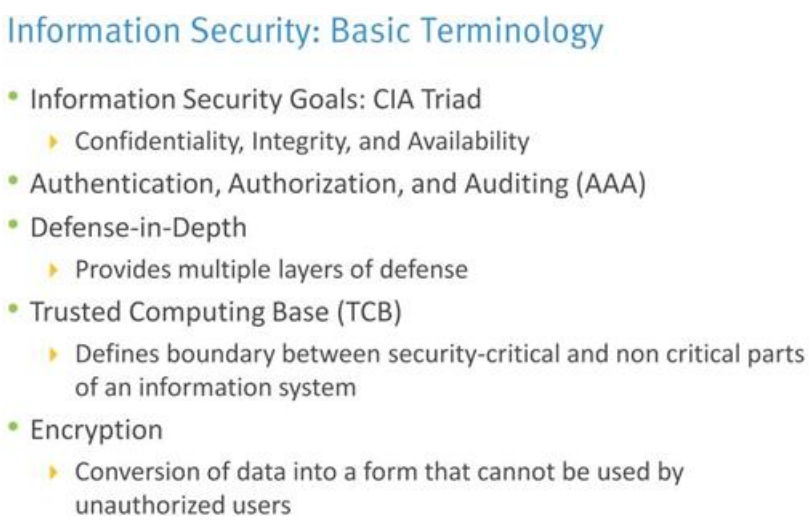
UNIT-6

Cloud Security and Migration to cloud

This module focuses on security concerns and counter measures in a VDC and Cloud environment. It discusses key security concerns and threats. It describes various infrastructure security mechanisms in VDC and Cloud environment, including access control, identity management, governance, etc. Additionally, the module lists Cloud security best practices.



This lesson covers the basic information on security concepts.



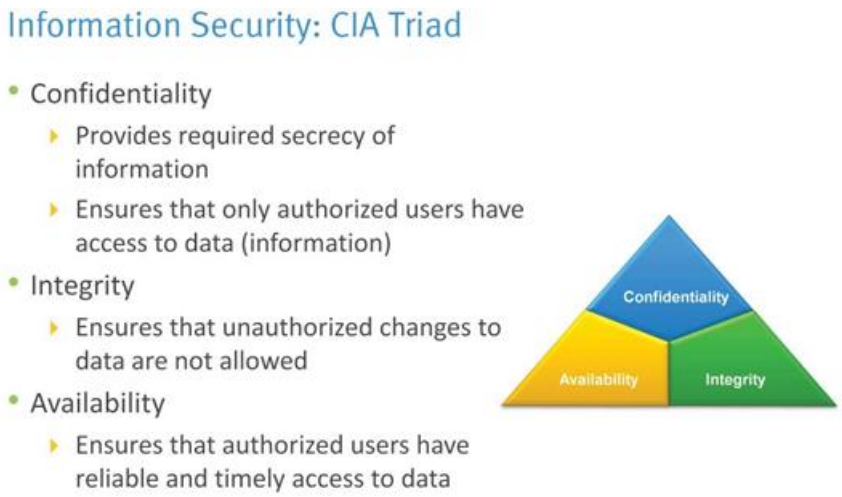
Basic information security terminology, which will assist in understanding VDC and Cloud security include:

- **CIA triad:** A security framework for an information system has three primary goals: Confidentiality, Integrity, and Availability of physical and logical resources. This is commonly known as CIA triad.
- **AAA:** The security framework for an information system should provide authentication and authorization capabilities. Auditing assesses the effectiveness of security mechanisms.
- **Defense-In-Depth:** It is a risk management strategy which provides multiple layers of defense against attacks.

• **Trusted Computing Base (TCB):** TCB of an information system is the set of all components that are critical to its security. Vulnerabilities occurring inside the TCB might jeopardize the security of the entire system. This way, TCB essentially defines a boundary for security-critical and non-critical parts of an information system; for example, kernel services of an OS are more critical for its security than application level services. Therefore, kernel services are part of a TCB for an OS, whereas application level services need not be a part of it.

• **Encryption:** It is the conversion of data into a form that cannot be easily understood by unauthorized users. Decryption is the process of converting encrypted data back into its original form. Encryption is used to enforce confidentiality, privacy, and integrity.

The following slides provide details on some of the above mentioned concepts.

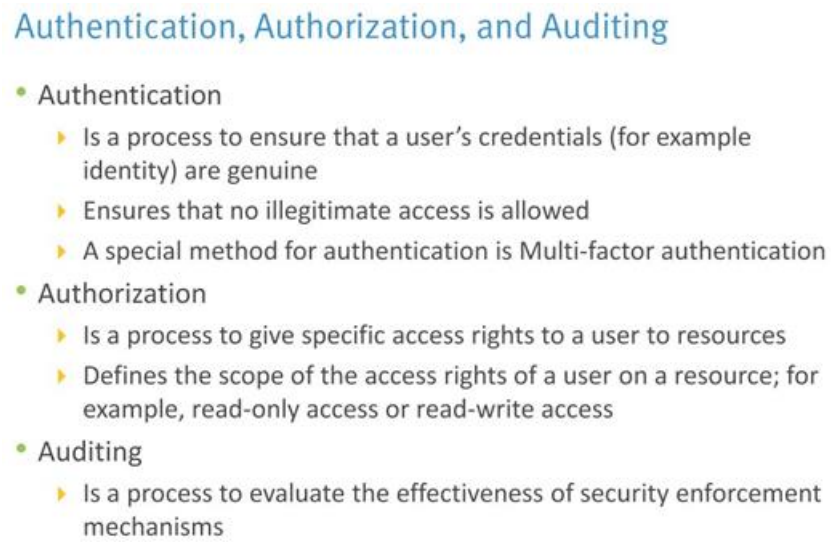


The CIA Triad includes:

• **Confidentiality**: Provides the required secrecy of information and ensures that only authorized users have access to data. In addition, it restricts unauthorized users from accessing information.

• **Integrity**: Ensures that unauthorized changes to information are not allowed. The objective of this goal is to detect and protect against unauthorized alteration or deletion of information.

• **Availability**: Ensures that authorized users have reliable and timely access to the compute, storage, and network resources. Availability also requires transferring data to different location(s) to ensure its availability if a failure occurs in any location.



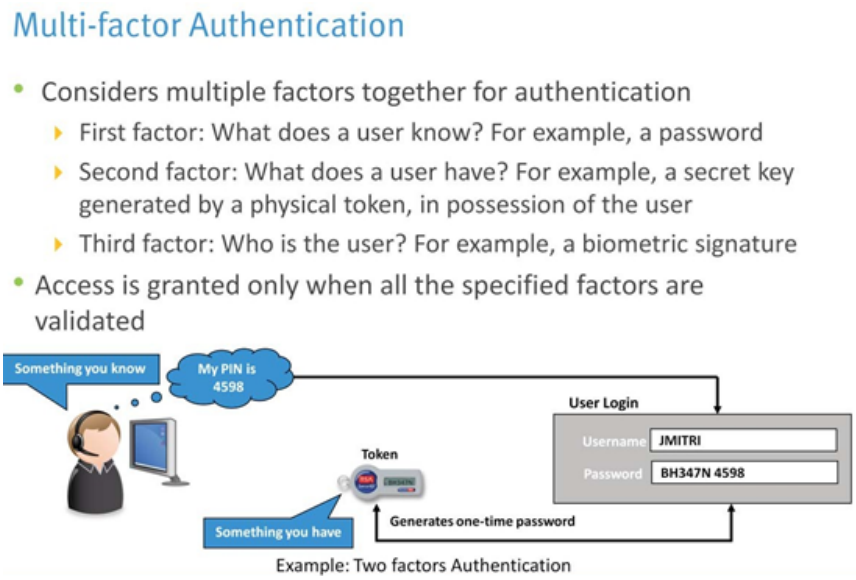
In an information system security framework, authentication and authorization

capabilities are required to ensure legitimate access to data.

Authentication is a process to ensure that a user's or asset's credentials (for example identity) are genuine so that no illegitimate access to information is allowed. Multi-factor authentication is a special method for authentication, which considers multiple factors together for authenticating a user.

Authorization is a process to grant specific access rights to a user on resources. Authorization defines the limits of the access rights of a user on a resource; for example, read-only access or read-write access on a file.

Finally, auditing is a process to evaluate the effectiveness of security enforcement mechanisms.



Example: Two factors Authentication

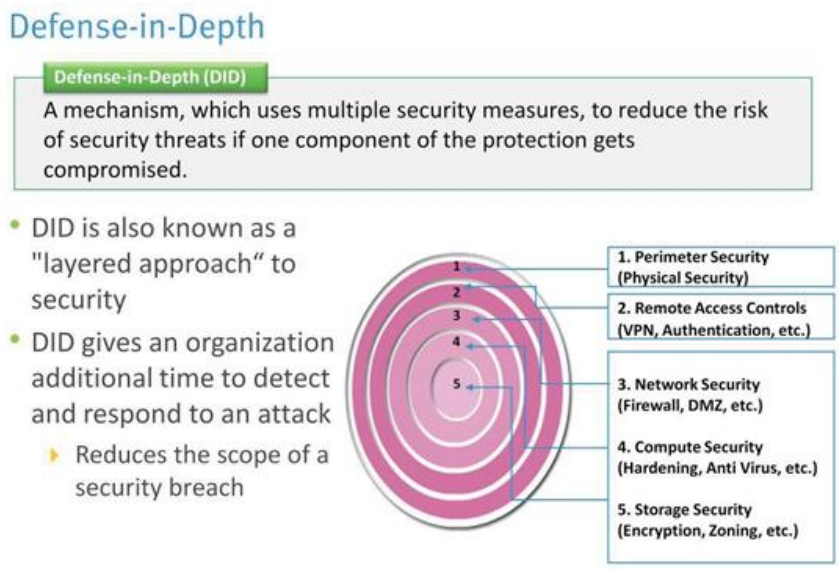
Multi-factor authentication considers multiple factors before permission to access a resource is

granted to the user. Typically, the factors considered are:

- **First factor:** What does a user know? For example, a password for a log on session will be what a user is required to know.
- **Second factor:** What does a user have? For example, a user needs to provide a secret key, generated by a physical device (token), which is under the user's possession.
- **Third factor:** Who is the user? For example, a biometric signature of a user can be considered as an example of who a user is.

Additional factors can also be considered, for example, "key phrases" unique to the user's past activity.

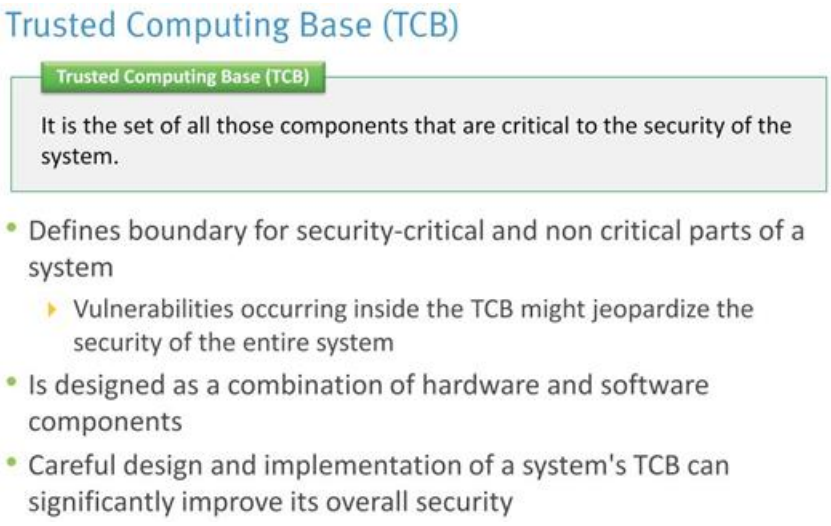
A multi-factor authentication scheme may consider any combination of these factors for authentication. User access is granted only when all the required factors are validated. For example, a two-factor authentication could be based upon the first and second factors discussed above and depicted in the diagram on this slide.



Defense-in-depth represents the use of multiple security defenses to help mitigate the risk of security threats if one component of the defense is being compromised. An example could be an antivirus software installed on individual VM when there is already a virus protection on the firewalls within the same environment. Different security products from multiple vendors may be deployed to defend different potential vulnerable resources within the network.

Defense-in-depth is an information assurance strategy in which multiple layers of defense are placed throughout the system. For this reason, it is also known as a "layered approach to security".

Because there are multiple measures for security at different levels, defense-in-depth gives additional time to detect and respond to an attack. This reduces the scope of a security breach. However, the overall cost of deploying defense-in-depth is often higher, compared to single- layered security mechanisms.

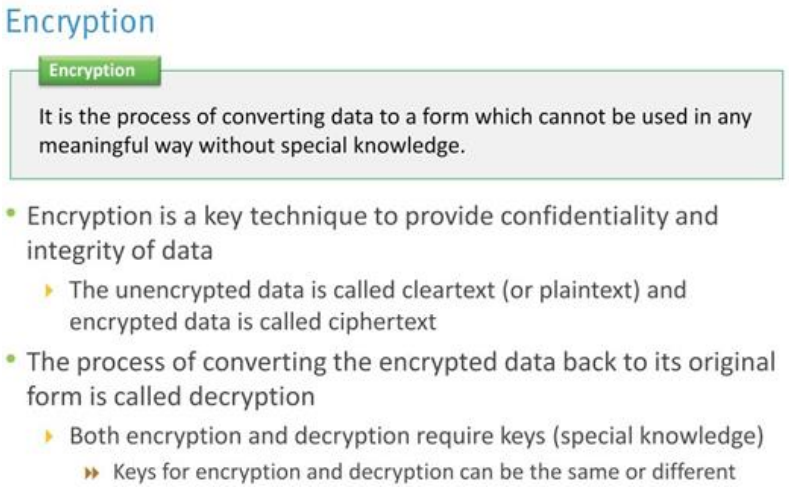


An important design concept while designing security framework for an information system is the concept of Trusted Computing Base (TCB). TCB of an information system is the set of

all components that are critical to its security, in the sense that vulnerabilities occurring inside the TCB might jeopardize the security of the entire system.

