MODULE – 2

DATA CENTER ENVIRONMENT

前情回顾

- 以下哪一项是结构化数据的示例？
  - A. 镜像
  - B. PDF 文档
  - C. Datebase 
  - D. XML文档

- 关于大数据，以下哪项是正确的？
  - A. 数据规模超过10G
  - B. 单个数据规模超过10G
  - C. 使用传统软件工具高效捕获
  - D. 数据大小超出传统软件的处理能力 

知识测验 – 2

以下哪一项准确描述了虚拟化？

A. 提供按需计量的虚拟服务
B. 将物理资源抽象化为逻辑资源
C. 共用逻辑资源以提供数据完整性
D. 支持跨数据中心的分散式管理

常见数据中心主机

塔式计算系统	机架式计算系统	刀片式计算系统
		

Module 2: Data Center Environment

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

- **Describe the core elements of a data center**
- **Describe virtualization at application and host layer**
- **Describe disk drive components and performance**
- **Describe host access to storage through DAS**
- **Describe working and benefits of flash drives**

Module 2: Data Center Environment

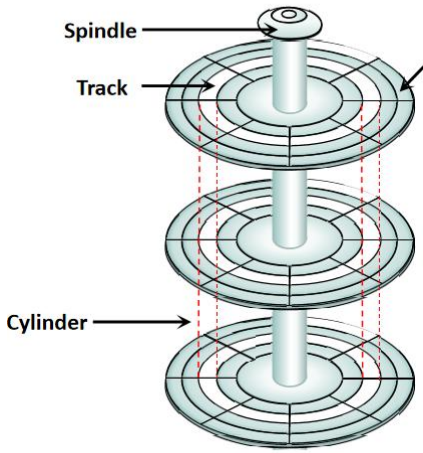
Lesson 1: Application, DBMS, and Host (Compute)

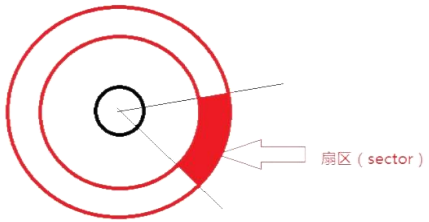
During this lesson the following topics are covered:

- Application and application virtualization
- DBMS
- Components of host system
- Compute and memory virtualization

Application

- A software program that provides logic for computing operations
- Commonly deployed applications in a data center
  - Business applications – email, enterprise resource planning (ERP), decision support system (DSS)
  - Management applications – resource management, performance tuning, virtualization
  - Data protection applications – backup, replication
  - Security applications – authentication, antivirus
- Key I/O characteristics of an application
  - Read intensive vs. write intensive
  - Sequential vs. random
  - I/O size





Application Virtualization

Application Virtualization

It is the technique of presenting an application to an end user without any installation, integration, or dependencies on the underlying computing platform.

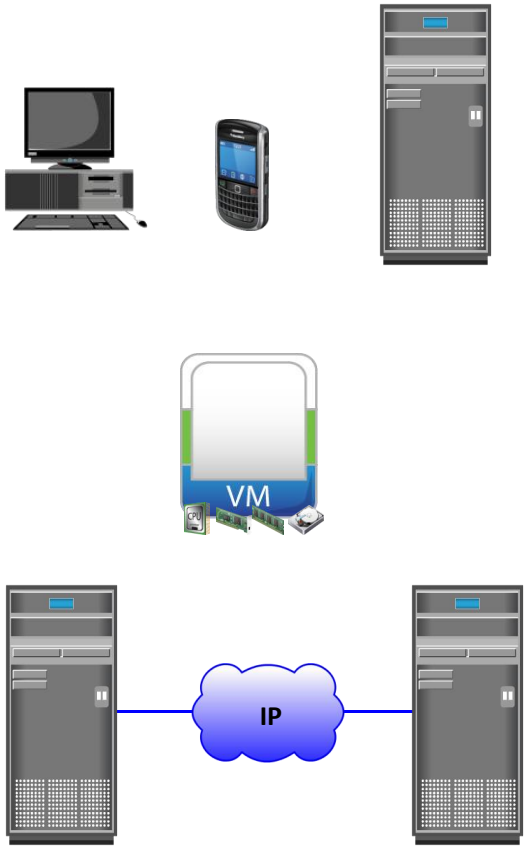
- Allows application to be delivered in an isolated environment
  - Aggregates Operating System (OS) resources and the application into a virtualized container
  - Ensures integrity of Operating System (OS) and applications
  - Avoids conflicts between different applications or different versions of the same application

Database Management System (DBMS)

- Database is a structured way to store data in logically organized tables that are interrelated
  - Helps to optimize the storage and retrieval of data
- DBMS controls the creation, maintenance, and use of databases
  - Processes an application's request for data
  - Instructs the OS to retrieve the appropriate data from storage
- Popular DBMS examples are MySQL, Oracle RDBMS, SQL Server, etc.

Host (Compute)

- Resource that runs applications with the help of underlying computing components
  - Example: Servers, mainframes, laptop, desktops, tablets, server clusters, etc.
- Consists of hardware and software components
- Hardware components
  - Include CPU, memory, and input/output (I/O) devices
- Software components
  - Include OS, device driver, file system, volume manager, and so on

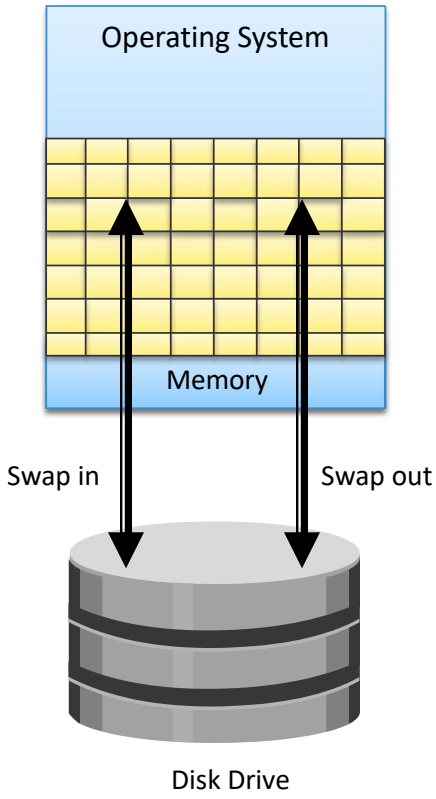


Operating Systems and Device Driver

- In a traditional environment OS resides between the applications and the hardware
  - Responsible for controlling the environment
- In a virtualized environment virtualization layer works between OS and hardware
  - Virtualization layer controls the environment
  - OS works as a guest and only controls the application environment
  - In some implementation OS is modified to communicate with virtualization layer
- Device driver is a software that enables the OS to recognize the specific device

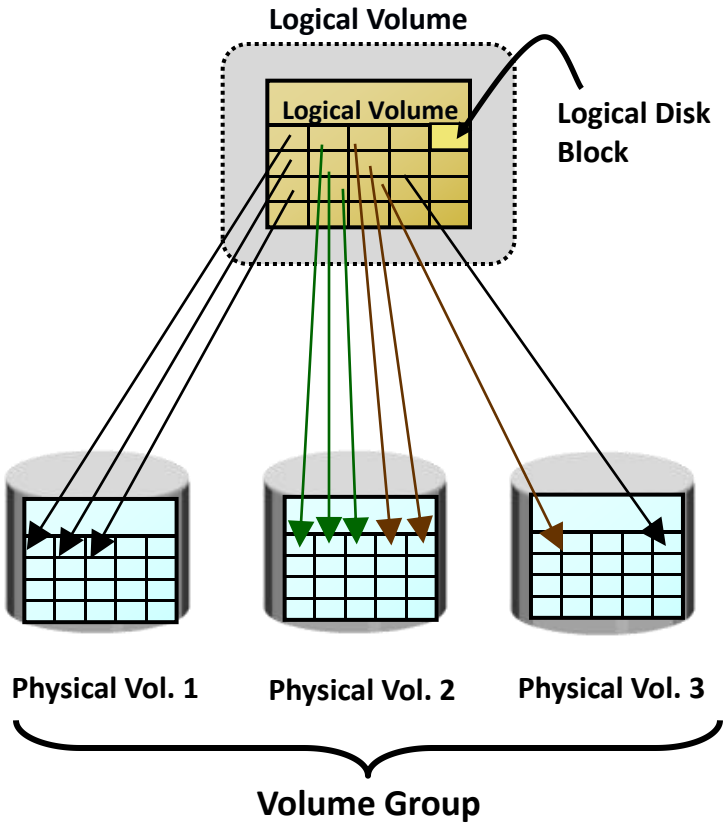
Memory Virtualization

- An OS feature that presents larger memory to the application than physically available
  - Additional memory space comes from disk storage
  - Space used on the disk for virtual memory is called ‘swap space/swap file or page file’
  - Inactive memory pages are moved from physical memory to the swap file
  - Provides efficient use of available physical memory
  - Data access from swap file is slower – use of flash drives for swap space gives best performance

