# ISM

## Unit - 1

Data → Collection of raw facts & figures.
Information → useful derived data.

DBMS vs File Processing :-
1) DBMS removes redundancy & inconsistency.
2) DBMS provides better security.
3) DBMS has concurrency control.

## Topic-1 Introduction to Information Storage :-

* Information → Business uses Processed data to derive information that is critical to their day to day activities.
* Storage → Storage is a repository that enables users to store and retrive the digital data.

(i) Data :- Data is a collection of raw facts from which conclusions may be drawn.
Example → Handwritten notes, printed books, family photograph, a movie on video tape, bank ledgers, etc.

Before the advent of computers, procedures and method adopted for data creation and sharing were limited to fewer forms such as paper and files.

After the advancement of computer, data can be generated from ebok, email message, digital movie, etc.

With the advancement of computer, the rate of data generation and sharing has increased exponentially. Following are the factors contributed towards this growth :-

(i) Increase in data processing capabilities :- Modern computer provides increase in processing & storing capacities. This enables conversion of various type of content from conventional form to digital form.

(ii) Lower cost of digital storage :-

(iii) Affordable & faster communication technology :-

(iv) Proliferation of application & smart devices :-

## (ii) Types of Data :-

Data can be classified as structured and unstructured based on how it is stored and managed.

| Structured Data | Unstructured Data |
|---|---|
| → It is that data which can be organised in rows & columns. | → It is that data which can not be orsaganised in rows & columns. |
| → It is easy to access the required data. | → It is difficult to access. |
| → DBMS is an example of structured data. | → Sticky notes, e-mail, business cards, audio, video, forms, instant message, invoices, etc are examples |

(iii) Big Data :- Big Data is a new and evolving concept which refers to the data sets whose size are beyond the capability of commonly used software tools.
It includes both structured and unstructured data.

Big Data Ecosystem Consist of :-
(i) Devices which collect data from multiple locations and also generate new data about the data
(ii) Data collectors who gather various offline data :-
(iii) Data aggregators that compile the collective data to abstract meaningful information.
(iv) Data users & buyers who benifits from information collected & aggregated by others.

(iv) Information :- Data, whether structured or unstructured, does not fulfil any purpose for individuals or business unless it is presented in a meaningful form. Information is the intelligence and knowledge derived from data.

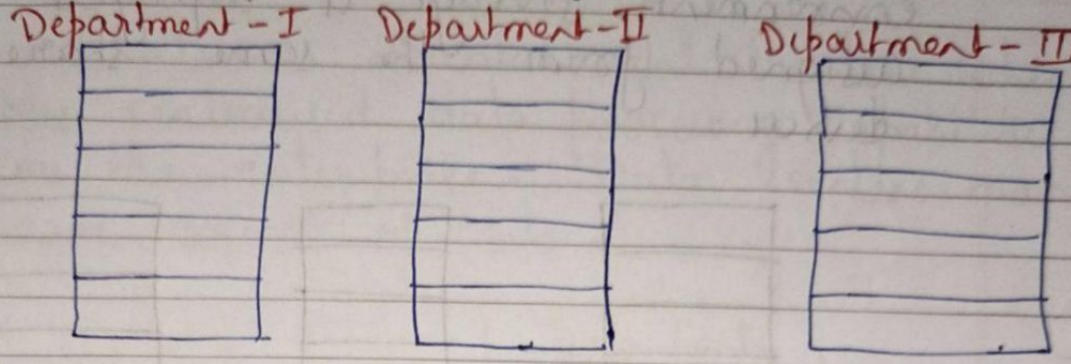
(v) Storage :- Data created by individuals or businesses must be stored so that it is easily accessible for further processing. In computing environment devices designed for storing are termed as storage devices. The type of storage used varies based on type of data and the rate at which it is created and used.

Eg. DVD, CD-ROM, memory in phone, hard disk, etc.

## Topic-2 Evolution of Storage Architecture :-

Initially, we used (i) Server-Centric Storage Architecture.
Nowadays, we are using (ii) Information-Centric Storage Architecture.

## (i) Server - Centric Storage Architecture :-




In this architecture, the storage was typically internal to the server and these storage devices could not be shared by other servers.

Drawbacks :-

i) Each server has limited no. of storage devices and any administrative task such as maintainence of server or increasing capacity might result in unavailability of information.