TCB is often the only part of the entire system which runs in a privileged mode, thus enhancing the inherent security of the system. TCB controls and authenticates access to the system resources and verifies system integrity. TCB is usually designed as a combination of hardware and software components. TCB essentially defines a boundary for security-critical

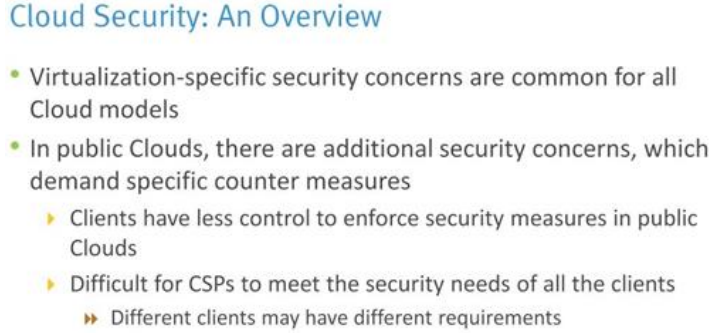
and non-critical parts of an information system. Careful design and implementation of a system's TCB is of great significance to its overall security.



Encryption is the process of converting data (information) to a form that cannot be used in any meaningful manner without special knowledge. Encryption is a key technique to enable the confidentiality and integrity of data. The non-encrypted data, which is given to the encryption process as an input, is called plaintext (cleartext) and the encrypted data, which is an outcome of the encryption process, is called ciphertext. The process of converting the encrypted data back into its original form is called decryption.

Encryption (and decryption) requires key(s) (special knowledge) or process to apply on data. When the keys for encryption and decryption are the same, it is known as symmetric encryption. When these keys are different (but related), it is known as asymmetric encryption. For data encryption, most often, symmetric encryption is used. Asymmetric encryption is most commonly used to secure separate end points of a connection, for example, Web browser and Web server (using https), VPN client and server, or for transferring a symmetric key.

This lesson covers security concerns and threats in a VDC and Cloud environment from Cloud Service Provider's (CSP's) and user's perspectives.



Security concerns, which arise specifically due to virtualization, are common to all Cloud models. As compared to VDC or Private Clouds, in Public Clouds, there are additional security concerns which demand specific counter measures. This is because, in a VDC or a Private Cloud, a client has complete control over the resources and can enforce security policies. In Public (and hybrid) Clouds, however, clients usually do not have that much control over resources and therefore, enforcement of security mechanisms for a client is comparatively

difficult. For CSPs also, it is not easy to enforce all the security measures that meet the security needs of all the clients, because different clients may have different security demands based upon their objective of using the Cloud services. In public Clouds, there are additional security concerns, which demand special measures.

From a security perspective, both Cloud users as well as service providers have several concerns and face multiple threats. Some of the concerns and threats are common to both of them. From a CSP perspective, majority of the concerns and threats are common to all Cloud-deployment models. Therefore, in cases where a security concern or threat is specific to a Cloud model (for example Public Cloud, it will be explicitly mentioned.

Security concerns and threats in a Cloud environment can be classified for CSPs and Cloud users based upon whom that concern or threat affects more.

For a CSP, the key security concerns are multitenancy and 'velocity-of-attack'. Though these are also concerns for Cloud users, it is the CSP who needs to adopt suitable counter measures to address these concerns.

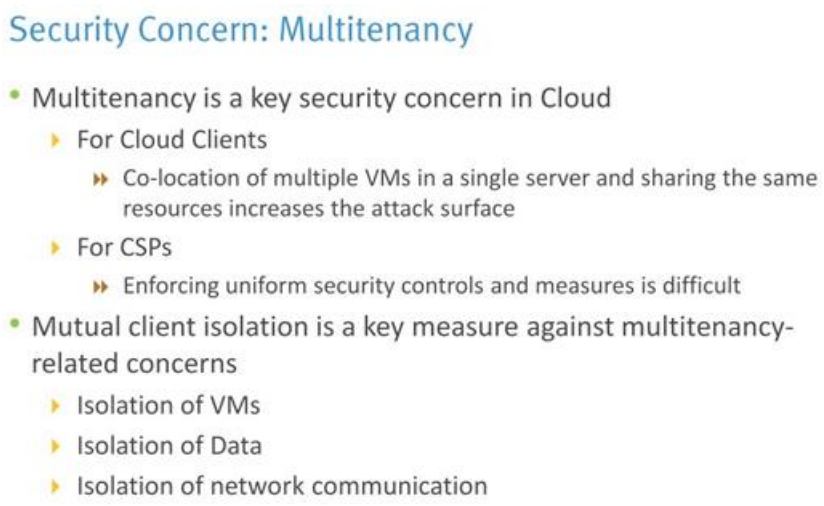
Multitenancy refers to the fact that the Cloud infrastructure, by virtue of virtualization, enables multiple independent clients (tenants) to be serviced using same set of resources. This consequently increases the risks for data confidentiality and integrity. These risks are especially more severe in case of Public Cloud environment. This is because, in Public Cloud, services can be used by competing clients as compared to Private Clouds. Also, the number of Cloud users are much higher in Public Clouds. The 'Velocity-of-attack' in the Cloud refers to a situation where any existing security threat in the Cloud spreads more rapidly and has larger impact than that in the Classic Data Center (CDC) environments.

Information assurance, data privacy, and ownership are among the key Cloud security concerns for its users. Information assurance covers many related aspects of ensuring that data in a Cloud is "safe". Data privacy and ownership concerns specifically relate to the risk of an unauthorized data disclosure.

The key security threats for VDC and Cloud infrastructure include:

- **VM theft:** It involves unauthorized copying or movement of a VM.
- **VM escape:** Guest OS or an application running on it breaks out and starts interacting directly with the hypervisor.
- **HyperJacking:** An attacker installs a rogue hypervisor or Virtual Machine Monitor (VMM) that can take complete control of the underlying server.
- **Data leakage:** Refers to the fact that confidential data of a client stored on a third party Cloud is potentially vulnerable to unauthorized loss or manipulation. It is primarily a threat for a Cloud user.
- **Denial of Service:** Denial of Service attacks prevent legal users of a system from accessing its services. In Cloud, it could occur when a malicious VM is installed on the same server and consumes all the server resources, thus preventing the other VMs from functioning properly.

These Cloud security concerns and vulnerabilities will be discussed next.



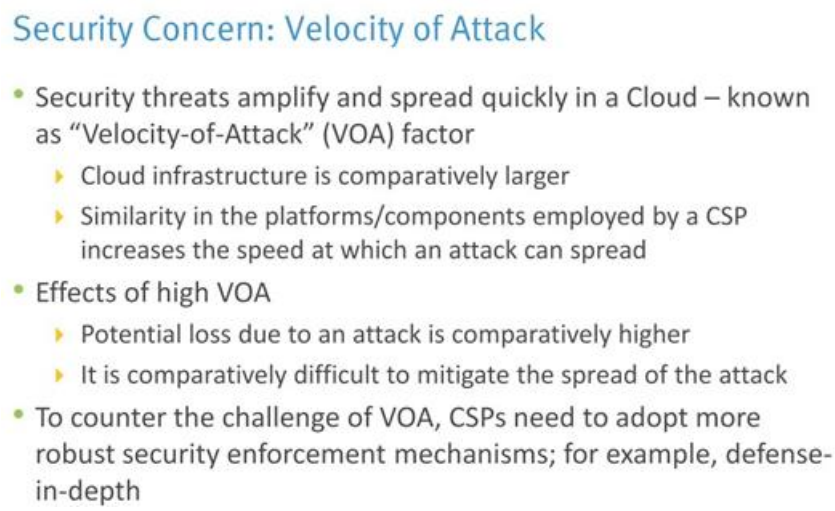
In spite of the benefits offered by multitenancy to a CSP, it is a major security concern for the Cloud clients. This is because Cloud infrastructure and services are, by nature, shared among multiple business entities, for example, multiple business units in an organization or different companies. Co-location of multiple VMs in a single server and sharing the same resources increase the attack surface. This gives rise to a potential security concern for the Cloud users because it makes private data of one client vulnerable to theft by other competing clients who run applications using the same resources. There also exists a danger that in the absence of adequate security controls, a tenant application might disrupt operations of other tenants, for example, by launching DoS attacks. An already compromised VM might enable an attacker to compromise the security of other VMs (running on the same server) or of the hypervisor. A compromised guest OS in a VM can also impact other guest OSs in other VMs.

For CSPs also, multitenancy makes it harder to enforce uniform security controls and counter

measures for all the clients.

Isolation of VMs, data, and network communication related to different clients is a key measure against multitenancy-related concerns.

Note: An attack surface refers to various access points/interfaces that an attacker can use to launch an attack; for example, the code within a compute system, which can be executed without requiring any authorization.

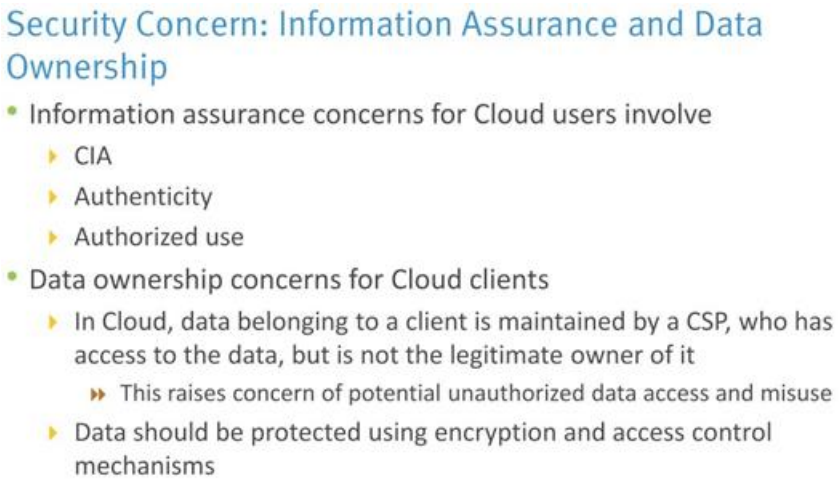


Clouds harness the power of relatively large compute, storage, and network infrastructure, compared to individual data centers. A Public Cloud might consist of thousands of physical servers, network connectivity across large geographic boundaries, and very high

capacity storage arrays. There also exists homogeneity (similarity) in the platforms and components (for example, the virtualization software, guest OS) employed by the Cloud service providers. Owing to these factors, security threats are amplified more and spread quickly, which is considered as the "velocity of attack" factor in the Cloud. Owing to this factor, security threats can cause much higher levels of losses to the Cloud service providers and to its clients, compared to those in a CDC. Mitigating the spread of a threat in a Cloud is also comparatively more difficult than in a CDC environment.

Because of the potentially high velocity-of-attack, CSPs need to adopt and deploy stronger and robust security enforcement and containment mechanisms, for example, defense-in-depth.

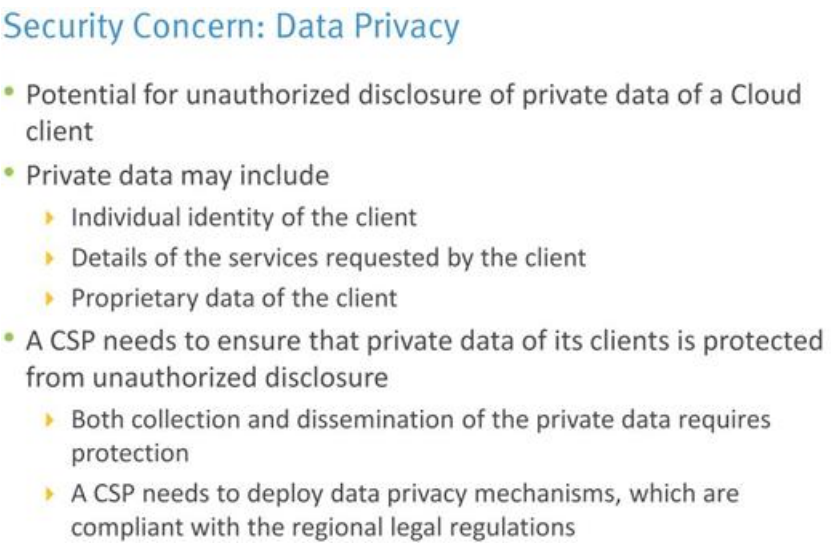
Cloud users also need to carefully assess the security services deployed by a CSP before moving operations to the Cloud.



The core information assurance concerns for Cloud users include:

- **CIA:** Ensuring confidentiality, integrity, and availability of data in the Cloud.
- **Authenticity:** Ensuring that all the users operating on the Cloud are genuine (i.e., their identities have not been fabricated).
- **Authorized Use:** Ensuring that the Cloud users (clients and CSP administrators) can operate only with legitimate rights and scope.

Ownership of data is yet another concern for Cloud clients. Ownership issues arise in public Clouds because data belonging to a client is maintained by a CSP who has access to the data but is not the legitimate owner of it. This raises concern of potential misuse of the data for the users because they do not have much control over the data handled by CSP. Data should be protected using encryption. The access control mechanisms for Clouds must be designed to ensure ownership based access rights.



Data Privacy is a major concern in Cloud. A CSP needs to ensure that Private and Personally Identifiable Information (PII) about its clients is legally protected from any unauthorized disclosure.

Private data may include:

- Individual identity of a Cloud user
- Details of the services requested by a client
- Proprietary data of the client

A CSP needs to ensure that private data of its clients is protected from unauthorized disclosure. Both collection and dissemination of the private data needs protection from any possible unauthorized disclosure, in particular, as per the regional legal requirements.