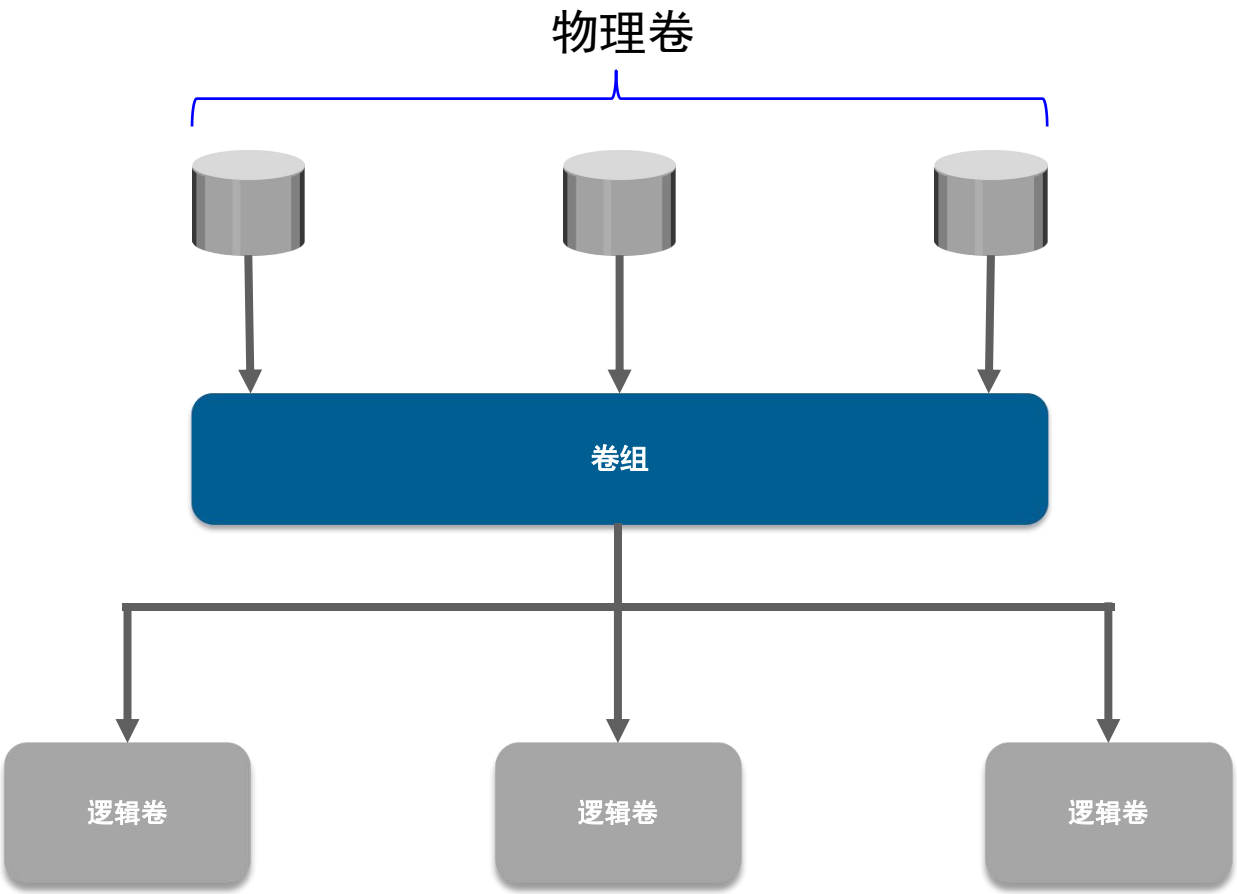


Logical Volume Manager (LVM)

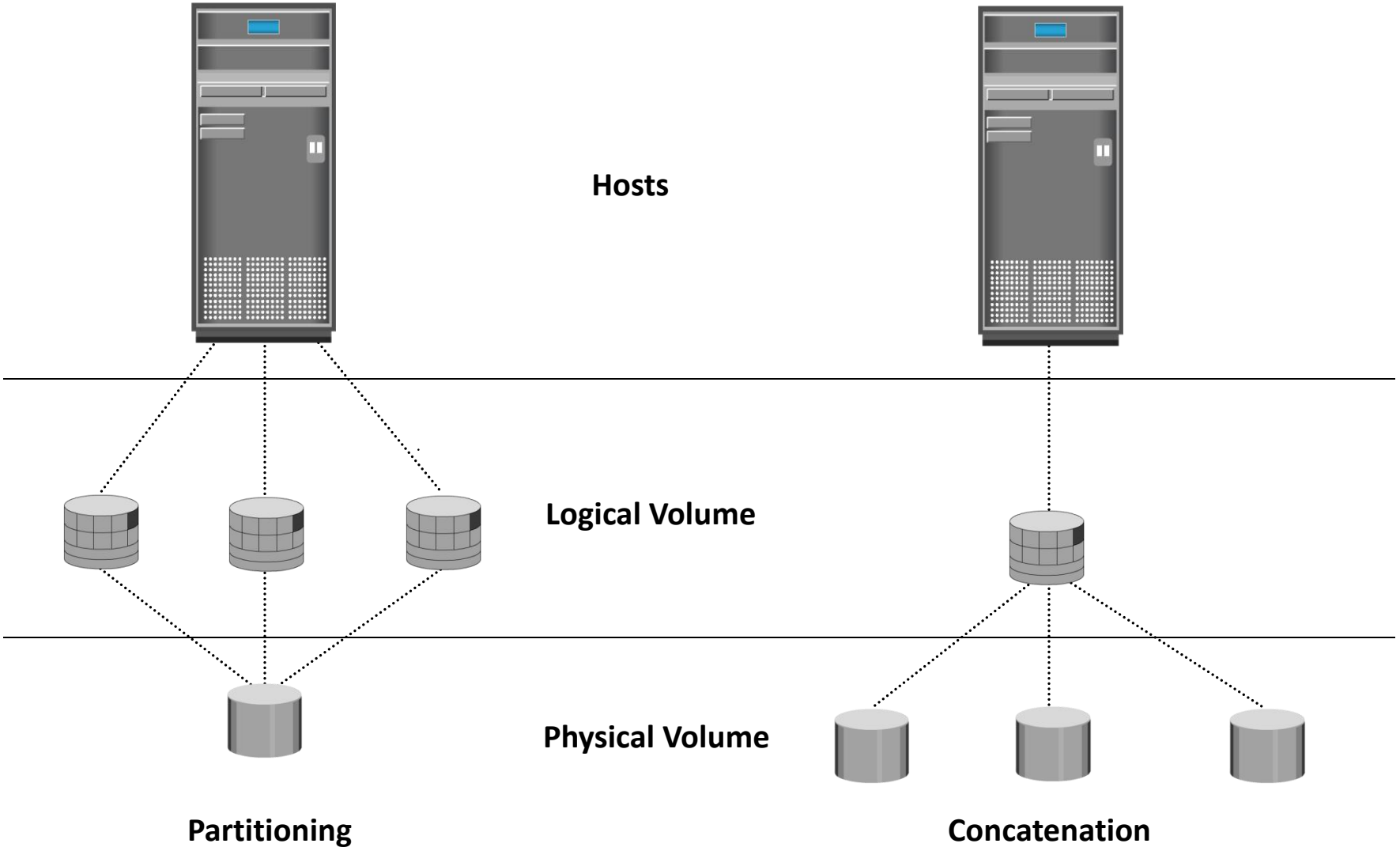
- Responsible for creating and controlling host level logical storage
  - Physical view of storage is converted to a logical view
  - Logical data blocks are mapped to physical data blocks
- One or more Physical Volumes form a Volume Group
  - LVM manages Volume Groups as a single entity
- Logical volumes are created from the volume group