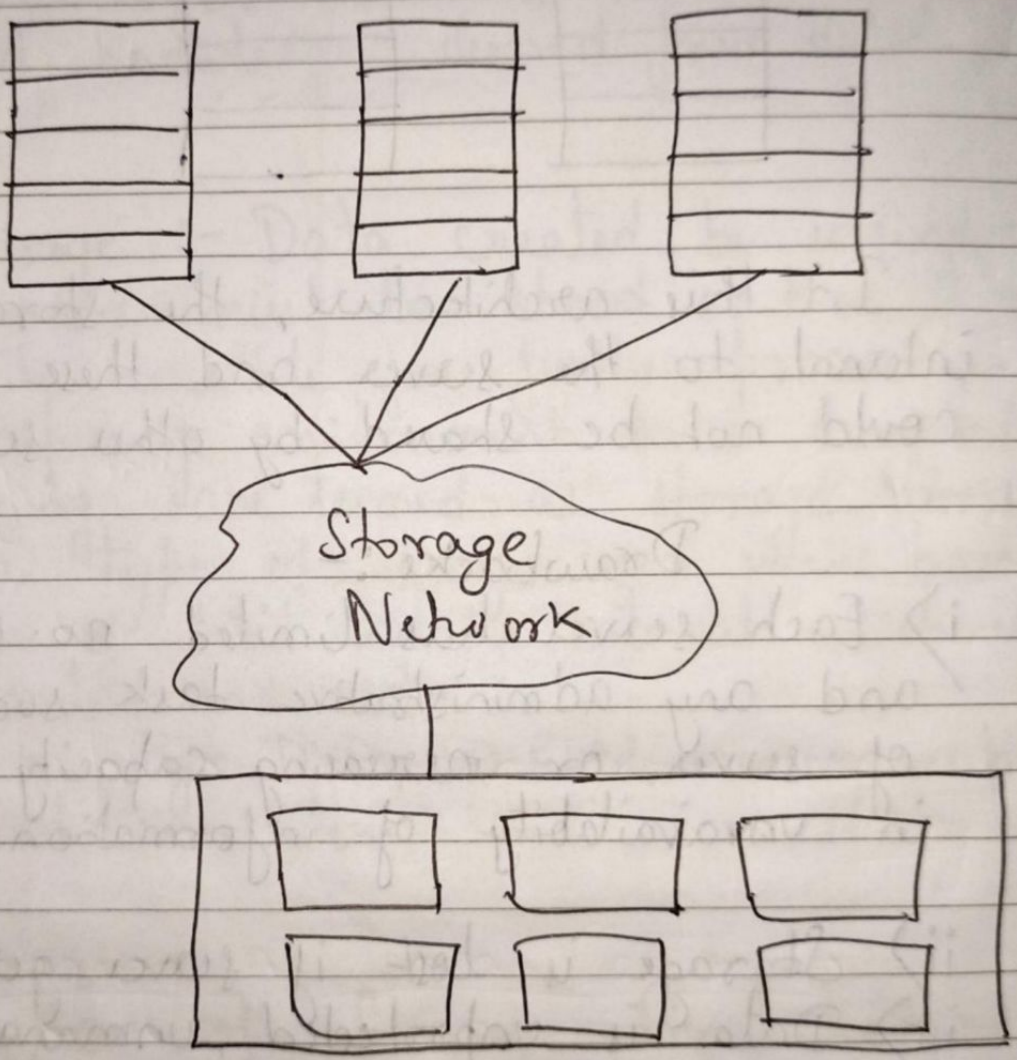
ii) Storage is lost if server gets down.

iii) Data is unprotected, unmanaged and cost of operation is high.

## (ii) Information - Centric Storage Architecture :-

In this architecture, all servers shares same common data storage system. Basically data is managed centrally & independent of server.

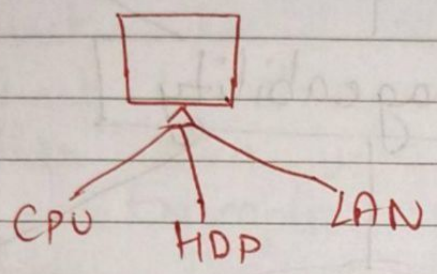
Whenever, new server is deployed in the environment, storage is ~~deployed in the~~ assigned from the same shared storage device.




Adv :- Just reverse of all the disadvantage we studied in Sans-Central.

# Topic-3 :- Data Center Infrastructure

Organizations maintain data centers which provides centralised data processing capabilities across the enterprise. Data centers manages a large amount of data.