Security Threat: VM Vulnerabilities

- VMs are vulnerable to attack when they are running and when they are powered-off
  - A powered-off VM is still available as an image file, which is susceptible to malware infections
  - Unprotected VM migration is vulnerable to network attacks
  - Encryption of VM image files is required as a protection measure when it is powered-off or during its migration
- VM templates are also vulnerable to attacks
  - When new unauthorized VMs are created from a template
  - When a template is modified to infect the VMs
  - To protect, VM templates must be kept encrypted and access should be restricted to privileged users (administrators)

VM templates are vulnerable when they are running and when they are powered-off. A powered-off VM is still available as a VM image file that is susceptible to malware infections and patching. If adequate security measures are not deployed, VM migration could expose the migrating VM to various network attacks, for example, eavesdropping and modification.

VM templates are also vulnerable to attacks; for example, by creating new unauthorized VMs from a template or modifying the templates to infect the VMs that will be provisioned using those templates.

Protecting VMs from these vulnerabilities require encryption of the VM image files when they are powered off, or while they are migrated. This is beyond the well-defined isolation of VMs, which is generally performed at the hypervisor level. Also, access to the VM templates should be restricted to a limited group of privileged users.

Security Threat: VM Theft

VM Theft

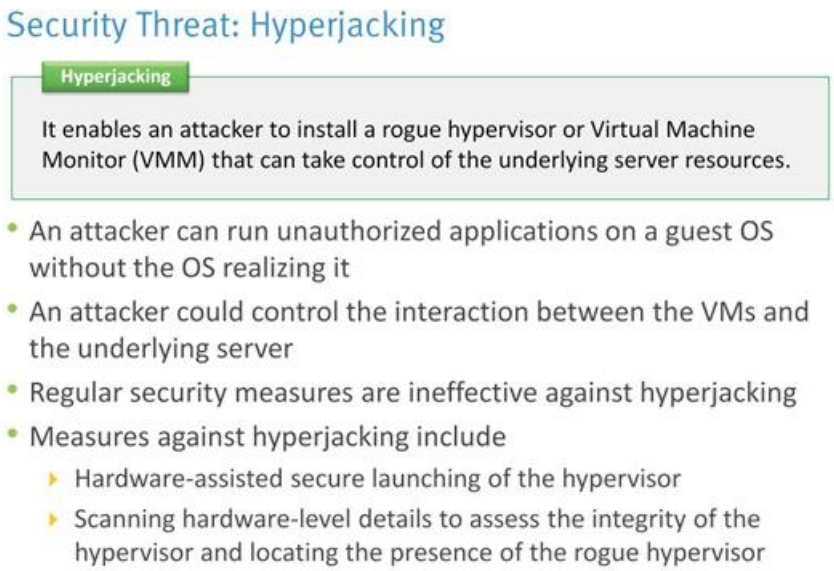
A vulnerability that enables an attacker to copy or move a VM in an unauthorized manner.

- Is a result of inadequate controls on VM files allowing unauthorized copies or move operations
- Copy and Move restrictions are essential to safeguard against VM theft
  - These restrictions bind a VM to a specific physical machine
    - A VM with copy and move restriction cannot run on a hypervisor installed on any other server
  - These restrictions use a combination of virtualization management and storage management services for effective enforcement
  - Limit applying such restrictions to critical/sensitive VMs only

VM theft enables an attacker to copy and/or move a VM in an unauthorized manner. VM theft is a result of inadequate controls on VM files allowing their unauthorized copy or movement. VM theft can cause a very high degree of loss to a Cloud client if its files and data are sensitive in nature.

Copy and Move restrictions are essential to safeguard against VM theft. Such restrictions effectively bind a VM to a specific (secure) physical machine so that even if there is a forceful copy of the VM, it will not operate on any other machine. A VM with copy and move restrictions cannot run on a hypervisor installed on any other machine. These restrictions use a combination of virtualization management and storage management services for their effective enforcement. Even though these restrictions are essential to safeguard against VM theft, they, on the other hand, might limit the benefit of load balancing of VMs across physical servers. Therefore, it is advisable that copy and move restrictions be applied in a limited way, especially only on those VMs, which are considered security critical/sensitive or are relatively more vulnerable for theft.

Apart from VM theft, another threat on the VM level is known as 'VM escape'. Normally, virtual machines are encapsulated and isolated from each other. There is no straightforward way for a guest OS and the applications running on it to break out of the virtual machine boundary and directly interact with the parent hypervisor. The process of breaking out and interacting with the hypervisor is called a VM escape. Since the hypervisor controls the execution of all VMs, due to VM escape, an attacker can gain control over every other VM running on it by bypassing security controls that are placed on those VMs.



Hyperjacking is a rootkit level vulnerability that enables an attacker to install a rogue hypervisor or virtual machine monitor that can take complete control of the underlying physical server. A rootkit is a malicious program which is installed before an hypervisor or VMM is fully booted on a physical server. This way, a rootkit runs with privileged access and remains invisible to the administrators. After a rootkit is installed, it allows an attacker to mask the ongoing intrusion and maintain privileged access to the physical server by circumventing normal authentication and authorization mechanisms employed by an OS.

Using such rogue hypervisor, an attacker can run unauthorized applications on a guest OS without that OS realizing the presence of such an application. With hyperjacking, an attacker can control the interaction between the VMs and the underlying physical machine. Regular security measures are ineffective against this rough hypervisor because:

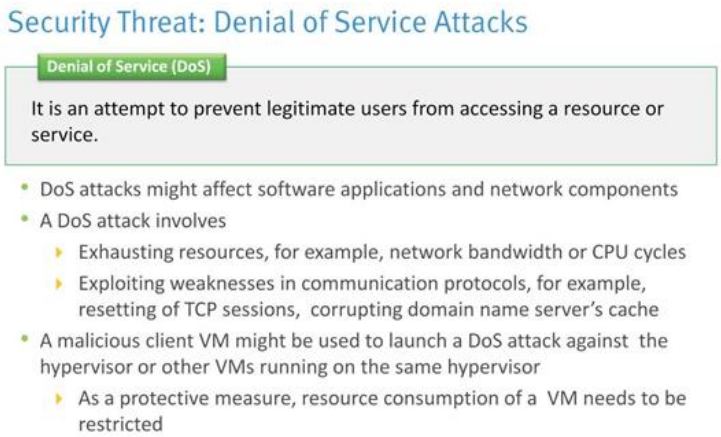
- Guest OS would remain unaware of the fact that the underlying server has been attacked and
- The antivirus and firewall applications cannot detect such rogue hypervisor because they are installed over the server itself Measures against hyperjacking include:
- Hardware-assisted secure launching of the hypervisor so that rootkit level malicious

programs cannot launch. This involves designing and using a TCB so that the hypervisor gets support at the hardware level itself.

- Scanning the hardware-level details to assess the integrity of the hypervisor and locating the presence of rogue hypervisor. This scanning may include checking the state of the memory and registers in the CPU.

Confidential data stored on a third party Cloud is potentially vulnerable to unauthorized loss or manipulation. Third party Cloud refers to the Cloud that is used by a service provider to provide services to its end clients. In such a case, the service provider might not have total control over the Cloud to ensure the confidentiality and integrity of clients' data. An attack on the service provider's control systems (for example, password lists) could make all of the clients' data vulnerable and expose the data to malicious uses. To mitigate the risk of such data theft, cloud users must evaluate the end-to-end data protection measures adopted by all the concerned parties that have any level of access on their data.

"Side Channel Attacks" (SCA) can also be used to check data leakage in Cloud. An SCA extracts information by monitoring indirect activity, for example, keystroke activity, cache data, etc. A cross VM SCA involves one VM being used to launch an SCA on another VM running on the same server. A cross VM SCA could reveal information on a client's critical business activity to a malicious client that runs its VMs on the same server. Protection against such cross VM SCA requires a placement policy that permits clients to install only non-conflicting VMS on the same server. This will, however, reduce resource optimization that virtualization technology offers. It could also add extra service cost to the Cloud clients who demand such privileged services.



A Denial of Service (DoS) attack is an attempt to make resources or services of an information system unavailable to its authorized users. A DoS attack prevents legitimate users from accessing a resource or service. DoS attacks could be targeted against software applications (example OS) and network components including routers or servers.

Often a DoS attack involves either one or a combination of the following:

- Attack aims to exhaust computing resources, for example, network bandwidth and CPU cycles. An attack may involve massive quantities of data sent to the target with the intention of consuming bandwidth/processing resources.
- Attack involves exploiting weaknesses in a protocol to target network resources; for example, resetting of TCP sessions, IP address spoofing, or corrupting DNS server cache.

A Distributed DoS (DDoS) attack is a special type of DoS attack in which several systems launch coordinated DoS attack on their target(s), thereby causing denial of service to the users of the targeted system(s). In a DDoS attack, the main attacker is able to multiply the effectiveness of the DoS attack by harnessing the resources of multiple collaborating systems. These collaborating systems serve as attack platforms. Typically a DDoS master program is installed on one compute system using a stolen account. Then, at a designated time, the master program communicates to any number of "agent" programs installed on computers anywhere on the network. When the agents receive the command, they initiate the attack.

In a virtualized environment, a rough VM could be used to launch DoS attack against the hypervisor or other VMs running on the same hypervisor. Such a rough VM could use the internal virtual network for launching the DoS attacks. To protect against such VM based DoS attacks, the resource consumption of a VM should be restricted to specific limits.

Module 10: Cloud Security

Lesson 3: Security Mechanisms

Topics covered in this lesson:

- Security at compute level, including securing server, hypervisor, VM, guest OS, and applications
- Security at network and storage levels, including virtual firewall, demilitarized zone, and data shredding
- Intrusion detection in VDC and Cloud
- Physical Security of the premises
- Access Control and Identity management services in Cloud

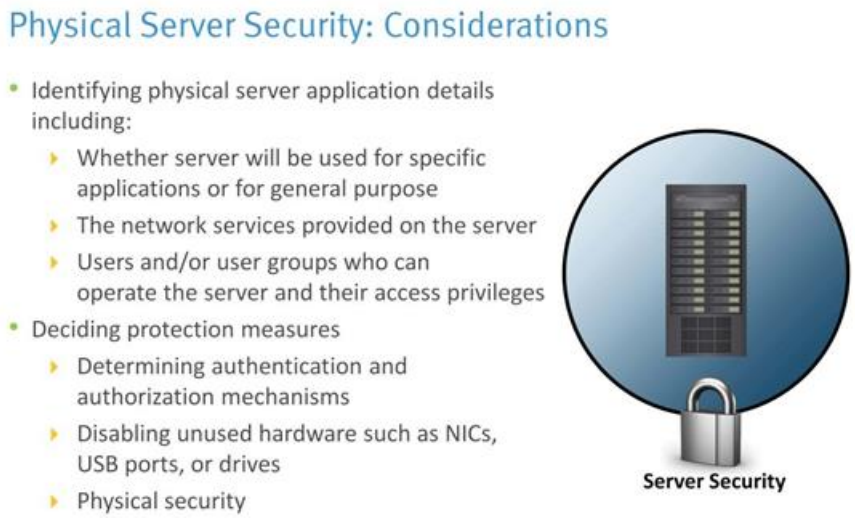
This lesson covers Cloud infrastructure security mechanisms at the compute, network, and storage levels. Under compute level security, the lesson discusses security for server, hypervisor, VM, guest OS, and applications. Next, security at the network level discusses virtual firewall, demilitarized zone, and intrusion detection. Also covered are aspects related to storage security, including securing data-at-rest and data shredding. Next, the lesson discusses measures for physical security of the premises. Finally, the lesson discusses access control mechanisms, including role based access control and identity management techniques such as one-time password, identity federation, and OpenID.

Security at Compute Level

- Securing a compute system includes
  - Securing physical server
  - Securing hypervisor
  - Securing VMs
    - VM Isolation
    - VM Hardening
  - Security at guest OS level
    - Guest OS Hardening
  - Security at application level
    - Application Hardening

Securing a compute infrastructure includes enforcing security of the physical server, hypervisor, VM, and guest OS. Security at the hypervisor level primarily aims at securing

hypervisor from the rootkits and malware based attacks and protection of the hypervisor management system. VM isolation and hardening are two key techniques for securing VMs. Security at the guest OS level uses sandboxing and hardening as two key methods. Application hardening is used to reduce vulnerability of the applications from getting exploited by malicious attackers. All these security methods are explained in the following slides.



Server security considerations include identifying server application details such as:

• Deciding whether the server will be used for specific applications (for example backup applications) or for general purpose.

• Identifying the network services to be provided on server; for example, LAN connection, wireless connection, etc.

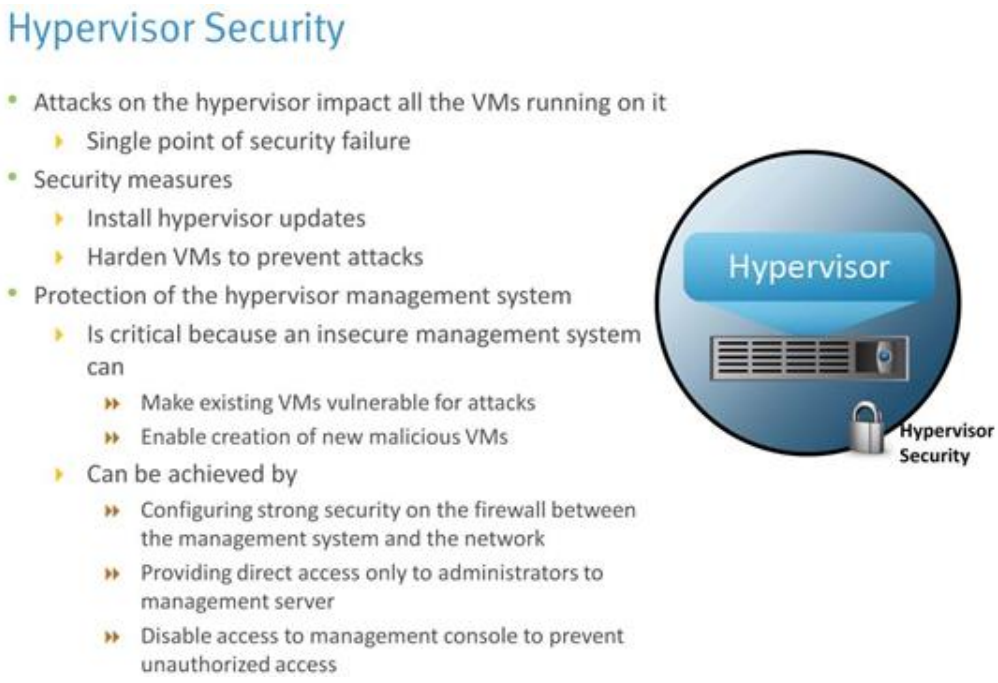
• Identifying users and/or user groups who will be given access rights on the server. This also includes determining their specific access privileges.