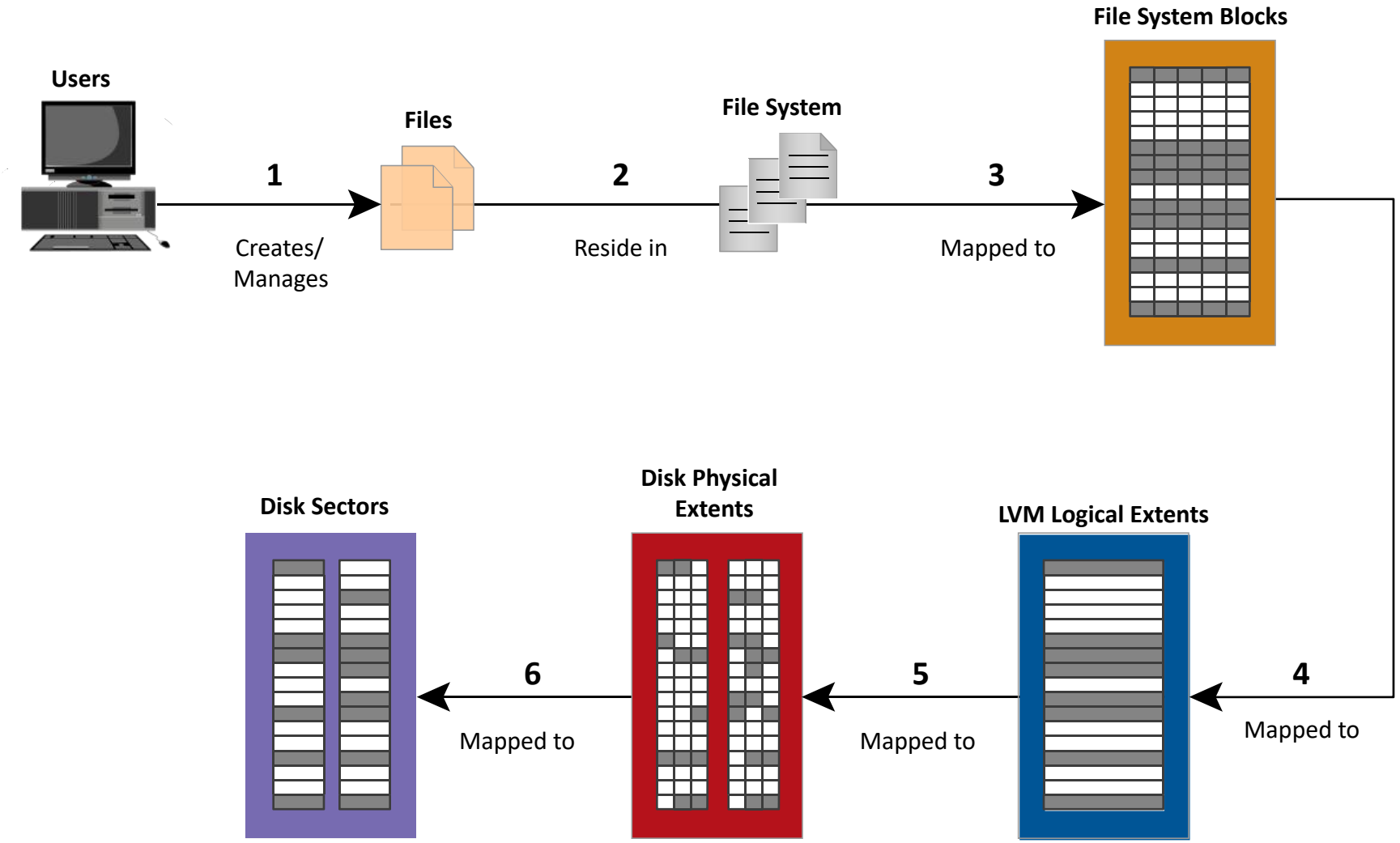
实际LVM



LVM Example: Partitioning and Concatenation



File System

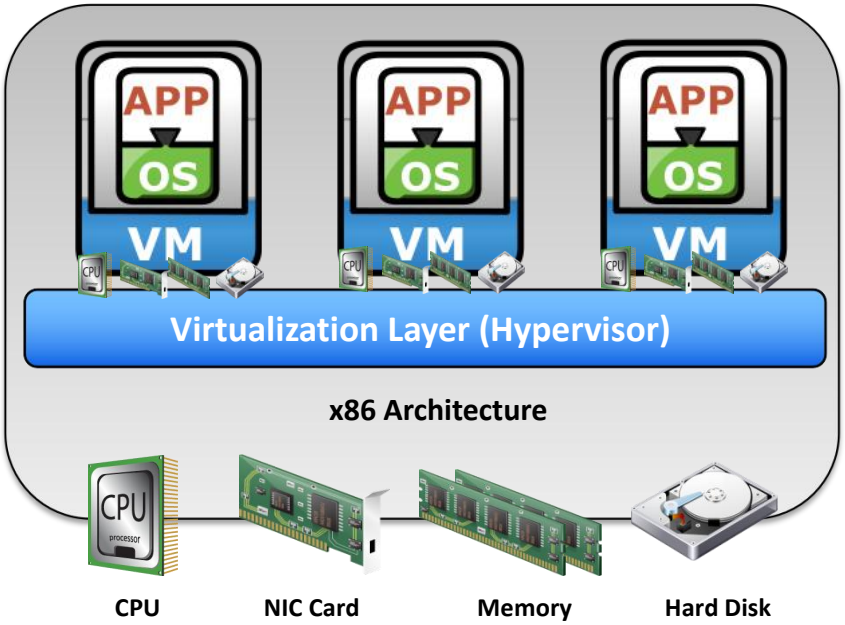

Users
Files
File System
File System Blocks
1
Creates/
Manages
2
Reside in
3
Mapped to
Disk Sectors
Disk Physical
Extents
LVM Logical Extents
6
Mapped to
5
Mapped to
4
Mapped to

Compute Virtualization

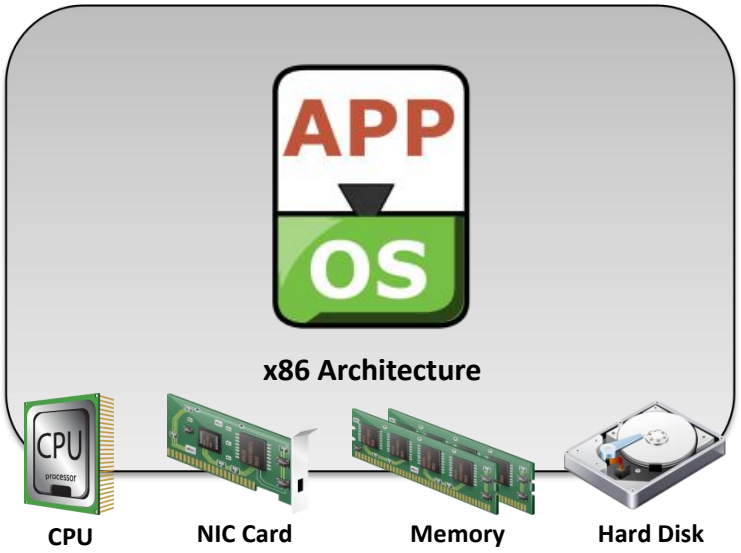
Compute Virtualization

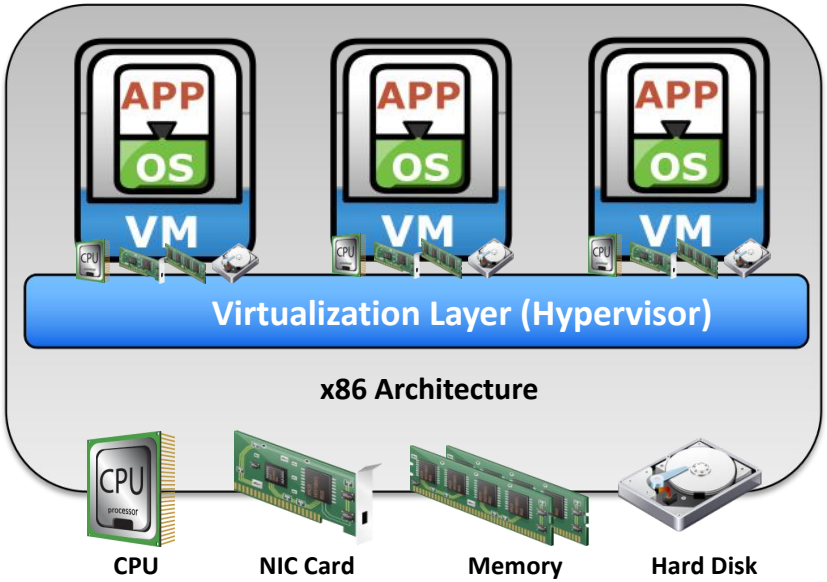
It is a technique of masking or abstracting the physical compute hardware and enabling multiple operating systems (OSs) to run concurrently on a single or clustered physical machine(s).