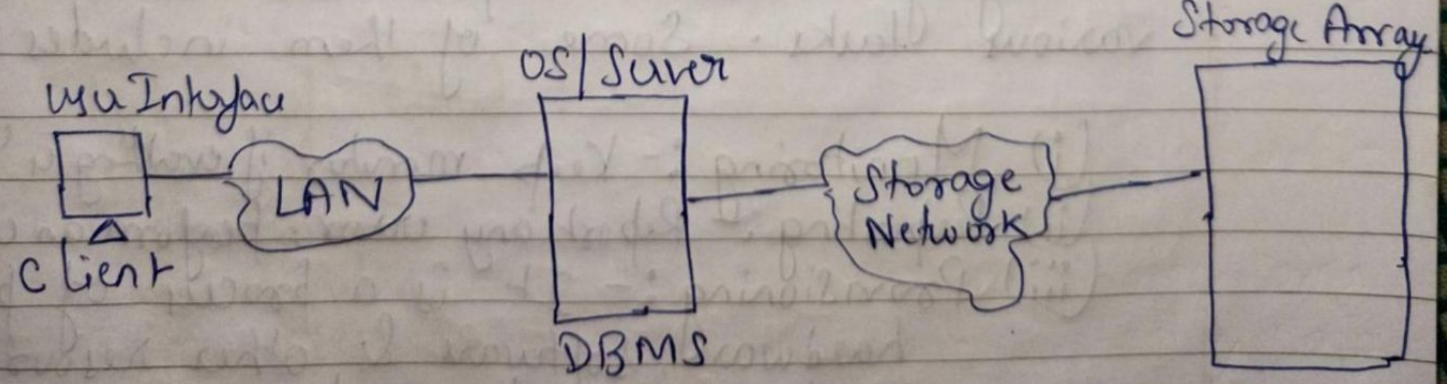



Data Center :-
Server, Storage devices & Power backup. → H/W component
S/W, App / OS → S/W component
Fire suppressor, Ventilator, AC, etc. → Environmental condition

Core-Element of Data Center :-

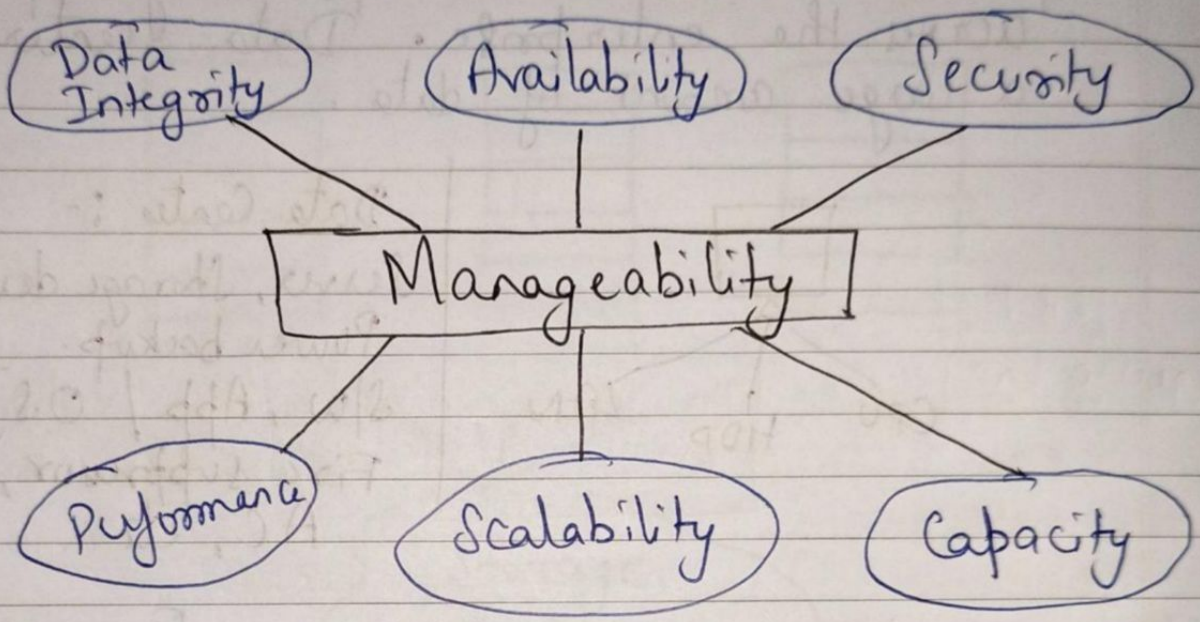
(i) Computer or Host :-
(ii) DBMS :-
(iii) Storage devices :-
(iv) Network :-
(v) Application :-

→ You can Explain it easily in your own words.