Based upon these details, suitable protection measures need to be decided including the following:

• Determine user authentication and authorization mechanisms.

• If the server has unused hardware components such as NICs, USB ports, or drives, they should be removed or disabled. This should also be done in the VM (template).

• Adequate physical security protection including safety of the premises where the server will be housed.

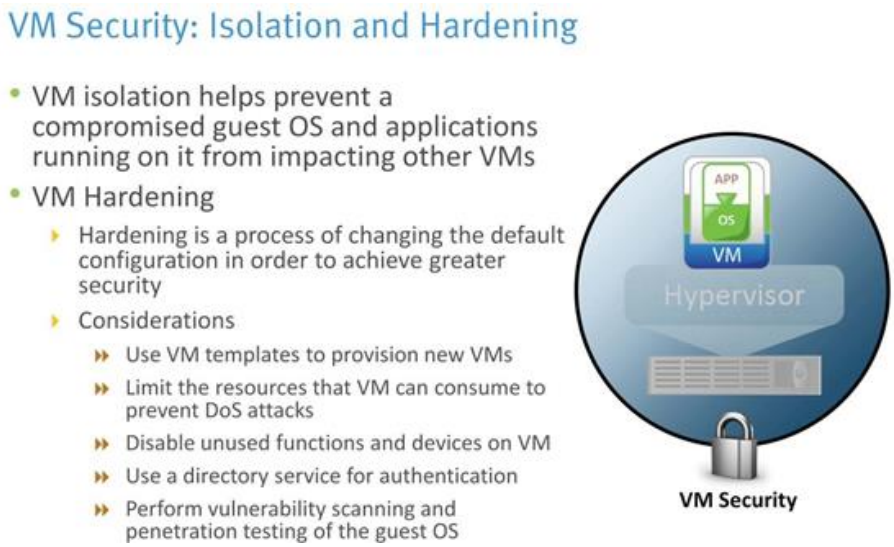


Hypervisor in a virtualized environment presents a single point of security failure for all

the VMs running on it. A single breach of the hypervisor places all the guest OSs on these VMs at high risk. Rootkits and malware installed below the OS make detection difficult for the antivirus software installed on the guest OS. To protect against attacks, security-critical hypervisor updates should be installed at the earliest possible and the VMs hosted on it should be hardened (VM hardening measures are discussed next). Hypervisor services like clipboard, if not used, should be disabled.

The hypervisor management system must be protected. Malicious attacks and infiltration to the management system can impact all the existing VMs and allow attackers to create new VMs.

Access to the management system should be restricted only to authorized administrators. Levels of access should be restricted to selected administrators. Furthermore, there must be a separate firewall with strong security installed between the management system and the rest of the network. Yet another measure is to disable access to the management console to prevent unauthorized access.



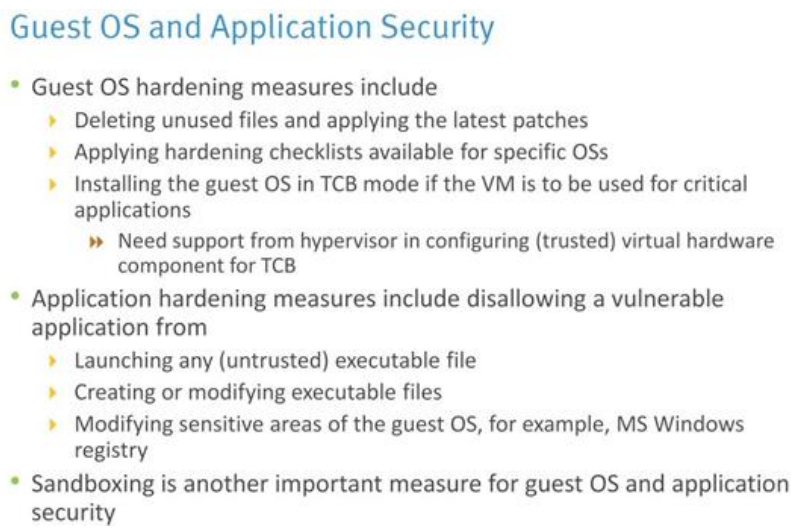
VM isolation is a key measure that helps in preventing a compromised guest OS from impacting

other guest OSs. VM isolation is implemented at the hypervisor level.

Apart from isolation, VMs should be hardened against security threats. Hardening is a process of changing the default configuration in order to achieve greater security. Some of the key measures for hardening a VM (especially for a VDC or private Cloud environment) include the following:

- Use VM templates to deploy VMs. When using templates, harden the VM image so that all the deployed VMs have a known security baseline. VM templates should be created with up-to-date VM patches and security updates.
- In order to avoid DoS attacks, the VM management software should limit the VM's resources so that a single VM is not allowed to consume all of the server's resources.
- Increase in the number of services, ports, and applications running on the VM also increases the area of attack surface. Therefore, unneeded functions and unused devices should be disabled.
- Configure access permissions for the selected group of administrators. Avoid account sharing by groups of users and strictly control root privileges. Employ directory based authentication to ensure the genuineness of credentials.
- Take VM backups on a regular basis and schedule point-in-time snapshots to restore a VM to a safe state, in case of an attack.
- Perform vulnerability scanning of the guest OS regularly to identify existing vulnerabilities. Perform a penetration test to determine the feasibility of an attack and the extent of business impact of the attack. Note that in Public Clouds, usually, penetration tests originating from outside the CSP network are forbidden by the CSP. Therefore, a Cloud user

should rely upon the CSP to perform these tests.



Apart from the measures to secure a hypervisor and VMs, VDC and Cloud environment also require further measures on the guest OS and application levels. Hardening is one such important measure which can effectively safeguard guest OS and the applications running on it.

OS hardening, for example, may involve actions such as configuring system and network components, deleting unused files, and applying the latest patches. There are hardening checklists available for major OSs which administrators should follow to harden the guest OSs deployed in VDC or Cloud. In cases where a VM is to be used for business critical applications, the guest OS should be installed in the TCB mode – an option available with many OSs. Such TCB mode installation, however, requires hypervisor support for configuring trusted virtual hardware components (for example virtual CPU).

Application hardening helps to prevent exploitation of vulnerabilities in software applications that have not been patched so far. The key steps for hardening a vulnerable application include:

- **Disallowing the application from Spawning executable files:** One of the methods used by attackers is to make a vulnerable application into spawning executable files of their choice. Therefore, an important action for hardening the application is to disallow it from launching other executables, except those that are trusted.
- **Disallowing the application from creating or modifying executable files:** Another technique, which is used by attackers, is to convert the vulnerable application into modifying or creating executable files of their choice in order to insert the malicious code into the system. The malicious code may eventually be executed and activated. Therefore, it is critical that applications are not allowed to modify or create any executable file when they are running.
- **Disallowing the application from modifying Sensitive Areas:** It involves disallowing the application from modifying sensitive areas of the guest OS, for example, registry keys in the Windows OS.

Apart from hardening, sandboxing is yet another important measure for guest OS and application security.

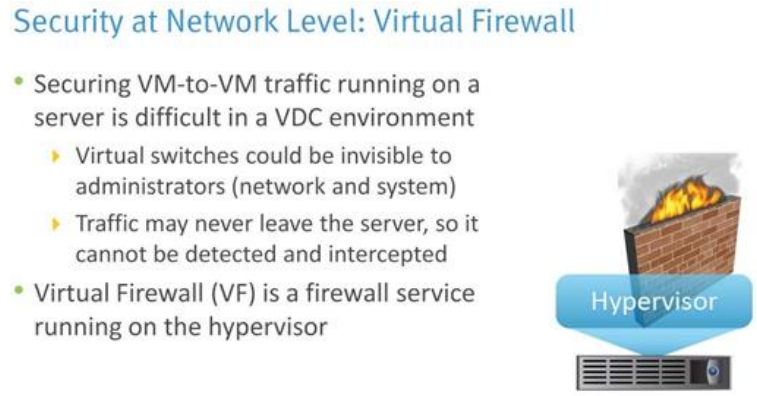
Sandboxing involves isolating execution of an application from other applications in order to restrict the resources that the application can access and the privileges it can have.

A sandbox is used for separating the execution of an untrusted application from unverified third- parties, suppliers, and untrusted users. A sandbox provides a tightly-controlled set of resources for the application to execute, such as scratch space on disk and memory. Network access and the ability to inspect the system components or read-from-input devices are either disallowed or restricted and closely monitored.

Sandboxing should be applied on a vulnerable or suspected guest OS or application. Guest OS sandboxing is achieved on the hypervisor level or at the OS kernel level, so that OS services or the malicious software can be subjected to the sandbox constraints. Because of sandboxing,

if one guest OS crashes due to an application fault or an attack, the other guest OSs running on that server will remain unaffected.

Sandboxing can also be as a security measure against side-channel-attacks because it disallows a malicious software from monitoring system components.

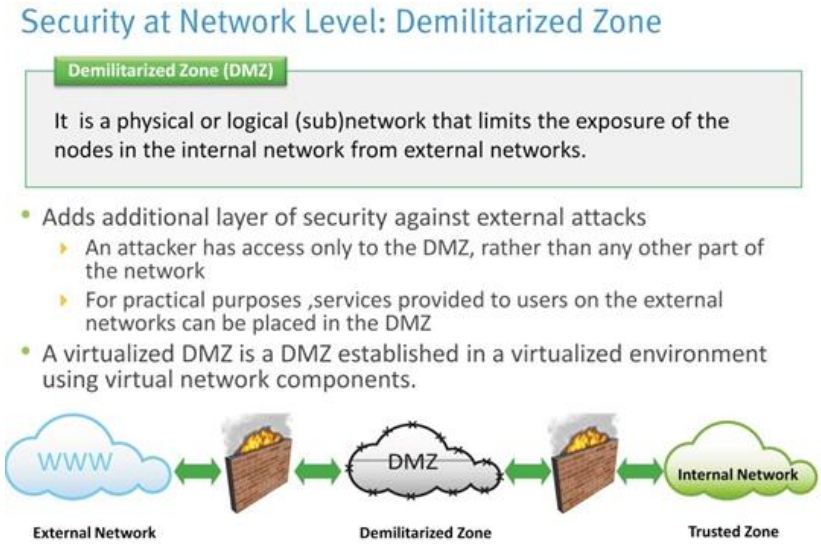


A firewall is a security technology designed to permit or deny network transmissions based upon a set of rules. A firewall is implemented on a compute level and limits access between networks and/or systems in accordance with a specific security policy. A firewall is used to protect networks from unauthorized access while permitting only legitimate communications.

In a VDC and Cloud infrastructure, a firewall can also be used to protect hypervisors and VMs; for example, if remote administration is enabled on a hypervisor, access to all the remote administration interfaces should be restricted by a firewall.

Securing the VM-to-VM traffic running on a server is a key security problem in a VDC environment. Securing this virtual network is a considerable challenge because virtual switches could be invisible to network and/or system administrators, who usually enforce security at the network level. Because the virtual network traffic may never leave the server, security administrators cannot observe VM-to-VM traffic, cannot intercept it, and so, cannot know what that traffic is for. Thus, logging of the VM-to-VM network activity within a single server and verification of virtual machine access for regulatory compliance purposes is relatively difficult. Inappropriate use of virtual network resources and bandwidth consumption in VM-to-VM are difficult to monitor or rectify.

Therefore, a firewall service, which can be used to monitor and control VM-to-VM traffic, is critically required. It is provided by a Virtual Firewall (VF), which is a firewall service running entirely on the hypervisor. VF provides the usual packet filtering and monitoring of the VM-to- VM traffic. VF gives visibility and control over VM traffic and enforces policies at the VM level.

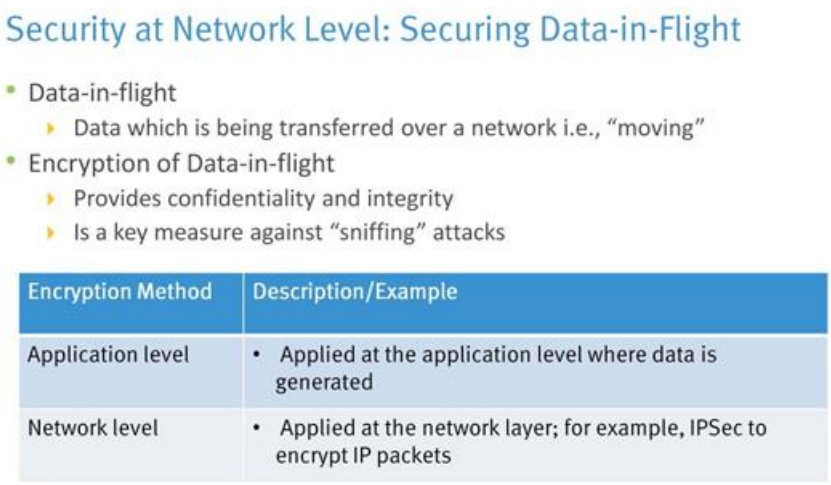


In a network, the nodes (compute systems) that are most vulnerable to an attack are

those that provide services to users outside of the network; for example, e-mail and Web servers. Therefore, these nodes are placed into their own sub-network in order to protect the rest of the network from intruders. Such a sub-network is known as Demilitarized Zone (DMZ), which is isolated from the rest of the network.