- 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



Need for Compute Virtualization





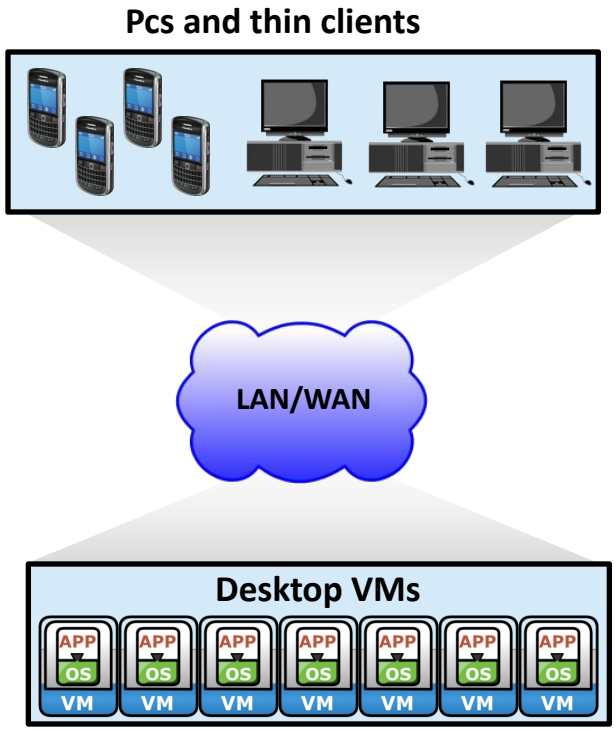
Before Virtualization	After Virtualization
• Runs single operating system (OS) per machine at a time • Couples s/w and h/w tightly • May create conflicts when multiple applications run on the same machine • Underutilizes resources • Is inflexible and expensive	• Runs multiple operating systems (OSs) per physical machine concurrently • Makes OS and applications h/w independent • Isolates VM from each other, hence, no conflict • Improves resource utilization • Offers flexible infrastructure at low cost

Desktop Virtualization

Desktop Virtualization

It is a technology which enables detachment of the user state, the Operating System (OS), and the applications from endpoint devices.

- Enables organizations to host and centrally manage desktops
  - Desktops run as virtual machines within the data center and accessed over a network
- Desktop virtualization benefits
  - Flexibility of access due to enablement of thin clients
  - Improved data security
  - Simplified data backup and PC maintenance



Module 2: Data Center Environment

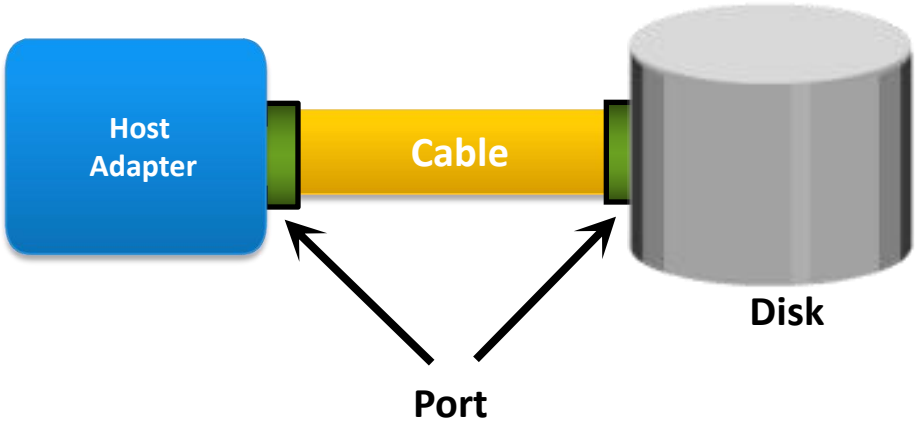
Lesson 2: Connectivity

During this lesson the following topics are covered:

- Physical components of connectivity
- Storage connectivity protocols

Connectivity

- Interconnection between hosts or between a host and peripheral devices, such as storage
- Physical Components of Connectivity are:
  - Host interface card, port, and cable
- Protocol = a defined format for communication between sending and receiving devices
  - Popular storage interface protocols: IDE/ATA and SCSI

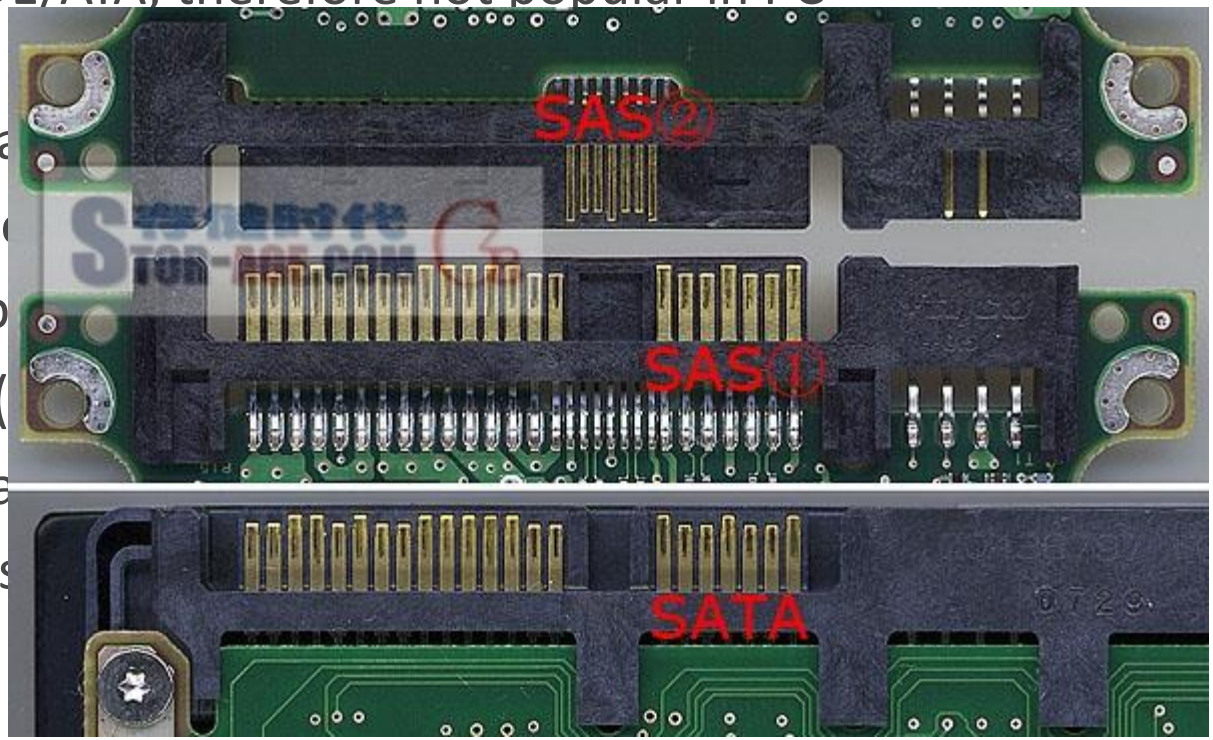


IDE/ATA and Serial ATA

- Integrated Device Electronics (IDE)/Advanced Technology Attachment (ATA)
  - Popular interface used to connect hard disks or CD-ROM drives
  - Available with varity of standards and names
- Serial Advanced Technology Attachment (SATA)
  - Serial version of the IDE/ATA specification that has replaced the parallel ATA
  - Inexpensive storage interconnect, typically used for internal connectivity
  - Provides data transfer rate up to 6 Gb/s (standard 3.0)

SCSI and SAS

- Parallel Small computer system interface (SCSI)
  - Popular standard for connecting host and peripheral devices
    - Commonly used for storage connectivity in servers
  - Higher cost than IDE/ATA, therefore not popular in PC environments
  - Available in wide va
  - Support up to 16 d
  - Ultra-640 version p
- Serial Attached SCSI (
  - Point-to-point seria
  - Supports data trans



Fibre Channel and IP

- Fibre Channel (FC)
  - Widely used protocol for high speed communication to the storage device
  - Provides a serial data transmission that operates over **copper wire** and/or **optical fiber**
  - Latest version of the FC interface '16FC' allows transmission of data up to 16 Gb/s
- Internet Protocol (IP)
  - Traditionally used to transfer host-to-host traffic
  - Provide opportunity to leverage existing IP based network for storage communication
    - Examples: iSCSI and FCIP protocols

Lesson 3: Storage

During this lesson the following topics are covered:

- Various storage options
- Disk drive components, addressing, and performance
- Enterprise Flash drives
- Host access to storage and direct-attached storage

Storage Options

- Magnetic Tape
  - Low cost solution for long term data storage
    - Preferred option for backup destination in the past
  - Limitations
    - Sequential data access
    - Single application access at a time
    - Physical wear and tear
    - Storage/retrieval overheads