Example of an order processing system

# Key Characteristics / Requirement of Data Center :-




Explain these in your own words. Easy

## Managing a Data Center :-

Managing a modern, complex data center involves various tasks. Some of them includes :-

(i) Monitoring :- Keep monitor if everything is working.
(ii) Reporting :- Report any error, performance.
(iii) Provisioning :- It is a process of providing hardware, software & other resources needed by data center.

# Topic 4 → Cloud Computing & Virtualization

Cloud Computing :- Cloud computing is the delivery of computing resources, servers, databases, storage and intelligence over the internet
It is the on-demand resource server over the internet. It is use of remote server over internet to store, manage and process data rather than using local server.
Examples of cloud servic providers includes Google cloud, AWS, IBM cloud, Microsoft Azure, etc.

## Benifits of cloud Computing :-

(i) Reduced Investment :- No need to buy private serva.

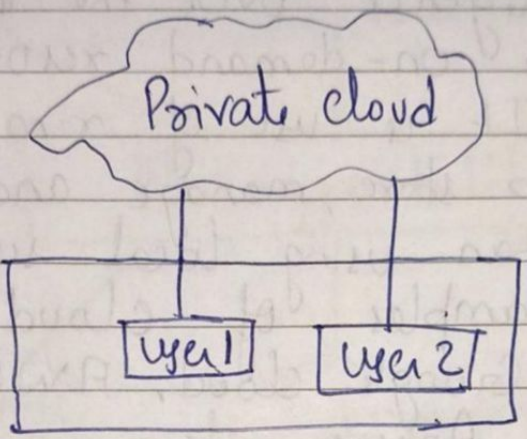
(ii) Increased scalability :- Rent out or release according to need.

(iii) Increased availability & Reliability :- fault tolerance

## Types of cloud :-

(i) Private
(ii) Public
(ii) Hybrid
(iv) Community.

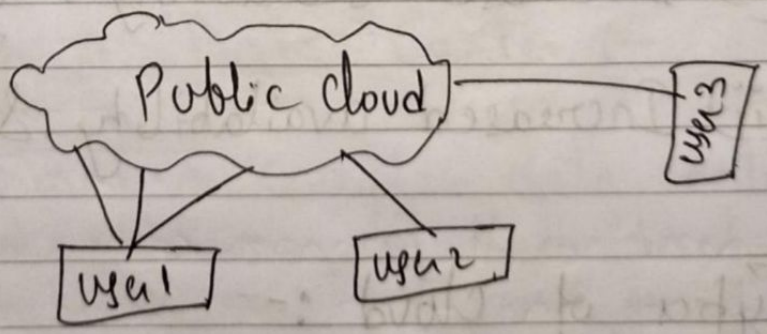
(i) Private :- Those cloud in which services are available within an organization only are called private cloud. They are managed by either third party or particular organization.




user3
can't access

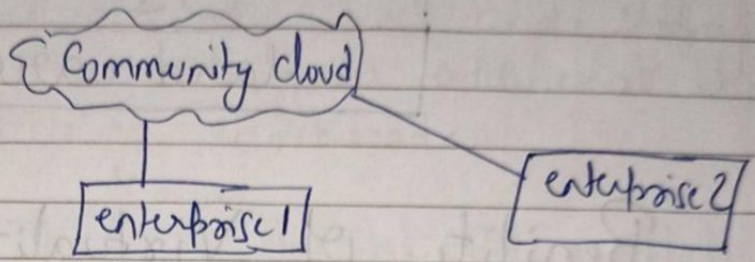
Benifit :- i) Security
ii) More customizable than public

(ii) Public :- Anyone can access. One user can access multiple resources.




(iii) Hybrid :- Public + Private
↓ for non-critical activity
↓ for critical activity.

(iv) Community :- Group of organization form a cloud and used by any enterprise.