DMZ adds an additional layer of security because an external attacker has access only to the DMZ interface. A firewall controls the traffic between the DMZ nodes and the internal network clients. The remaining (internal) network, apart from the DMZ, which does not directly interface with the external network, is commonly referred to as trust zone. Nodes in the trust zone are not exposed to the external networks. For practical purposes, any service that is provided to users on the external network can be placed in the DMZ.

A virtualized DMZ is a network DMZ established in a virtualized environment using virtual network infrastructure. A virtualized DMZ network can fully support and enforce multiple trust zones.



Data-in-flight refers to the data transferred over a network, and means that the data is "moving".

Encryption of data-in-flight is the key method for providing confidentiality and integrity services. Encryption makes the data indecipherable to an unauthorized user who otherwise may have access to the (encrypted) data. Encryption is indeed a key security measure against "sniffing" attacks. In a sniffing attack, a non-recipient malicious device/user accesses the data transmitted over the network.

Methods used for encrypting data-in-flight include:

- **Application-level encryption:** Encryption is applied at the application level where the data is generated. Encrypting at the application level protects data against unauthorized access; for example, Transport Layer Security (TLS) protocol allows client/server applications to enforce encryption service.
- **Network-level encryption:** Encryption is applied at the network layer; for example, IPSec, which can be used to encrypt and authenticate each IP packet transmitted over an IP network. The benefit of a network level encryption is that it is independent of the underlying guest OS.

Intrusion Detection

Intrusion Detection (ID)

It is a process of detecting events and/or entities that could possibly compromise the security of the system.

Types of Intrusion Detection System (IDS)	Description
Server based IDS	• Analyzes activity logs, including system calls, application logs, etc. • Better view of the monitored system but high vulnerability for an attack on IDS itself
Network based IDS	• Analyzes network traffic and communicating nodes • Poorer view of the system and low vulnerability for an attack on IDS itself
Integrated IDS	• Combination of server and network based approaches

Intrusion Detection (ID) is the process of detecting events that can compromise the confidentiality, integrity, or availability of a resource. An ID System (IDS) automatically analyses the events to check whether an event or a sequence of events match a known pattern for anomalous activity, or if it is (statistically) different from most of the other events in the system. It generates an alert if an irregularity is detected.

There are three main types of technologies currently in use for ID:

- **Server Based IDS:** It analyses activity logs including system calls, application logs, file- system modifications, etc. IDS is installed as another application and executes on the host computer. IDS has good visibility of the state of the monitored system and can analyze applications running on the computer. But the server based IDS is vulnerable to an attack itself because it is running on the same computer with the malicious application and therefore, is less isolated for an attack.
- **Network Based IDS:** It analyses network traffic and communicating nodes. It can monitor network traffic for port scans, DoS attacks, known vulnerabilities, etc. IDS resides on the network and therefore, is relatively isolated from malicious applications. Hence, it is relatively less vulnerable to attacks. But a network based IDS has poor view of the state of the monitored systems, especially on the activities of malicious applications running on these systems. If the network traffic is encrypted, there is no effective way for the Network based IDS to decrypt the traffic for analysis.
- **Integrated IDS:** It uses a combination of server and network based methods. It analyses both system activity logs and network traffic.

Firewalls limit the access between networks to prevent intrusion. In comparison to that, an IDS evaluates a suspected intrusion that may have already taken place and signals an alarm.

Intrusion Detection in VDC and Cloud

- ID in a VDC environment
  - ID at Guest OS level
  - ID using separate VM
  - ID at hypervisor level
  - ID at virtual network level
  - ID at physical network level
- ID in Cloud
  - ID in a SaaS Model
    - Provided by the CSP
  - ID in a PaaS Model
    - ID at local level provided by the CSP
  - ID in a IaaS Model
    - ID is set up by the client

In a VDC environment, ID can be performed at different levels:

- **ID at the Guest OS level:** It allows monitoring the activity of the applications running on the guest OS and detecting and alerting on issues that may arise.
- **ID using Separate VM:** A separate VM is used for ID that has access to all the traffic between a group of VMs and that can communicate with each other.
- **ID at the hypervisor level:** It allows monitoring not only the hypervisor but anything traveling between the VMs on that hypervisor. It is a more centralized location for an ID on a single server, but there may be issues in keeping up with performance or dropping some information if the amount of data is too large.
- **ID at the virtual network level:** Deploying ID to monitor the virtual network running within a single server allows monitoring the network traffic between the VMs on the server as well as the traffic between the VMs and the server.
- **ID at the physical network level:** It allows monitoring, detection, and alerting traffic that passes over the traditional network infrastructure. However, it cannot help when it comes to attacks within a virtual network that runs entirely within the hypervisor.

How much a client can implement and control ID in Cloud depends upon the underlying Cloud service model:

- **ID in a SaaS Model:** SaaS clients must rely almost exclusively on their CSPs to perform ID. Clients may have the option of getting some logs and deploying a custom monitoring and alerting on that information, but most ID activities will be done by their CSPs.
- **ID in a PaaS Model:** Since IDS typically run independent of any application, a client must rely on its CSP to deploy IDS in a PaaS. A client can, however, configure its applications and platforms to log onto a central location where it can then set up monitoring and alerting (i.e., where ID can be performed).
- **ID in a IaaS Model:** This is the most flexible model for ID deployment by a client. Unlike the other two service models, IaaS gives a client more options as a consumer.

Note that the IDS/IDP provided by a CSP would effectively be transparent to the client. Therefore, the client needs to perform IDS/IDP as feasibly as possible. By choosing not to do so may result in vendor lock-in (means dependence on the CSP services).

Major threats to storage system in a VDC and Cloud environment arise due to compromises at compute, network, and/or physical security levels. This is because an access to storage systems needs to be made by using compute and network infrastructure. Therefore, adequate security measures need to be in place at compute and network levels to ensure storage security. Storage Area Networks (SAN) have their own unique vulnerabilities which can be used to compromise their integrity. These include unauthorized device gaining fabric connection, DoS attack, WWN spoofing which would enable a device to masquerade as a

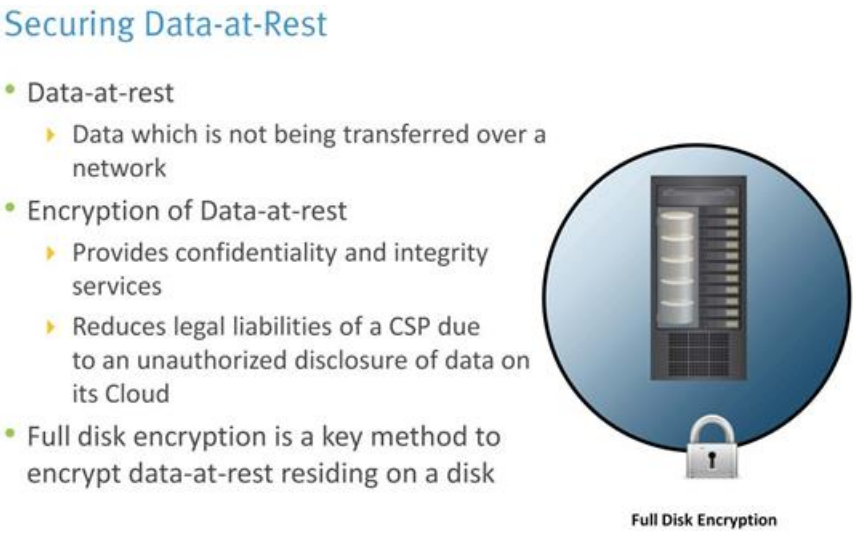
different entity, zoning bypass, etc.

Security mechanisms that might help to protect storage includes:

- Access control methods to regulate which users and processes access the data on the storage systems.
- Zoning and LUN masking
- Encryption of data-at-rest (on the storage system) and data-in-transit. Data encryption should also include encrypting backups and storing encryption keys separately from the data.
- WWPN and WWNN LUN masking for restricting access to the storage arrays in a SAN.
- Data shredding, which removes the traces of the deleted data. It may be performed by both CSP and also by the client.
- Backup and recovery in case of loss of data.

Apart from these mechanisms, physical isolation of storage devices from the other hardware and also an isolation of storage traffic from other types of traffic using VLANs and FC zoning would help in protecting storage.

Moreover, security protection is required for the storage utilized by the hypervisor for VMs, for example, a CFS supporting multiple VMs within a cluster. This may include using separate LUNs for VM components and for VM data and segregating VM traffic from hypervisor storage and management traffic.



Full Disk Encryption

Data-at-rest refers to the data which is not being transferred over a network i.e., is “not moving”. It includes data that resides in databases, file systems, flash drives, memory, networked storage, etc.

Encryption of Data-at-rest is the key method for providing confidentiality and integrity. Encryption makes the data indecipherable to unauthorized users. Encryption also reduces legal liabilities of a CSP due to an unauthorized disclosure of data on its Cloud because even if the encrypted data becomes accessible to an unauthorized user, it cannot be used in any meaningful way.

Full disk encryption is the key method used for encrypting data-at-rest residing on a disk. Full disk encryption employs software application or built-in hardware capability to encrypt every bit of data that goes on a disk or disk volume. Disk encryption thus prevents unauthorized access to data storage. For additional security, it can be used in conjunction with file system encryption.

Full disk encryption may use either a single key for encrypting the complete disk or different keys for encrypting different partitions.

Data Shredding

- Data which is deleted by a Cloud client or a process, but which leaves traces on the system, can be a potential source of attacks
  - Traces of deleted VMs can provide vital information to an attacker
  - Partially recoverable "deleted data" may reveal client details
- Data shredding permanently removes all the traces of the deleted data
  - Is a critical feature for data security in a Cloud infrastructure
  - Traces of the deleted data include
    - Logs of VM or application executions
    - Logs of old files, folders, and other resources
    - Logs of data communication

Unlike data stored in a privately controlled storage or in a CDC, data and information in a Cloud remain vulnerable even if deleted by the client or a process. This is because this data and information may still have recoverable traces about it on the system, and therefore, can be a potential source of attack. For example, traces of deleted VMs can provide vital information to an attacker about their clients. Partially recoverable "deleted data" from a Cloud storage may also reveal client details.

Data shredding is therefore a critical measure for data security in a Cloud infrastructure. A data which is shredded cannot be recovered any more. Data shredding removes all the traces of the deleted data including:

- Logs of deleted VMs including its configuration files and application executions
- Logs of old files, folders, and other resources
- Logs of data communication involving deleted VMs

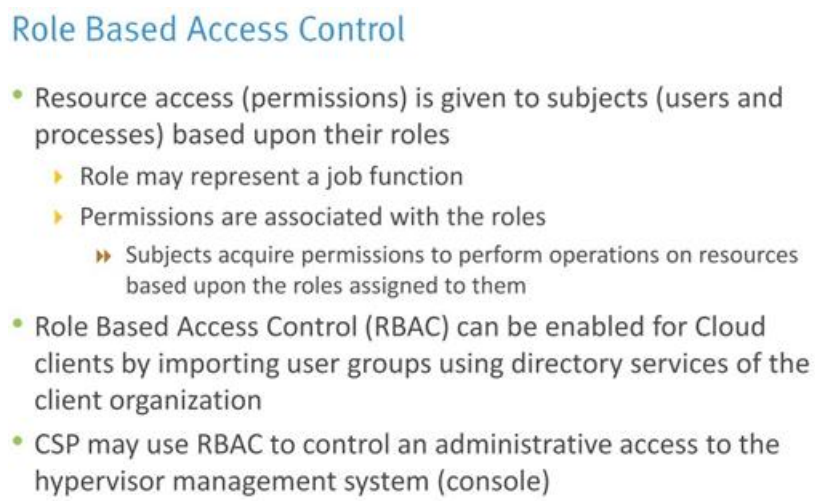
Physical Security in VDC and Cloud

- Restricted Port Access
  - Leave unused ports in disabled state
  - Bind specific devices to designated ports
- CCTV based video surveillance
- 24/7/365 onsite guarded security
- Biometric authentication based physical access

Cloud customers essentially lose control over physical security when they move to the Cloud, because the actual servers can be anywhere the provider decides to put them. Since physical Cloud infrastructure supports many Cloud clients together, its security is very critical both for the CSP as well as its clients. Policies, processes, and procedures are critical elements of successful physical security that can protect the equipment and data housed in the hosting center.

Typical security measures that must be in place for securing physical Cloud infrastructure include:

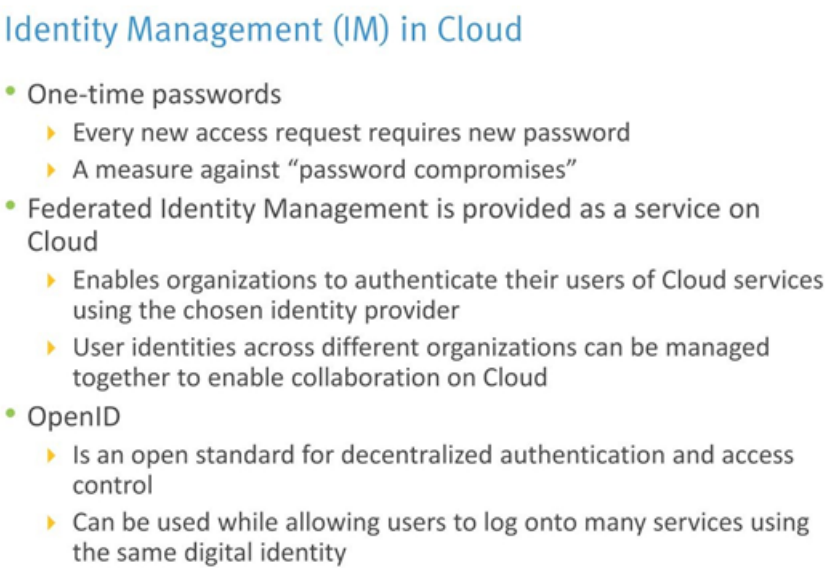
- Leaving a port in unconfigured or disabled state so that unknown devices or components cannot connect to the infrastructure. Additionally, bind specific devices to designated ports. Apply MAC/WWPN binding and VLAN restrictions to physical Ethernet switches.
- 24/7/365 onsite security for the premise where the Cloud physical infrastructure is hosted
- Biometric authentication based access to the premises
- Closed circuit TV cameras to monitor activity throughout the facility



In a Role Based Access Control (RBAC) model, resource access rights (permissions) are given to subjects (users and processes) based upon their roles. A role may represent a job function, for example, an administrator. Permissions are associated with the roles and subjects are not given any direct permissions. Subjects acquire permissions to perform operations on resources based upon the roles assigned to them.