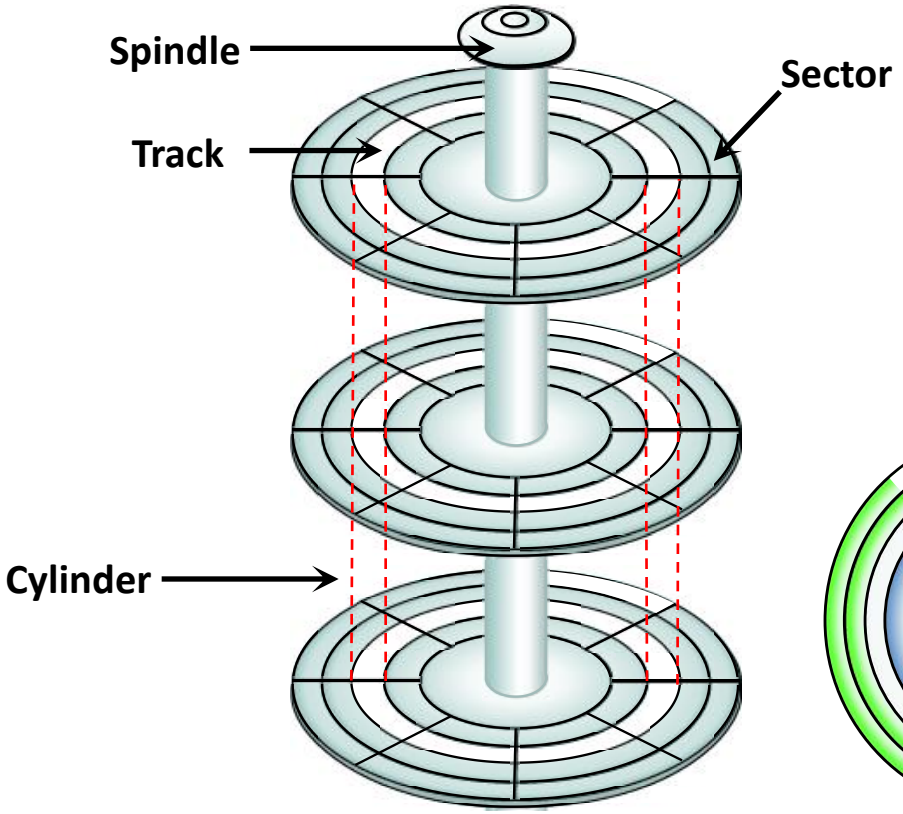
结绳 ⇨ 石块 ⇨ 刻字 ⇨ 竹简 ⇨ 纸张 ⇨ 穿孔卡 ⇨ 穿孔纸带 ⇨ 磁带 ⇨ 磁盘软盘 ⇨ 光盘 ⇨ Flash芯片 ⇨ 阵列 ⇨ 网络阵列

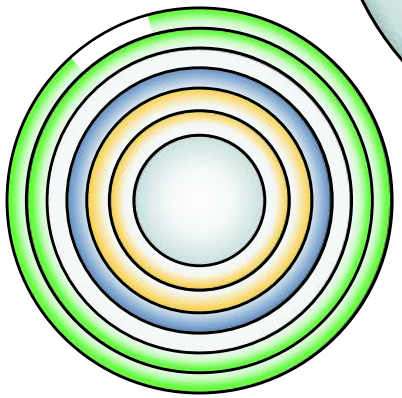
Storage Options (contd.)

- Optical discs
  - Popularly used as distribution medium in small, single-user computing environments
  - Limited in capacity and speed
  - Write once and read many (WORM): CD-ROM, DVD-ROM
  - Other variations: CD-RW, Blu-ray discs
- Disk drive
  - Most popular storage medium
  - Large storage capacity
  - Random read/write access
- Flash drives
  - Uses semiconductor media
  - Provide high performance and low power consumption