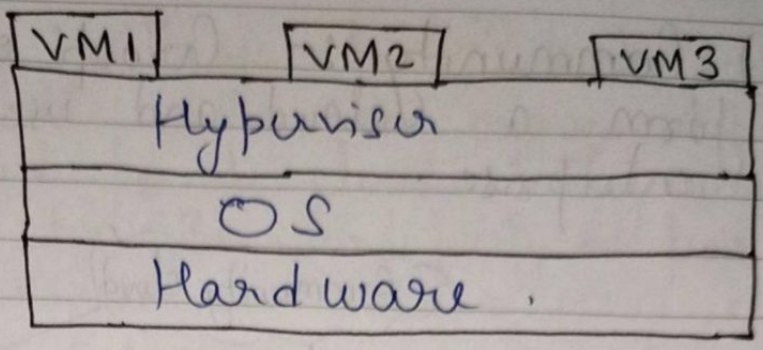



Drawback :- i) Data sharing.

Virtualization :- It is a technique which allows to store single physical instants of an application/resource among multiple organization or customers.
All virtual resources will work independently.

Hypervisor is the software that perform virtualization. Eg. VM Player.

* Host machine :- It is a machine on which virtual machine is going to be built.
* Guest Machine :- It is a virtual machine.
* Hypervisor :- It is a software that creates and run the virtual machines. It is used to create virtualization on physical machines. It is also known as virtual Machine Monitor (VMM).




## Benifits of Virtualization :-

(i) Better resource utilization
(ii) Lower the count of IT infrastructure.
(iii) Remote access
(iv) Pay-per view of IP infrastructure on demand.
(v) Enable running multiple OS.
(vi) If one VM is not working or having any problem others will not be affected.

## Topic-5 Connectivity :-

It is the interconnection between a host or between a host and peripheral devices such as printer or storage device.

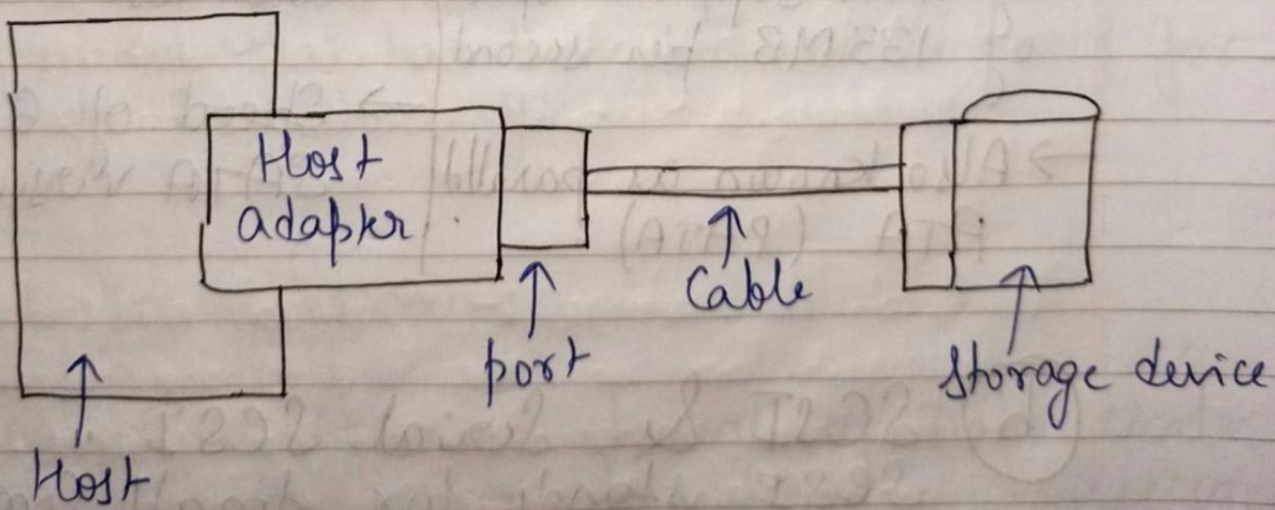
### Physical Component of Connectivity :-

It is the hardware elements that connects the host to storage.

(i) Host Interface device :- It is an application specific integrated circuit board that performs I/P-O/P operation interface function between the host and storage delivering the CPU from additional input output processing workload.

(ii) Ports :- It is a specialized outlets that enables connectivity between host and the external devices.
An Host Burst adapter contain one or more ports to connect to storage devices.

(iii) Cable :- Cable connects host to internal or external devices using copper or fibre optic medium.