For RBAC, users need to be grouped together into roles. As an example, a set of IT administrators can be given permissions to start, stop, and delete VMs that are running in the Cloud. However, there could be a specific subset of production servers that even the IT administrators are not allowed to control.

To handle sensitive data and compliance requirements, a Cloud needs RBAC capabilities. To achieve this, user groups can be imported using LDAP based Directory Services of the client organization for client installations in Cloud. CSPs may also use RBAC to control administrative access to the hypervisor management system (console).



Identity management is an administrative process that deals with identifying users of an information system. Additionally, identity management also controls access to system resources by placing restrictions using user identities . The key identity management-related aspects in Cloud are as follows:

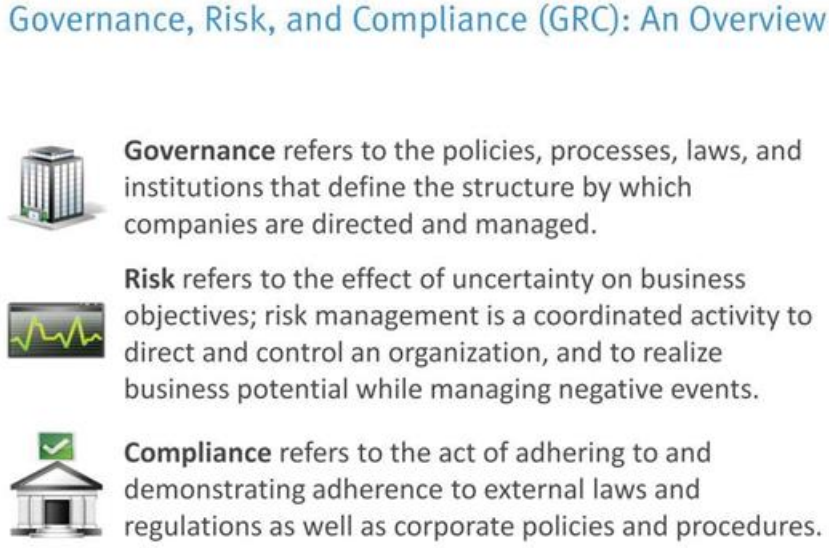
• **One-time Password:** Because passwords can be compromised, they must be protected. The One-Time Password (OTP) concept demands that a new password be used for each new log on and thus provides necessary security against password compromises. Time-dependent password generated by hardware security tokens (for example, RSA SecureID) is an

example of OTP. OTP is specifically useful in situations where likelihood of password compromises are high, for example, remote login based network access.

• **Federated Identity Management:** Federation is the process of managing the trust relationships among distinct organizations beyond the internal network or administrative boundaries. A federation is an association of organizations that come together to exchange information about their users and resources to enable collaborations and transactions. In a Cloud, Federated Identity Management (FIM) can play a vital role in enabling organizations authenticate their users of Cloud services using the organizations' chosen identity provider. This would involve exchanging identity attributes between the CSP and the identity provider in a secure way.

• **OpenID**: It is an open standard that describes how users can be authenticated in a decentralized manner, obviating the need for services to provide their own ad hoc systems and allowing users to consolidate their digital identities. OpenID enables maintaining single-user credentials for access control and can be used while allowing users to log onto many services using the same digital identity. It allows users to log in once and gain access to resources across participating systems. For example, Google Apps provides cloud identity services to its enterprise customers using OpenID.

This lesson covers governance, risk, and compliance aspects in Cloud. Governance aspects include SLAs and information flow regulations. Risk assessment includes identification of critical assets, potential risks, and the classification of critical assets into risk categories. Compliance includes internal and external policy compliance.



These slides defines the elements of Governance, Risk, and Compliance (GRC).

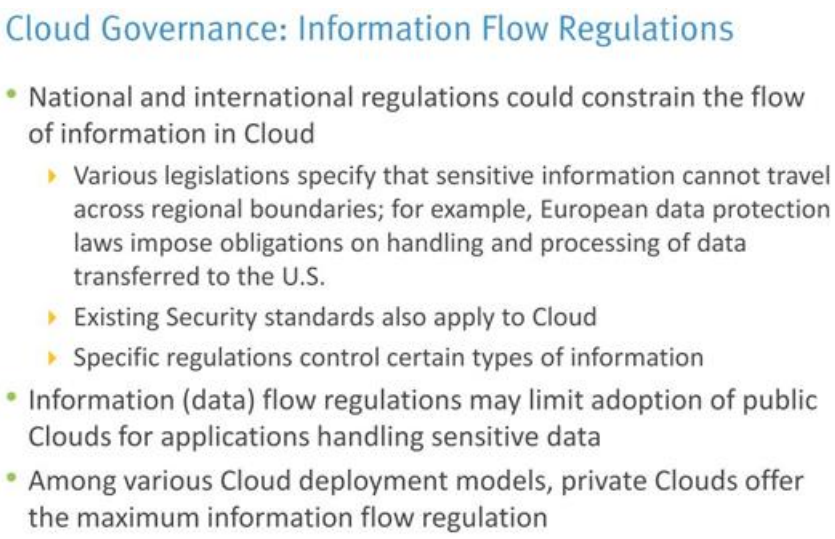
• **Governance** refers to policies, processes, laws, and institutions that define the structure by which companies are directed and managed. In Cloud, the contractual agreement between a Cloud user and the CSP defines the terms and conditions for both Cloud user as well as the

CSP.

- **Risk** is the effect of uncertainty on business objectives; risk management is a coordinated activity to direct and control an organization to realize its business potential while managing negative events.

- **Compliance** is the act of adhering to, and demonstrating adherence to external laws and regulations as well as corporate policies and procedures.

An Enterprise GRC (eGRC) solution allow a business to build an efficient, collaborative enterprise governance, risk and compliance (eGRC) program across IT, finance, operations, and legal domains. Such a program enables a business to manage risks, demonstrate compliance, automate business processes, and gain visibility into corporate risk and security controls.



National and international regulations could constrain the flow of information in Cloud. For example, Cloud offers location independence to its clients in terms of where the actual data resides. This might also lead to difficult legal issues; for example, a governmental data on a Cloud, handled abroad by private parties, fall under the foreign jurisdiction. There are various regulations, which specify that sensitive information cannot travel across regional boundaries; for example, European data protection laws impose additional obligations on the handling and processing of data transferred to the U.S.

Another aspect is that existing security standards may also apply to Cloud; for example, it is recommended that CSPs follow ISO 27001, which is an Information Security Management System standard and formally specifies requirements to bring information security under explicit management control.

Furthermore, there are regulations that control certain specific types of information. For example, the Administrative Simplification provisions of Health Insurance Portability and Accountability Act (HIPAA) specify security and privacy aspects dealing with medical data in the USA.

Such information flow regulations may limit adoption of Public Clouds for applications handling sensitive data. For example, governmental agencies might find it difficult to use public Clouds, which could potentially host their data in some other country due to cost advantages.

Among various Cloud deployment models, private Clouds offer the maximum information flow regulation because the control over data remains fully with the organization.

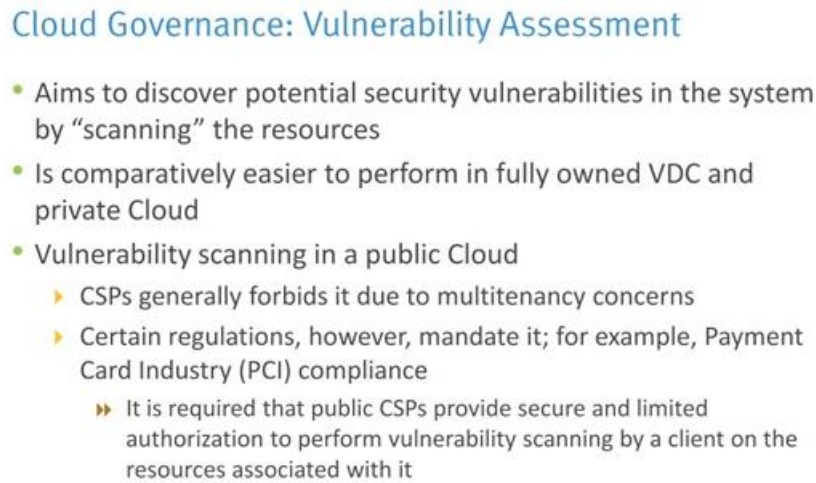
A client needs to assess implications when the existing business with a CSP needs to be terminated.

There might exist many situations where such a contract termination is required:

- The CSP goes out of business and winds up its services
- The CSP cancels the contract
- There is a natural closure for the contracted services

In order to avoid the negative impact of such a closure, there needs to be a formal agreement between a Cloud client and the CSP, which is referred as termination agreement. In situations which involve the closure of the Cloud services for a client, the following points should be considered:

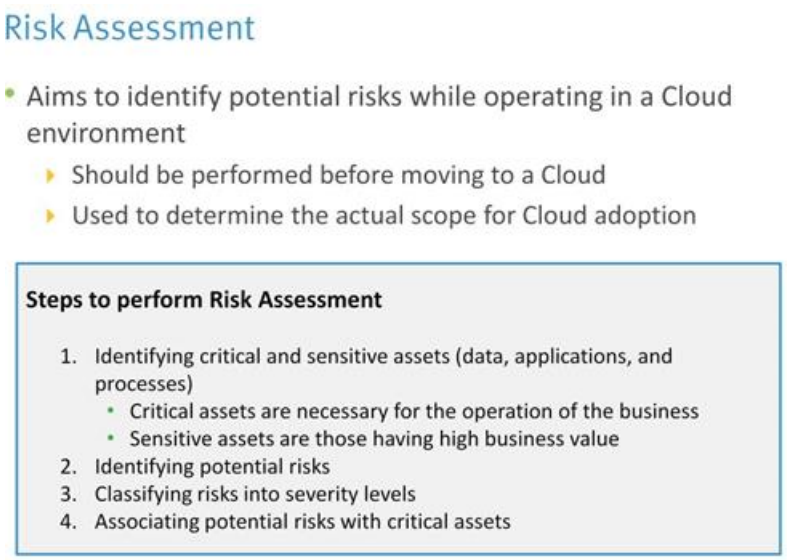
- A contingency plan for handling data in the Cloud
- A process for migrating the data back into the organization or to another Cloud
- A prior assessment as to whether the data can be moved over the network in a reasonable amount of time or whether it is necessary to make special arrangements for a physical transfer
- The plan for data destruction (storage, clones, backups) using shredding after the data has successfully moved from the Cloud



Vulnerability assessment or testing aims to discover potential security vulnerabilities in the system by "scanning" its resources (compute, storage, and network). It is comparatively easy to perform in a fully-owned VDC and Private Clouds, compared to Public or Hybrid Clouds.

Performing vulnerability scanning for a Public Cloud is often not feasible for a client because most of the Cloud providers explicitly prohibit such a scan through their terms. Any scan employed by a business has the potential to disrupt the services to other customers. On the other hand, there are certain regulations which mandate that business perform

vulnerability assessment while dealing with data. An example is Payment Card Industry (PCI) compliance for handling credit card data. For a Cloud client to be able to perform vulnerability scanning, it is required that CSPs provide secure and limited authorization to a client on the resources associated with it.



A risk assessment process aims to identify potential sources of risk while operating in a Cloud environment. This critical process is required before deciding to move operations to Cloud, especially if the decision is related to the Public Clouds.

Various steps involved in this process are:

1. Identifying critical and Sensitive Assets: All the assets (data, applications, and processes) must be carefully evaluated to assess as to how critical or sensitive an asset is for the organization. Critical assets are necessary for the operation of the business. Sensitive assets are those having high business value for the organization, for example, Intellectual Property (IP), project plans, and Personally Identifiable Information (PII).

2. Identifying potential Risks: It is important to analyze and identify all the potential risks while operating in a Cloud environment. An example could be the legal situations under which a CSP might have to disclose data belonging to its clients.

3. Classifying Risk into Severity Levels: After comprehensive classification of the risks associated with operating in Cloud, these risks need to be classified into various severity levels i.e., very high risk, moderately high risk, high risk, low risk, and no risk. These severity levels could alternately be numbered in a certain range, for example, 0 to 5.

4. Associating potential Risks with critical Assets: This step involves associating the critical assets with potential risks; for example, employee's banking records can be identified as critical assets (in step 1), data disclosure could be a risk (identified in step 2) of very high severity level (in step 3). Based upon these, data disclosure risk may be associated with employee records.