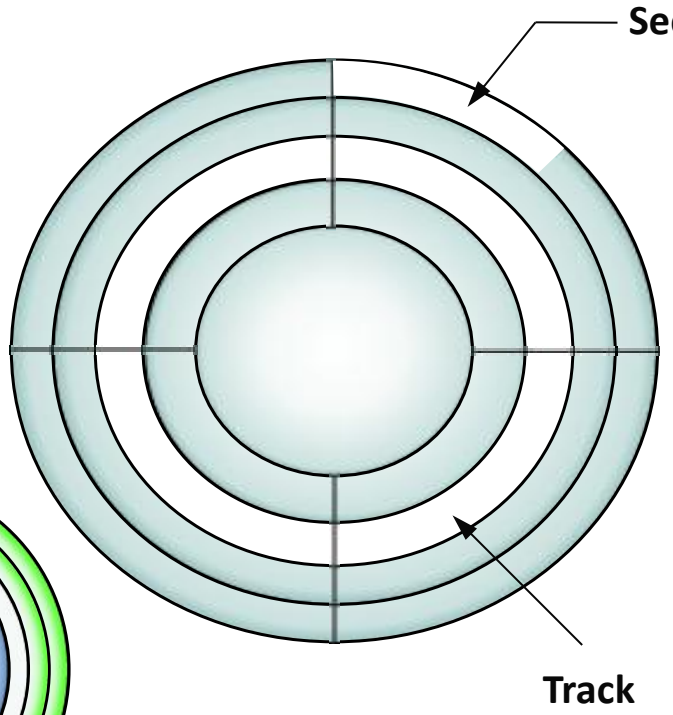
Disk Drive Components

Physical Disk Structure



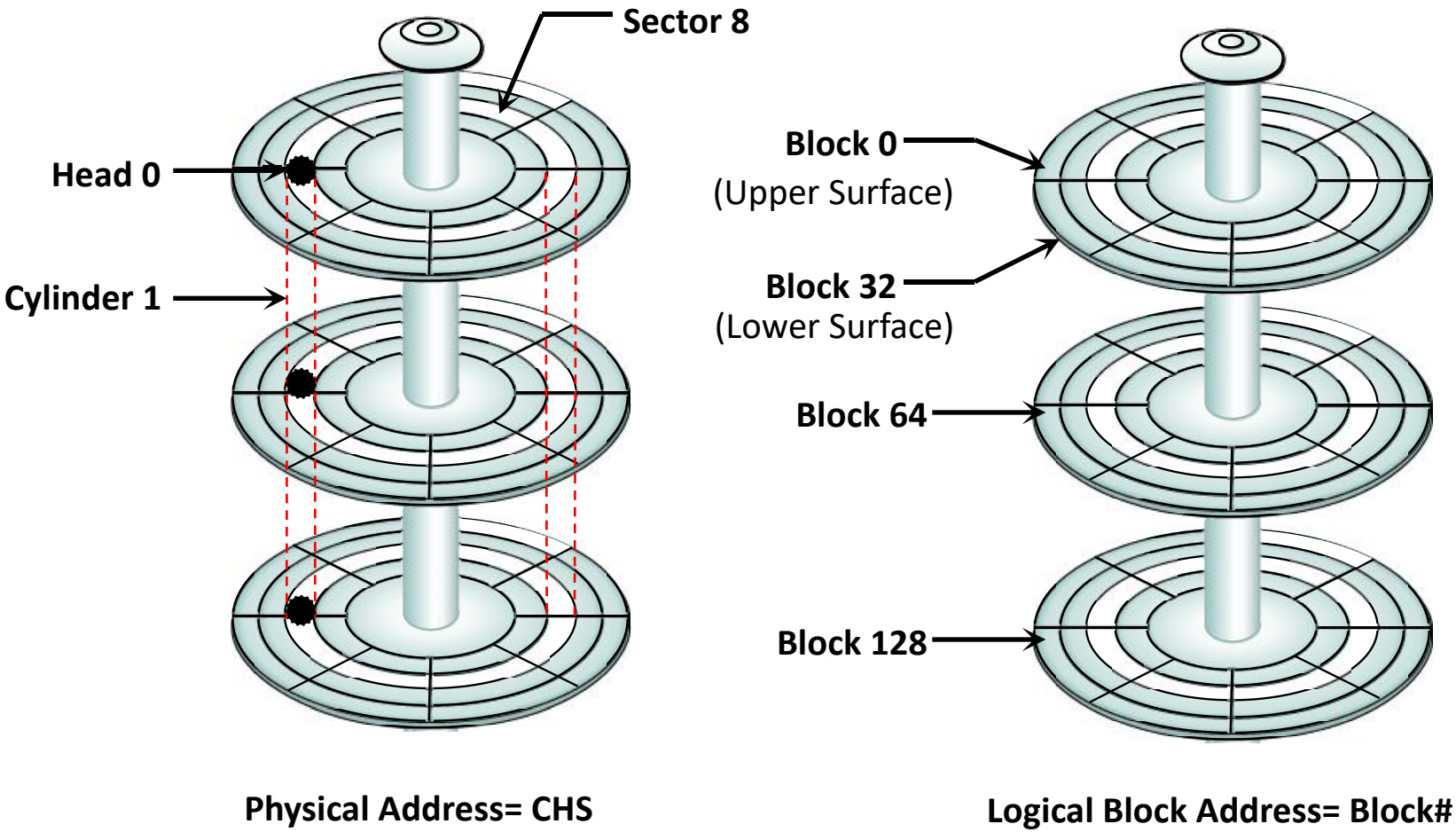


Platter With Zones



Platter Without Zones

Logical Block Addressing



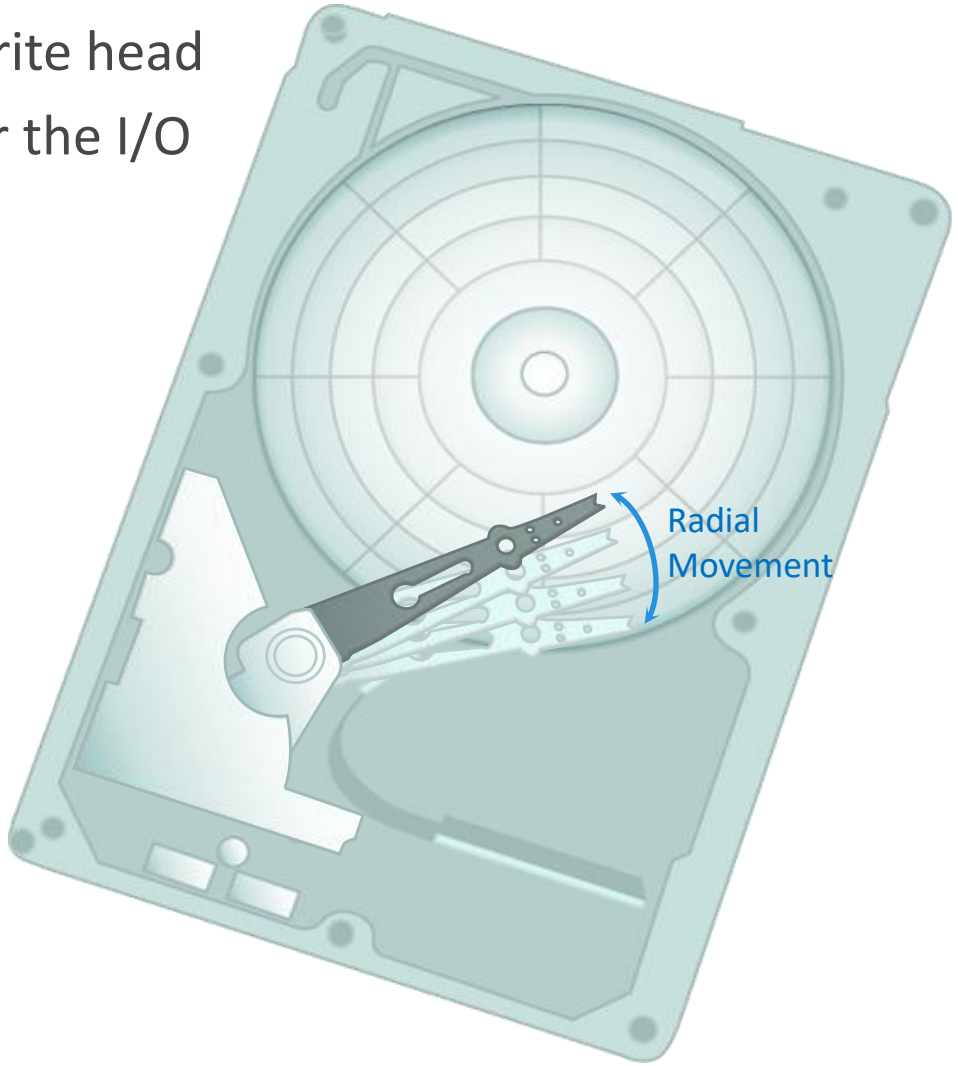
Disk Drive Performance

- Electromechanical device
  - Impacts the overall performance of the storage system
- Disk service time
  - Time taken by a disk to complete an I/O request, depends on:
    - Seek time
    - Rotational latency
    - Data transfer rate

Disk service time = seek time + rotational latency + data transfer time

Seek Time

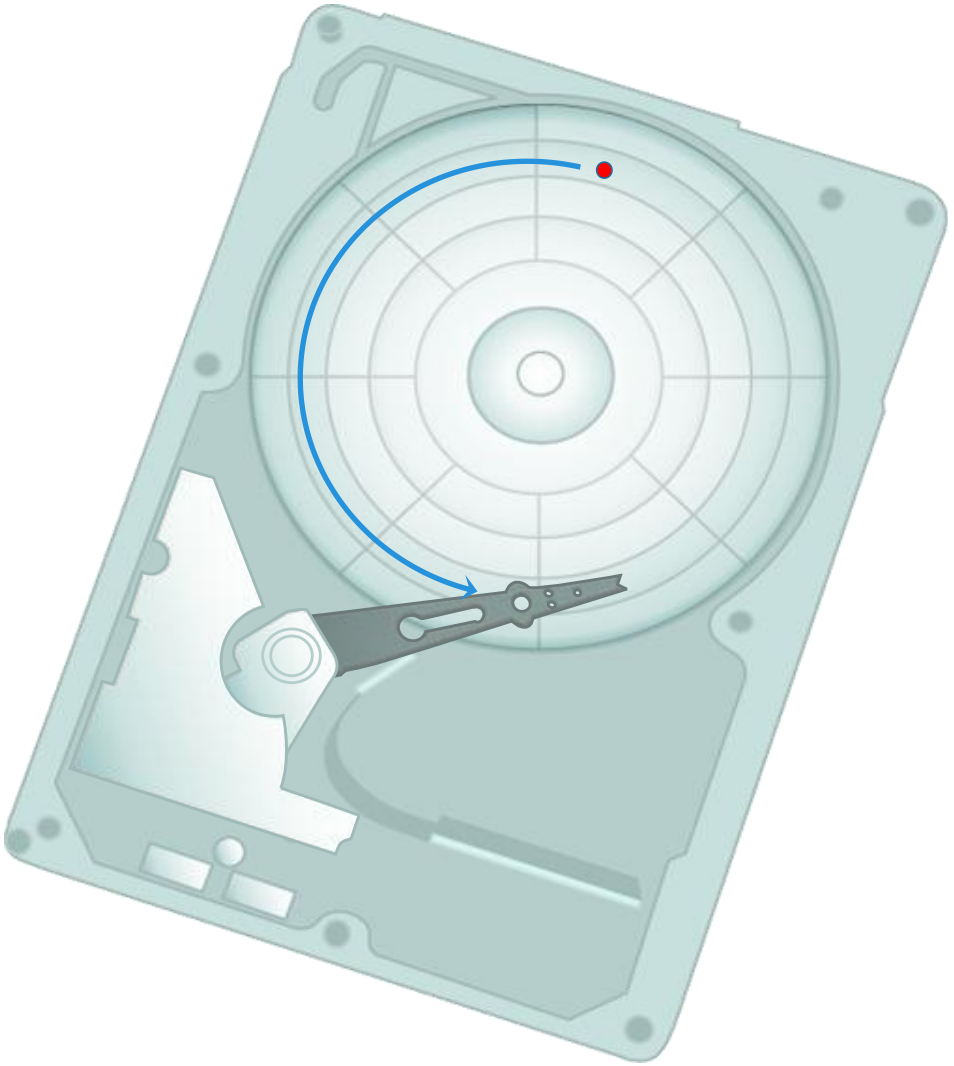
- Time taken to position the read/write head
- The lower the seek time, the faster the I/O operation
- Seek time specifications include
  - Full stroke
  - Average
  - Track-to-track
- The seek time of a disk is specified by the drive manufacturer