# Interface Protocols :-

A protocol enables communications between the host and storage.
Following are the popular interface protocol used for host to storage communication :-

(a) IDE/ATA & Serial ATA :- It stands for Integrated Device Electronics / Advanced Technology Attachment.

| IDE/ATA | Serial ATA |
|---|---|
| → used when storage device is CDROM/DVD | → Supports single bit serial transmission |
| → Ultra DMA/133 version of ATA supports speed of 133 MB per second | → High performance & low cost |
| → Also known as parallel ATA (PATA) | → Speed of 6 GB/s in SATA version 3.0. |

(b) SCSI & Serial SCSI.
SCSI stands for Small Computer System Interface.

* This protocol supports parallel transmission and offers better performance, Scalability & compatibility than IDE/ATA.

* It supports upto 16 device on a single bus & data transfer rate of 640 Mbps.

Serial SCSI :- It is a point-to-point serial protocol that provides an alternative to ~~serial~~ parallel SCSI. A new version of Serial SCSI support data speed of 6 GBps.

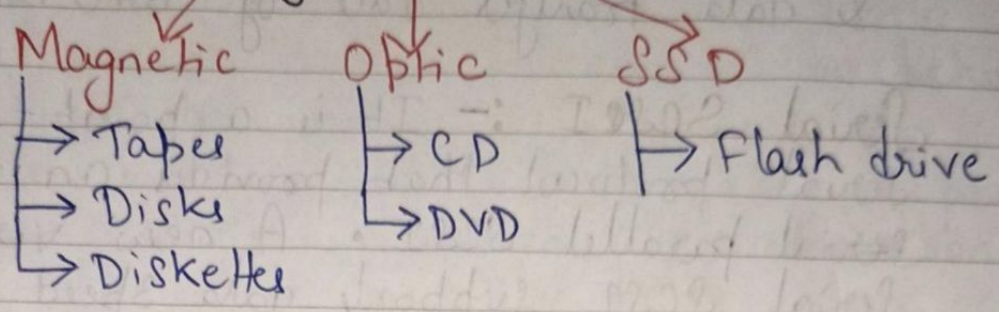
(c) Fibre Channel :- It is widely used for high-speed communication to storage device.
* It provide gigabit network speed.
* It provide a serial data transmission that operate over copper wire & optical cable fiber.
* Latest version of the Fiber channel interface allows transmission upto 16 GBps.

(d) Internet Protocol :- IP is a network protocol that has been traditionally used for host to host traffic.

## Topic-6 :- Storage

Storage is the core element of Data center used to store the data. Storage devices uses magnetic, optic or Solid State Media.

# Storage Devices




* In the past, tapes were the most popular storage option for backups due to their low cost. However, there were few drawbacks of this in terms of performance & management :-

(i) Takes more time to access data due to sequential search and retrival.

(ii) Two devices cannot simultaneously access the tape.

(iii) Over the period of time, the tape will be degraded or wear out as the read/write head touches the tape surface.

* Optical ~~Cables~~ :- Optical disc storage is popular in small, single user computing environments like games. It works on Writ Once Read Many technology.

Drawback :- Modification is not allowed.

* SSD :- fastest means of communication.

## Disk Drive Components :-

The key components of a hard disk drive are platter, spindle, read-writ head, actuator arm assembly & controller board.

### (i) Platter :-

* A typical HDD consist of one or more flat circular disk called platter. Data is recorded on those platter in Binary form.
* Set of rotating platters ~~on those~~ is sealed in a case called as Head Disk Assembly (HDA).
* A platter is rigid, round disk coated with magnetic material on both sides. Data can be written to or read from both surface of the platter.
* The no. of platter and storage capacity of each platter determine total capacity of drive.

### (ii) Spindle :-

* It connects all the platter and is connected to a motor. The motor of the spindle rotates with a constant speed.
* The disk platter spins at the speed of several thousands ~~per~~ revolution per ~~secon~~ minute.
* Common spindle speeds are 5400 rpm, 7200 rpm, 10,000 rpm, 15000 rpm.

## (iii) Read-Writ Head :-

* Read/writ head reads and writes data from or to platters. Drive has 2 read/writ head per platter, one for each surface of platter.
* While reading the data, the head detects the ~~fr~~ magnetic polarization on the surface of the platter.
* During read and write the R/w head senses the magnetic polarization and never touches the surface of the platter.
* When spindle is moving there is a microscopic air gap maintained between the R/w head and platter known as head flying height. This air gap is removed when the spindle stops rotating and R/w head rests on a special area on the platter near the spindle. This area is called landing ~~area~~ zone.
* The landing zone is coated with lubricant to avoid friction between the head & platter.
* If the drive malfunctions the read writ head accidently touches the surface the platter ~~the~~ outside the landing zone, a head crash occur.
* In head crash, the magnetic coating on platter is scratched and may cause damage to R/w head. A head crash generally results in data loss.