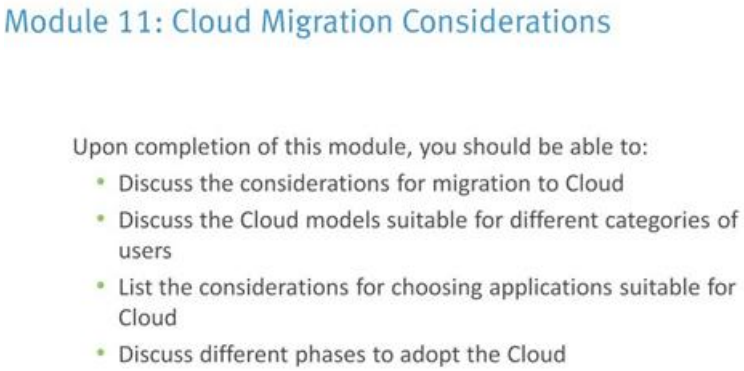
Based upon the risk assessment for the assets, a client could consider formulating terms and conditions of the contractual agreement with CSP; for example, a client might insist on having its data placed within certain geographical regions by the CSP.

Cloud service adoption and operation for enterprise businesses should abide by compliance policies. There are primarily two types of policies controlling IT operations in an enterprise that requires compliance even after moving operations to Cloud.

Internal Policy Compliance: Controls the nature of IT operations within an organization. A Cloud client organization needs to maintain same compliance even when operating in Cloud. This would require clear assessment of the potential difficulties in maintaining the compliance in Cloud and a process to ensure that this is effectively achieved.

External Policy Compliance: Includes legal legislations and industry regulations. These external compliance policies control the nature of IT operations related to the flow of data out of an organization. However, they may differ based upon the type of information (for example, source code versus employee records), business (for example medical services versus financial services), etc.

Meeting all the varied client compliance requirements is generally difficult for a CSP. Compared to Private Clouds, the Public Cloud environment makes compliance more challenging. Therefore, many enterprises may prefer to adopt the hybrid Cloud deployment model so that they can ensure all the necessary policy compliances.



This module focuses on considerations for migration to the Cloud. It details 'Cloud model' suitable for different categories of users. Further, it covers considerations for choosing candidate application and various other considerations for migration to Cloud. It also covers various phases to adopt the Cloud.

Note: All the discussions in this module are primarily focused on Public Cloud and external private Cloud, referred here as "Cloud". However, the discussed approach may be considered while migrating to Internal private Clouds.

Module 11: Cloud Migration Considerations

Lesson 1: Migration Considerations

Topics covered in this lesson:

- Considerations for Cloud model and suitable application for Cloud
- Criteria for Cloud vendor selection
- Service Level Agreements (SLAs) and performance issues
- Cloud vendor lock-in
- Cloud open standards

This lesson covers the key points to consider while migrating an application to the Cloud. The key points to consider are: How Cloud fits in an organizations environment, Cloud model, candidate applications, criteria for Cloud vendor selection, Service level agreements, etc.

Cloud Migration – Key Questions

CIOs/IT Managers seeking to move to Cloud face several questions:

- How does Cloud fit into the organization's requirements?
  - Financial advantage, convenience, etc.
- Which are the applications suitable for Cloud?
- How do I choose the Cloud Vendor?
- Is the Cloud infrastructure capable of providing the required Quality of Service (QoS)?
  - Performance, availability, and security
- How will I address Change Management concerns?
- What can Cloud provide?
  - Application, platform and infrastructure

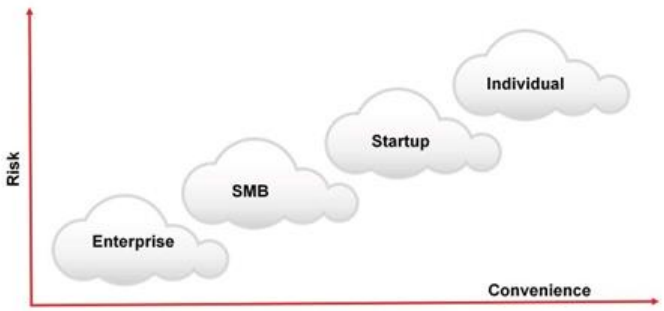
Organizations are not only looking to get a financial advantage with their first move into Cloud, but are also making a significant learning experience to expand their Cloud perspective. Businesses, determining to make their first move into the Cloud, always face a question "How does Cloud fit to the organization's environment?" This is so because there is a risk of introducing an evolving Cloud into an established system.

Most companies are not ready to abandon their existing IT investments to move all of their business processes fully to the Cloud at once. Instead, it is more likely to be a gradual shift in the business processes to the Cloud over time. The reason behind this cautious approach is that the Cloud providers are not assuring the same levels of security, controls, and performance that organizations have on premises. Lack of regulatory compliance and policies for both providers and consumers further slow down the adoption of Cloud.

It is important to understand the various Cloud migration considerations before migrating to the Cloud.

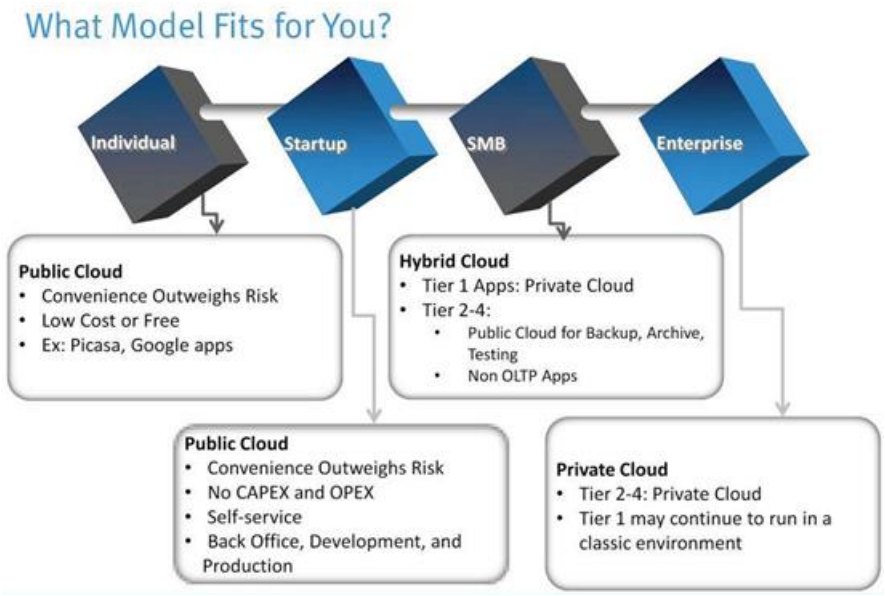
How does Cloud fit to your requirements?

- Current infrastructure and requirements
  - Consider from application, network, and security perspectives
- 'Risk vs. Convenience' profile
  - Based on this profile, choose 'Cloud model' for your organization



Studies based on the experience of early Cloud adopters suggest that moving to the Cloud without proper strategy and process does not yield expected benefits. The most important concern that needs to be evaluated before making a move to Cloud is 'How does Cloud Computing fit in the context of the organization's overall business strategy?". Sometimes, a Cloud may look attractive from an application perspective. However, in real time, it could be a challenge for network administrators. Further more, there could be security concerns about having data outside the firewall.

Risk versus convenience is a key consideration for deciding Cloud migration strategy. This consideration also forms the basis for choosing right Cloud model. Cloud benefits are well established, but data may reside outside the organization's perimeter causing risk. A balance must be evaluated to determine how much risk an individual or organization may handle for the benefit of convenience. This proportion varies among Cloud consumers, based on which they may be segmented into individual, business startup, small and medium business, and enterprise. Typically, individuals and startup businesses are ready to take high risk to get most of the convenience offered by a Cloud. Compared to that, SMBs and enterprises are more sensitive to risk and are unlikely to move their applications to Cloud.



Let us understand which Cloud model will be most suitable for an organization or an individual:

- Public Cloud is preferred by individuals who want to access Cloud services such as Picasa and Google apps, and are least concerned about the security or availability risks in Cloud for the most part. Here, cost reduction is the primary objective. Public Cloud enables the opportunity to access these applications for free or by paying minimum usage charges.
- People who start up businesses from small office or home typically opt for Public Cloud. A large investment to purchase IT resources is not affordable or may not give the required ROI. Therefore, for obvious reasons, convenience offered by the Cloud outweighs risk.
- Small and medium-sized businesses have a moderate customer base and any anomaly in customer data and service levels may impact their business. Hence, they may not be willing or be able to put Tier 1 applications, such as Online Transaction Processing (OLTP), in the Cloud. A hybrid Cloud model may fit in this case, which includes the organization's internal IT resources (Private Cloud) and external Public Cloud resources. Tier 1 application data should never cross the boundary of Private Cloud. Public Cloud enables cost savings and faster time to market and is typically used for tier 2, tier 3, and tier 4 applications such as backup, archive, and testing.
- Enterprises typically have a strong customer base worldwide. The priority is to maintain critical customer data and service levels, with strict enforcement of security policies. They are highly concerned with the risk and information access control in Cloud. They are financially capable of building a massive Private Cloud. Many enterprises may not even want to move any of

their applications to Cloud.

Choosing Applications for Public Cloud

- Proprietary and mission-critical application
  - Proprietary applications provide competitive advantage
  - Organization perceives high risk to move this application to Cloud
  - These applications are typically maintained in-house
- Non-proprietary but mission-critical application
  - Organization perceives high risk to move this application to Cloud
  - It can be moved to Cloud if:
    - Organization does not have adequate resources to maintain the application
- Non-proprietary and non-mission critical application
  - Perceived as good candidate for Cloud, unless it is performance sensitive

Not all applications are good candidates for Cloud, although it may depend on the capability of the Cloud infrastructure and the quality of service offered by Cloud providers.

When migrating applications to the Cloud, there are three general considerations that may be used to determine if the application can move to the Cloud. Proprietary and mission- critical applications are core and essential to the business. Often, they are applications that provide competitive advantages and are usually designed, developed, and maintained in- house. Typically, the perceived risk and effort to outsource these systems to the Cloud is high.

Give close consideration to applications that are non-proprietary, but are still mission-critical. Though the effort to migrate these applications to the Cloud may be minimal, the perceived risk to the business may be deemed high. If the organization does not have adequate resources to maintain the application or if the cost to maintain the application is high, then this may outweigh the risks.

The sweet spot for migrating applications to the Cloud is the non-proprietary and non-mission critical applications if they are not performance sensitive. These applications have good compatibility, standardized functionality, and interfaces, making the level of migration effort minimal in comparison to proprietary applications. Since these are non-proprietary and non-mission critical applications, moving to the Cloud poses minimum risk.

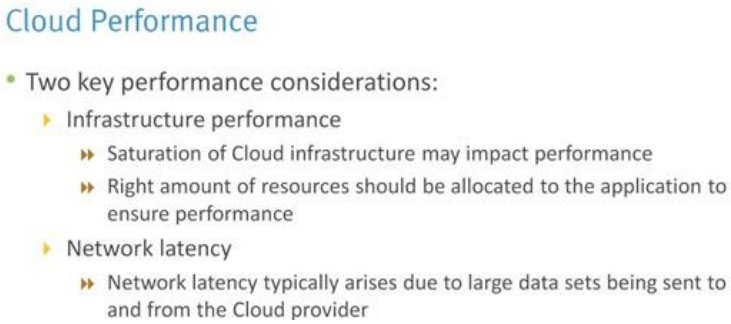
Selecting a Cloud Service Provider

- Some key questions to ask before selecting a Cloud Service provider:
  - How long has the provider been providing the services?
  - How well does the provider meet the organization's current and future requirements?
  - How easy is it to relinquish resources not in use and to reduce cost?
  - What tools does the provider offer (like virtual machine images) that would make it easy to move to another provider when required?
  - How easy is it to add/remove services?
  - Does the provider have good customer service support?
  - What happens when the provider upgrades their software? Is it forced on everyone? Can you upgrade on your own schedule?
  - Does the provider offer required security services?
  - Does the provider meet your legal and privacy requirements?

Cloud is an emerging technology and many Cloud players are just entering the market. Out of the several Cloud service providers, selecting a provider is a critical task. Some key provider selection criteria are listed on this slide.

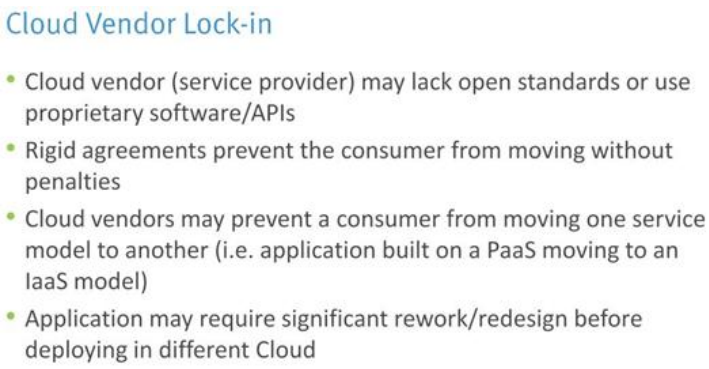
As consumers move towards Cloud, the quality and reliability of services become important considerations. However, the demands of the consumers vary significantly. It is not possible to fulfill all consumer expectations from the service provider's perspective, and hence, a balance needs to be made via a negotiation process. At the end of the negotiation process, the provider and the consumer commit to an agreement. This agreement is referred to as Service Level Agreement (SLA). This SLA serves as the foundation for the expected level of service between the consumer and the provider. The QoS attributes are generally part of an SLA (such as response time and throughput). However, these attributes change constantly, and to enforce the agreement, these parameters need to be closely monitored.

Strong Service Level Agreements (SLAs) from Cloud vendors are a must to ensure QoS. Without these agreements and penalties for failing to meet them, vendors have less incentive to maintain performance at the highest levels. SLAs can include factors such as network availability, performance, etc.



There are two key factors that impact Cloud performance, infrastructure performance and network latency.