Rotational Latency

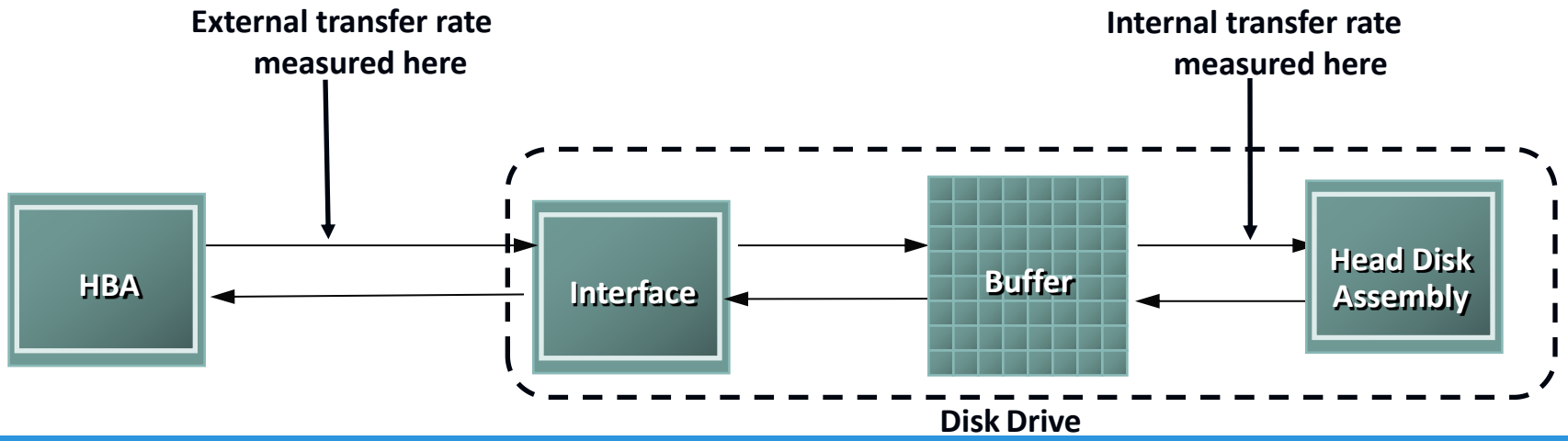
- The time taken by the platter to rotate and position the data under the R/W head
- Depends on the rotation speed of the spindle
- Average rotational latency
  - One-half of the time taken for a full rotation
  - For '*X*' rpm, drive latency is calculated in milliseconds as:

$$= \frac{1/2}{(X/60)}$$



Data Transfer Rate

- Average amount of data per unit time that the drive can deliver to the HBA
  - Internal transfer rate : Speed at which data moves from a platter's surface to the internal buffer of the disk
  - External transfer rate: Rate at which data move through the interface to the HBA

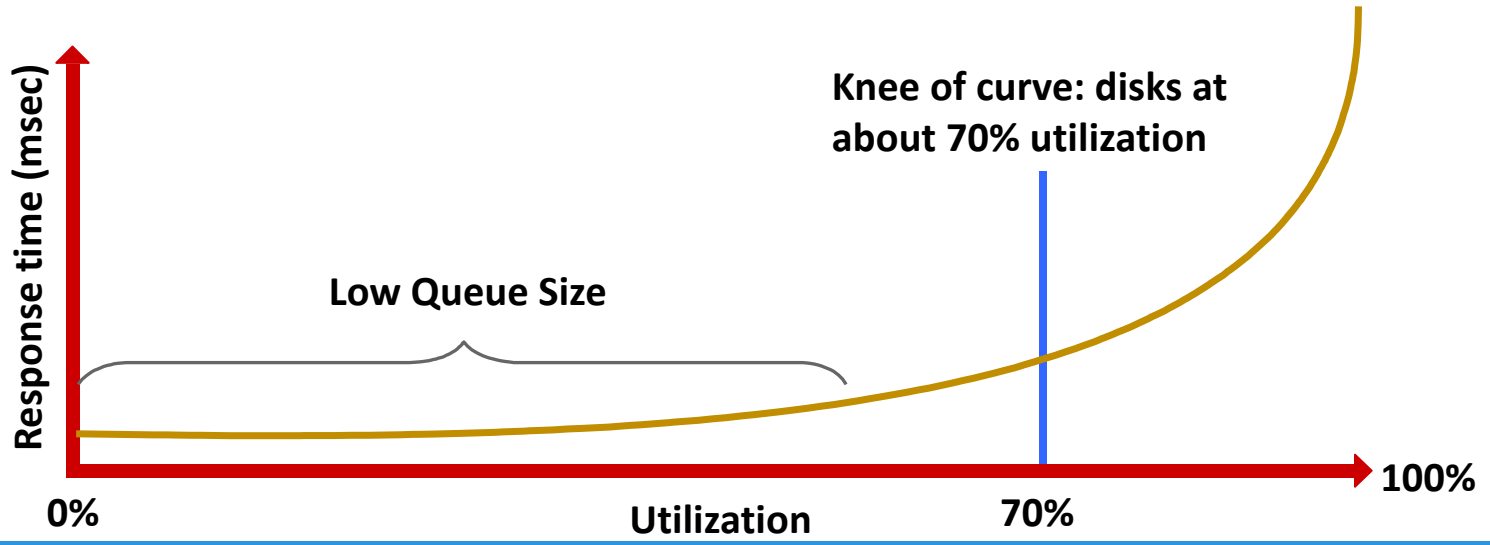


I/O Controller Utilization Vs. Response Time

- Based on fundamental laws of disk drive performance:

$$Av.\,Response\,Time = \frac{Service\,Time}{(1 - Utilization)}$$

  - Service time is time taken by the controller to serve an I/O
- For performance-sensitive applications disks are commonly utilized below 70% of their I/O serving capability



Storage Design Based on Application Requirements and Disk Drive Performance

- Disks required to meet an application's capacity need (D_C):

$$D_C = \frac{Total\ capacity\ required}{Capacity\ of\ a\ single\ disk}$$

- Disks required to meet application's performance need (D_P):

$$D_P = \frac{IOPS\ generated\ by\ an\ application\ at\ peak\ workload}{IOPS\ serviced\ by\ single\ disk}$$

Input/Output Operations Per Second

- IOPS serviced by a disk (S) depends upon disk service time (T_S):

$$T_S = Seek\ time + \frac{0.5}{(Disk\ rpm/60)} + \frac{Data\ block\ size}{Data\ transfer\ rate}$$

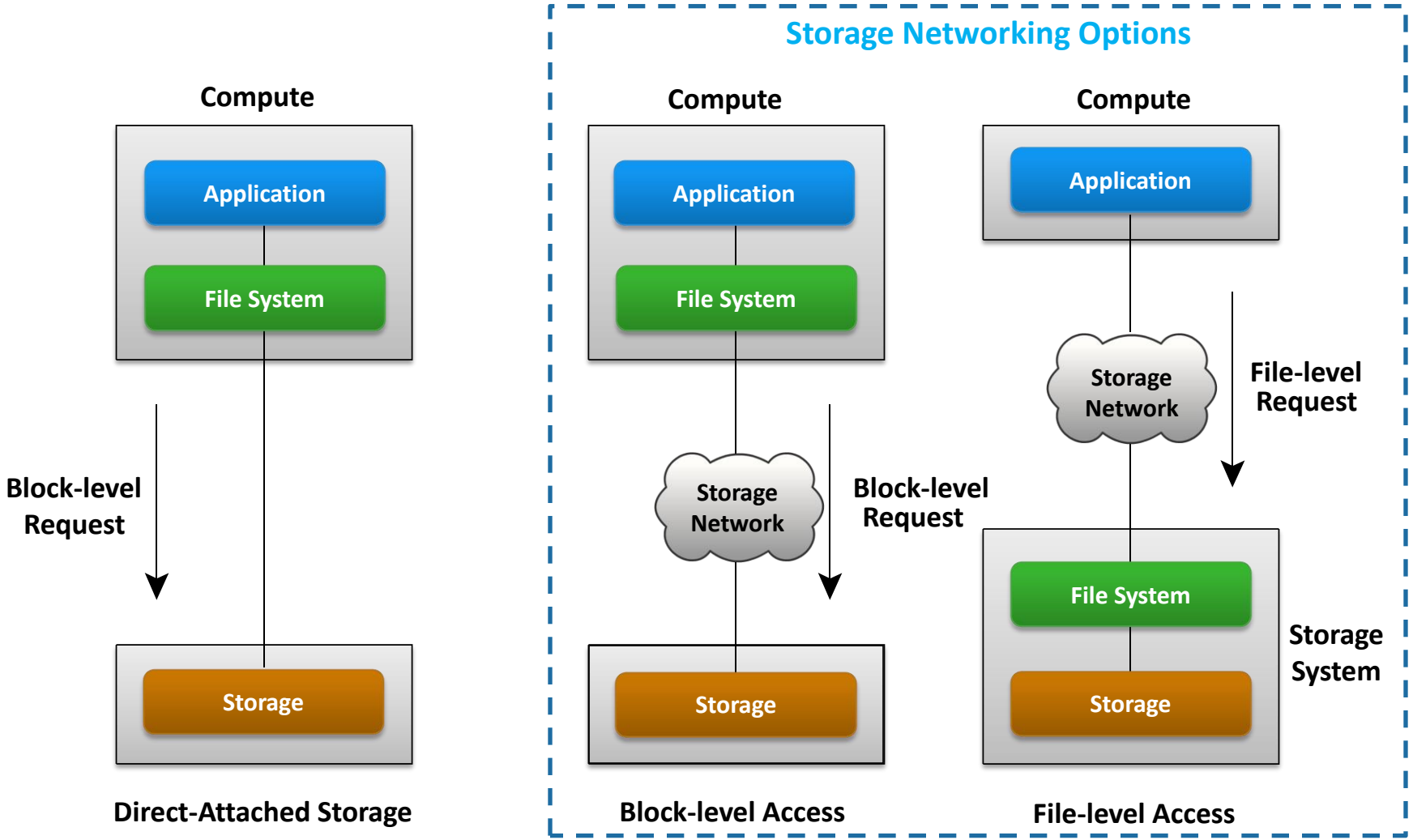
  - T_S is time taken for an I/O to complete, therefore IOPS serviced by a disk (S) is equal to ($1/T_S$)
    - For performance sensitive application (S) $= 0.7 x \frac{1}{T_S}$