(iv) Actuator Arm :- Read/write head are ~~accur~~ mounted on the actuator arm assembly, which position the Read/write head at the location on the platter where data needs to be written or read.

The read-write head of all platters on a drive are attached to one actuator arm assembly and move across the platter simultaneously.

(v) Controller Board :- It is a printed circuit board mounted at the bottom of disk drive. It consist of a microprocessor, internal memory circuitry and firmware.

The firmware control the power to the spindle motor and speed of the motor. It also manages the communication between the drive and the host. In addition, it control the R/w Operation by moving the actuator arm and switching between the different R/w head and performs optimization of data access.

## Physical Disk Structure :-

Explain previous Component in this including Cylinders.

Cylinder → It is a set of identical tracks on both surface of each drive platters. The location of R/w head is reffered to ~~as~~ by cylinder number not by the track number.

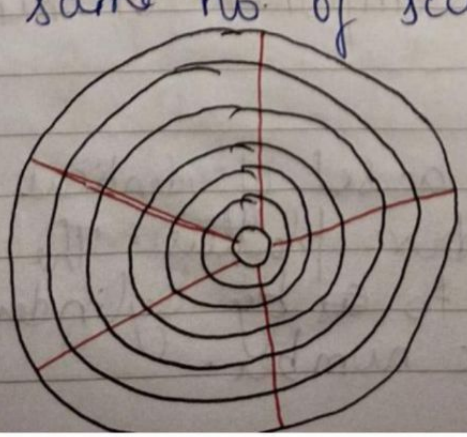
# Zoned Bit Recording :-

Platters are made of concentric tracks, the outer tracks can hold more data than the inner tracks because the outertracks are physically longer than the inner track.

On older disk, the outer track had the same no. of sectors than the inner tracks so the data density was low on outer tracks. This was an inefficient use of available space.

Zoned bit recording use the disk efficiently. This mechanism group tracks into zones based on their distance from the center of the disk. The zone are numbered, with outermost zone being zone 0. An appropiate no. of sector per tracks are assigned to each zone, so a zone near the center of the platter has fewer sector per track than a zone on the outer edge.

Tracks within a particular zone have the same no of sectors.



Without Zone

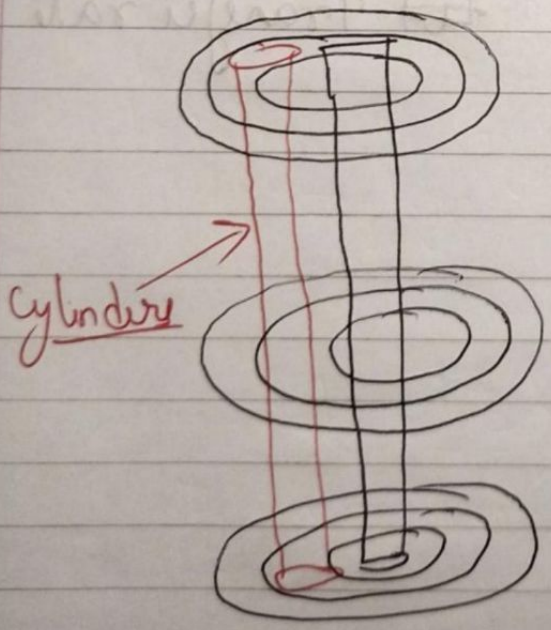
With Zone

# Logical Block Addressing :-

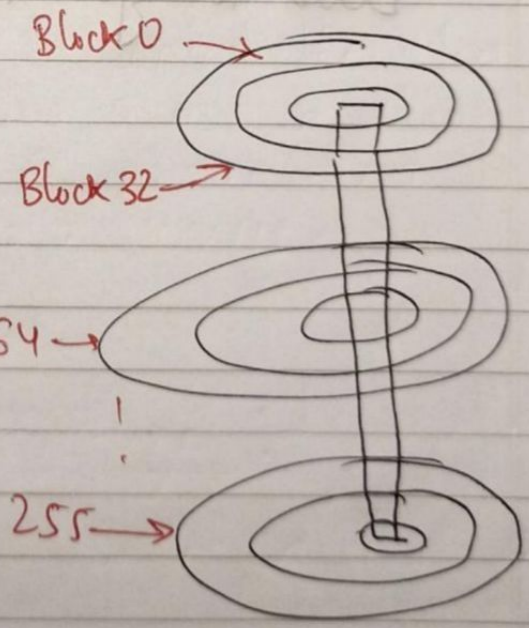
Earlier drive use physical addresses consisting of cylinders, Heads and Sector (CHS) number to refer to specific location on the disk.