- **Infrastructure performance:** Most Cloud platforms leverage a shared, multitenant, virtual infrastructure. An application may have its own virtual space or virtual machine, but it shares processors and storage space with several other applications on that Cloud infrastructure. It is possible that the Cloud infrastructure may become saturated from time to time, and thus impact performance. There is not much that can be done about this, other than work with the Cloud provider to ensure that the application gets the required performance. Typically, this is taken care in the SLA.
- **Network latency:** Performance related to network latency typically arises due to large data sets being sent to and from the Cloud provider. The larger the dataset, the more likely that the network performance issues come into play.



This slide lists some key reasons that may prevent moving an application from one Cloud to another.

Cloud lock-in refers to a situation where a Cloud consumer is unable to move out of the current Cloud vendor (service provider) due to the complexity/restriction imposed by the current Cloud vendor.

Vendor lock-in is seen as deterrent to moving services to the Cloud. This concern is magnified when looking at a federated Cloud. Lock-in may prevent an organization from moving their application from one Cloud to another to take advantage of geography and for potential performance improvement. It makes a customer dependent on a vendor because switching to another vendor may come with costs.

Cloud Open Standards

- Use proven and widely accepted technologies
- Prevent lock-in issues
- Open Virtual Machine Format (OVF) - an example of open standard

Open standards are the building block for multi-vendor, federated Clouds and can make vendor lock-in avoidable. Widely accepted standards provide interoperability and portability. Without open standards, it becomes difficult to connect public, private, and hybrid Clouds.

When using open standards, choose technologies that are widely accepted and proven. For example, Use of common APIs for an application may allow an organization to move it to other Cloud with minimal or no change. Similarly, use of Open Virtual Machine Format (OVF) is a common VM format that enables using a VM built in one Cloud to be deployed to another Cloud with minimum or no changes.

Module 11: Cloud Migration Considerations

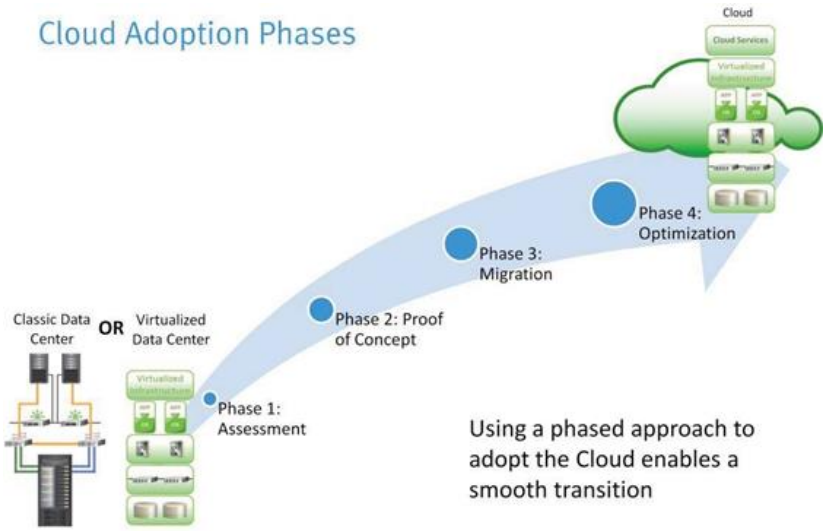
Lesson 2: Phases to Adopt the Cloud

Topic covered in this lesson:

- Different phases to adopt the Cloud

This lesson covers the different phases involved in adopting the Cloud. The adoption phases are: assessment phase, proof of concept phase, migration phase, and optimization phase.

Cloud Adoption Phases



After identifying the right application for the Cloud, it may be moved to the Cloud. Organizations typically adopt the Cloud in phases for smooth transition. The adoption process typically consists of four phases. They are assessment phase, proof of concept phase, migration phase, and optimization phase.

Phase 1: Assessment

- The assessment phase involves consideration of various factors
- Besides Cloud migration considerations discussed earlier, other key assessments are:
  - Financial
  - Security and compliance
  - Technical
  - Issues with licensed products

The first phase in Cloud adoption is the assessment phase. To ensure successful assessment, it is important to define and understand its objectives. Assessment involves consideration of various factors. Assessment should be performed for each application that is identified as a potential candidate for the Cloud. Besides considerations discussed in the previous lesson – Migration Considerations, other key assessments are: financial assessment, security and compliance assessment, technical assessment, and assessment of issues related to migration of licensed products.

Financial Assessment

- Provides cost comparison of in-house vs. service provider
  - TCO and ROI
- Requires cost consideration of the following elements:

Saving		Spending
CAPEX	OPEX	Other Cost
• Servers • Storage • Operating System (OS) • Application • Network equipments • Real estate	• Power and cooling • Personnel • Bandwidth • Maintenance • Support • Backup	• Migration • Compliance and Governance • Service cost

Financial assessment helps in determining Total Cost of Ownership (TCO) and Return on Investment (ROI) and in further building a business case. Financial assessment requires detailed and careful analysis, because it provides a clear picture of the cost involved in owning and operating a data center versus employing a Cloud based infrastructure.

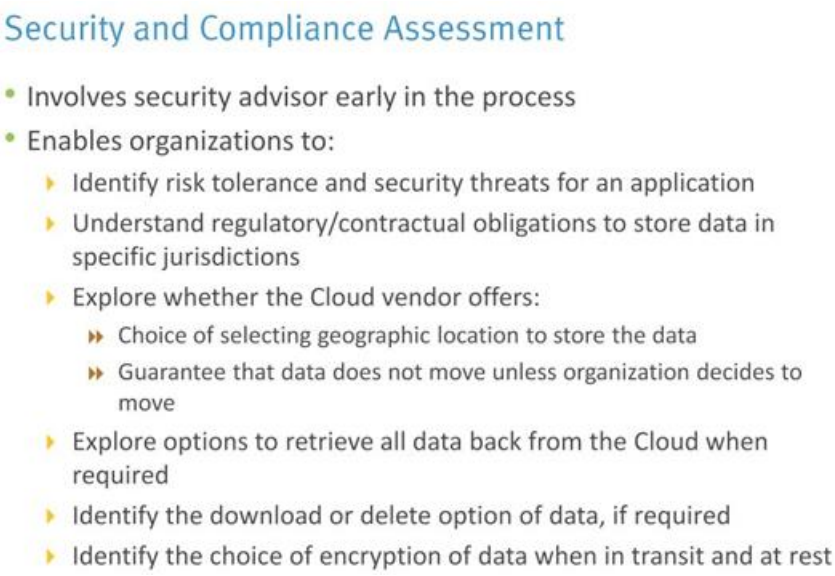
Financial assessment is not simple and requires considering CAPEX, OPEX, and Overhead cost.

For example, consider an organization that requires moving an application that has specific requirements, such as archiving data and retaining it for long period in the Cloud (compliance and governance) or specific security requirements. Then, the following are the cost components that should be considered to perform a financial assessment:

1. CAPEX includes cost of servers, storage, Operating System (OS), application, network equipments, real estate, etc. Moving an application to Cloud reduces CAPEX, compared to maintaining it on site.

2. OPEX includes the cost incurred for power and cooling, personnel, bandwidth, maintenance, support, backup, etc. Moving an application to Cloud reduces OPEX, compared to maintaining it on site.

3. Overhead includes migration cost and cost to ensure compliance and governance. In this case, the organization may have to pay an additional cost for archival service and storage capacity (to ensure compliance and governance). The organization may have to pay an additional cost to get the required security services for their application. Moving an application to Cloud may incur overhead cost over and above the subscription fees.

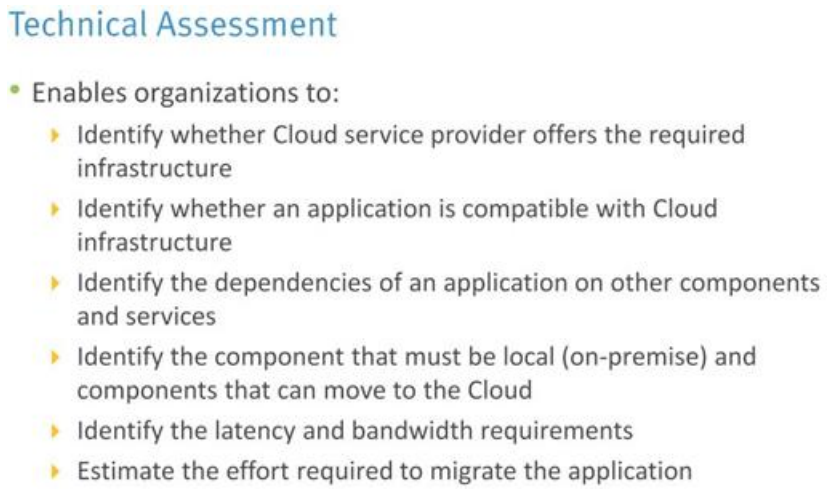


Data security can be a daunting issue if not properly understood and analyzed. Hence, it is important to understand risks and threats. Based on the sensitivity of data, classify the data assets into different categories (confidential, public, internal only, etc). This will help identify which data asset can be moved to the Cloud and which can be kept in-house.

If an organization has strict security policies and compliance requirements, it is recommended to involve the organization's security advisers and auditors early in the process.

This assessment will enable organizations to:

- Identify the overall risk tolerance for an application.
- Identify the security threats that have a likelihood of materializing into actual attacks.
- Understand the regulatory or contractual obligations to store data in specific jurisdictions.
- Explore whether the Cloud vendor offers a choice of selecting the geographic location to store the data and a guarantee that the data does not move unless the organization decides to move it.
- Explore options if the organization decides to retrieve all of the data back from the Cloud.
- Identify whether the Cloud vendor offers options to download or delete the data whenever required.
- Identify the choices offered by the Cloud vendor to encrypt the data while it is in transit and while it is at rest.



A technical assessment helps identify the applications that are more suited to the Cloud. It also helps the organization determine the applications that should move into the Cloud first, those that should move later, and those that should remain in-house.

Technical assessment identifies the dependencies of an application on other components and services. For example, a Web-based application depends on the database, login, and authentication services running on another system. In most cases, the best candidates for the Cloud are the applications that have minimum dependencies.

This assessment will enable organizations to:

- Identify whether the Cloud service provider offers all of the required infrastructure building blocks.
- Identify whether the application can be packaged into a Virtual Machine (VM) instance and be run on a Cloud infrastructure.
- Identify the component that must be local (on-premise) and components that can move to the Cloud.
- Identify the latency and bandwidth requirements.
- Estimate the effort required to migrate the application.

It is important to assess the issues related to migrating licensed software to the Cloud at an early stage.

There are two options available for migrating licensed software to the Cloud:

- **Use the existing License:** Cloud providers have partnered with several software vendors. Due to this partnership, software vendors permit organizations to use their existing product license on Cloud. This option offers the easiest path to move licensed software to the Cloud. In this option, the organization purchases license in the traditional way or uses the existing license in the Cloud.
- **Use SaaS Based Cloud Service:** Many software vendors offer their software with two

options, one that can be installed on-premise, and the other as a service. In this option (SaaS), the existing on-premise installed application is migrated to a hosted offering (SaaS) by the same vendor. If a software vendor does not offer its software as a service, explore and migrate the data on an equivalent offering by a different Cloud provider. In this case, the Cloud provider may charge a monthly subscription fee.

Phase 2: Proof of Concept

- Goal of this phase is to verify that an application runs as expected in the Cloud
- Proof of Concept enables organizations to:
  - Explore the capabilities of the Cloud
  - Explore the different business continuity and disaster recovery options offered by the Cloud vendor
  - Estimate the effort required to roll out the application
  - Identify applications that can be immediately moved after proof of concept

After a thorough assessment, identifying the right candidate for the Cloud, and estimating the efforts required for migration, it is time to test the application with a proof of concept. This phase helps to understand what an application can do and cannot do in Cloud.

The goal of this phase is to check whether an application runs as expected after migrating it to the Cloud. It is recommended to perform a thorough testing of the application during this phase. In this phase, the organization can validate the Cloud technology, test legacy software in the Cloud, perform necessary benchmarks, and set expectations.

This assessment will enable organizations to:

- Explore the capabilities of the Cloud
- Explore the different business continuity and disaster recovery options offered by the Cloud vendor
- Estimate the effort required to roll this proof-of-concept out to production
- Identify applications that can move after proof of concept

After this phase, the organization will gain hands-on experience with the Cloud environment, which in turn gives them more insight into what hurdles need to be overcome in order to move ahead.

Phase 3: Migration

Migration Strategies	Description
Forklift migration strategy	• Entire application is migrated at once instead of in parts • Good for tightly coupled or self contained applications
Hybrid migration strategy	• Some parts of application are moved into Cloud and some part remains in the data center • Good for application that have several components, and not tightly coupled

In this phase, applications are migrated to the Cloud. There are two application migration strategies:

- **Forklift Migration Strategy:** In this strategy, rather than moving applications in parts over time, all applications are picked up at once and moved to the Cloud. Tightly coupled applications

(multiple applications that are dependent on each other and cannot be separated) or self-contained applications might be better served by using the forklift approach. Self-contained Web applications that can be treated as a single entity and backup/archival systems are examples of systems that can be moved into the Cloud using this strategy.

- **Hybrid Migration Strategy:** In this strategy, some parts of the application are moved to the Cloud while leaving the other parts of the application in place. The hybrid migration strategy can be a low-risk approach to migration of applications to the Cloud. Rather than moving the entire application at once, parts of it can be moved and optimized, one at a time. This strategy is good for large systems that involve several applications and those that are not tightly coupled.

After migrating the application to the Cloud, run the necessary tests and confirm that everything is working, as expected. In this phase, focus on how to optimize the Cloud based application in order to increase cost savings.

Understand the usage pattern to optimize the resources consumed. To understand the usage pattern, monitor the resources consumed and the workload. Based on the workload, resources can be scaled up or scaled down. For example, if a customer-facing Website, deployed on a Cloud infrastructure, does not expect any traffic from certain parts of the world at a certain time of the day, the resources consumed by that region may be scaled down for that time.

Inspect the system logs periodically to understand the usage of the resources. Relinquish the idle resource.