Disk required for an application = max (D_C,D_P)

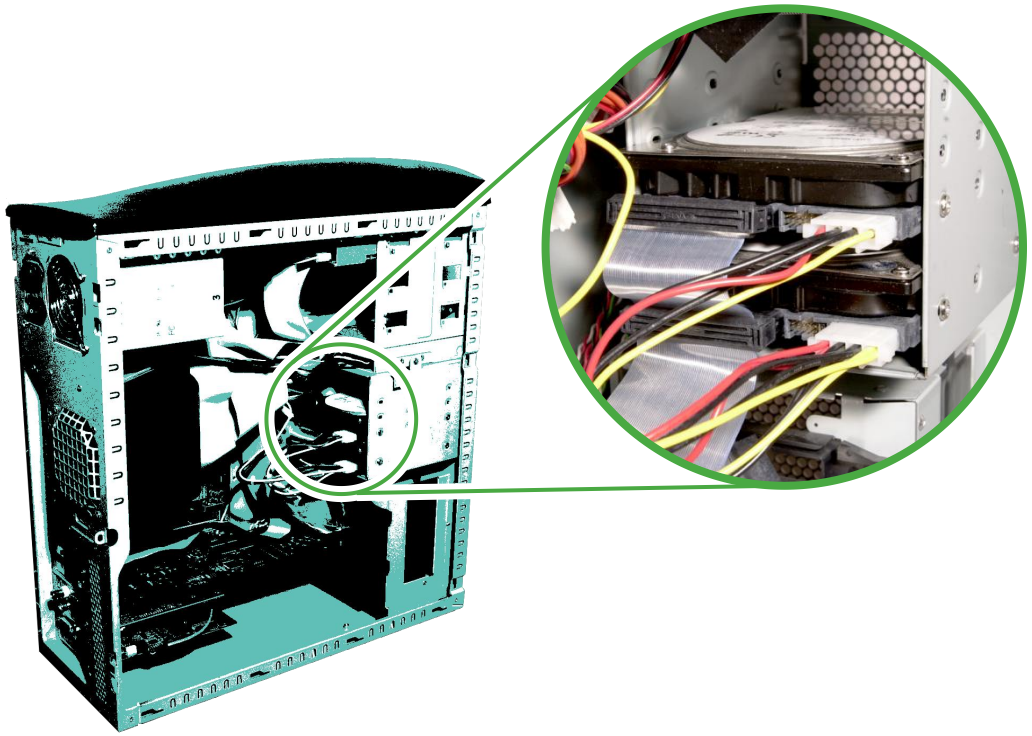
Enterprise Flash Drives

Conventional Hard Drives	Flash Drives
Mechanical delay due to seek time and rotational latency	Highest possible throughput per drive due to no mechanical movement
Limited performance and I/O serving capability	Very low latency per I/O and consistent I/O performance
More power consumption due to mechanical operations	High Energy efficiency • Lower power requirement per GB • Lower power requirement per IOPS
Low mean time between failure (MTBF)	High reliability due to no moving parts
Higher TCO due to more number of disks, power, cooling, and management cost	Overall less TCO

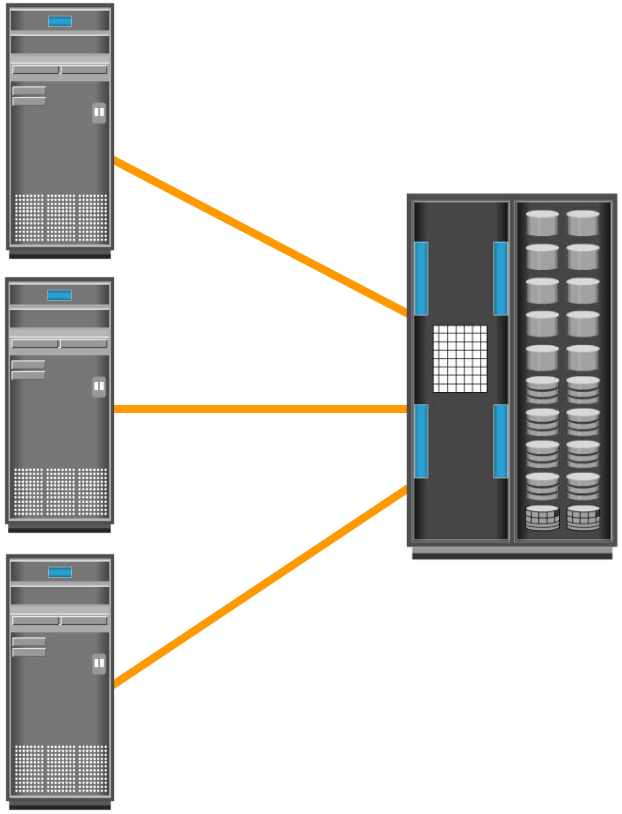
Host Access to Storage



Direct-Attached Storage (DAS)



Internal Direct Connect



External Direct Connect

Concept in Practice

- VMware ESXi

VMware ESXi

- Industry's leading hypervisor
  - Enable virtualization of x86 hardware platforms
- Physical machine that houses ESXi is called ESXi host
  - ESXi host abstracts physical compute resources to run multiple VMs concurrently on same physical server
- Two Components
  - VMKernel
    - Work similar to OS – responsible for process creation, resource scheduling, and so on
  - Virtual machine monitor
    - Performs binary translation for privileged OS instructions that can not be virtualized

Module 2: Summary

Key points covered in this module:

- Key data center elements
- Application and compute virtualization
- Disk drive components and performance
- Enterprise flash drives
- Host access to storage

Exercise: Design Storage Solution for New Application

$$D_P = \frac{IOPS\ generated\ by\ an\ application\ at\ peak\ workload}{IOPS\ serviced\ by\ single\ disk}$$

- Scenario
  - Characteristics of new application:
    - Require 5TB of storage capacity
    - Peak I/O workload 4900 IOPS
    - Typical I/O size is 12KB
  - Specifications of the available disk drives:
    - 15K rpm drive with storage capacity = 500 GB
    - Average seek time = 5ms
    - Data transfer rate = 120 MB/sec
  - As it is business **critical application**, response time must be within acceptable range
- Task
  - Calculate the number of disks required for the application

$$0.7 x \frac{1}{T_S}$$

$$T_S = Seek\ time + \frac{0.5}{(Disk\ rpm/60)} + \frac{Data\ block\ size}{Data\ transfer\ ra}$$

$$= \frac{1/2}{(X/60)}$$

知识测验 – 1

- 以下哪一项是计算虚拟化的好处？
  - A. 支持计算内存交换
  - B. 提高计算利用率 ♥
  - C. 将计算内存与应用程序隔离
  - D. 将计算 OS 与应用程序隔离

- 哪一项是对虚拟机 (VM) 的最佳描述？
  - A. 物理服务器上的所有虚拟机都必须运行同一 OS
  - B. 当虚拟机关闭时会删除虚拟机文件
  - C. 虚拟机是独立的文件组 ♥
  - D. 所有虚拟机平等共享可用资源

知识测验 – 2

- 什么是合并？
  - A. 将多个物理驱动器分组到逻辑驱动器
  - B. 将物理驱动器分为多个逻辑驱动器
  - C. 在逻辑驱动器上写入磁盘元数据的过程
  - D. 通过碎片整理向物理驱动器添加更多容量

- 哪些因素决定机械磁盘的总体服务时间？
  - A. 磁盘缓冲时间、全程和旋转延迟
  - B. 内部传输速度、外部传输速度和缓冲时间
  - C. 全程、平均寻道时间和道间寻道时间
  - D. 平均寻道时间、旋转延迟和数据传输速度

知识测验 – 3

- 哪一项是 DAS 环境的挑战？
  - A. 性能低
  - B. 可扩展性有限 ♥
  - C. 部署复杂性
  - D. 最初投资过多