The host operating system had to be aware of the geometry of each disk used.

Logical Block Addressing (LBA) simplifies addressing by using a linear address to access physical blocks of data. The disk controller translates the LBA to CHS address and the host only need to know the size of disk drive in term of no. of blocks.




CHS




Block 0-255 → 256 block

LBA

# Disk Drive Performance :-

Factors affecting disk drive performance:-

i) Seek time :- time taken by read/write head to reach the desired track.

ii) Rotational Latency :- Time taken to reach the desired sector

iii) Data Transfer rate :- Refers to the avg. amount of data per unit time that the drive can deliver to HBA.

Disk Access time = Seek time + Rotational Latency + Data transfer time

Data transfer time = $\frac{\text{Data to be transferred}}{\text{Transfer rate}}$.

→ Actuator Arm: R/W heads are mounted on actuator arm which positions the R/W head on location.

* Disk Drive Performance:

① Factors affecting the performance: → Disk Service Time.
→ Disk I/O Controller Utilization.

Disk Service time: Time taken by the disk to complete an I/O Request.
→ Seek time: time taken to position R/W and settle the arm & the head over the correct track. ↓ Seek time, faster the I/O operation.
→ Rotational latency: time taken by the platter to rotate and position the data under R/W head.
→ Data transfer rate: Avg. amount of data per unit time that the drive can deliver.

* Intelligent Storage Systems:

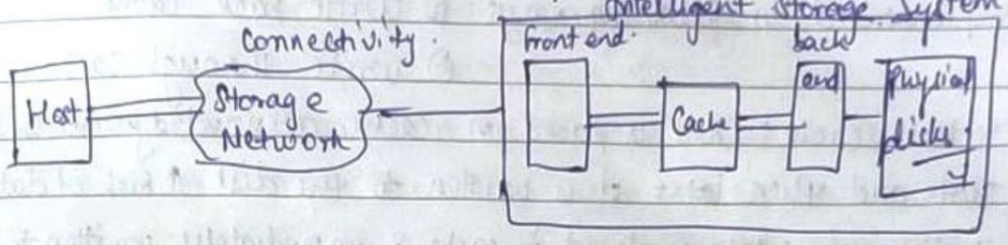
① Rich RAID arrays that provide highly optimized I/O processing capabilities.
② These storage systems are configured with a large amount of memory and multiple I/O paths.
③ Use sophisticated algorithms to meet the requirements of performance sensitive applications.
④ Have an operating environment that intelligently and optimally handles the management, allocation and utilization of storage resources.

* Components of Intelligent Storage System:

→ Front end
→ Cache
→ Back end
→ Physical disks

① An I/O request received from the host at the front end port

is processed through cache and back end, to enable storage and retrieval of data from physical disk. A read request can be serviced directly from cache if the requested data is found in the cache.




→ Front end: ① Provides the interface b/w storage system & host.
② Consits of 2 components: ⓐ Front-end ports ⓑ Front end controllers.
Each controller contains multiple ports that enable large numbers of hosts to connect to Intelligent Storage System.
③ Each front end controller has processing logic that executes appropriate transport protocol.
④ Front-end controllers route data to and from cache via Internal Data Bus

→ Cache: ① Semiconductor memory where data is placed temporarily to reduce the time required to service I/O requests.
② It improves storage system performance by isolating hosts from mechanical delays associated with rotating disks or Hard Disk Drives (HDD). Rotating disks are the slowest Components of SS.
③ Accessing data from cache is faster. It is organized into pages.

Read operation with cache:

→ When a host issues a read request, the storage controller reads the tag RAM to determine whether the required data is available.
→ If available, then it is known as read cache hit/read hit and the data is sent directly to the host.
→ If not available then it is cache miss and data must be read from the disk. The back end accesses the appropriate disk. The data is then placed in cache and then finally sent to the host through front end.

Write Operation with Cache:
① Provides performance advantages over writing directly to disks.
② Completed in less time.
③ Implemented in following ways: ⓐ Write back cache
ⓑ Write through cache.

Write back cache: Data is placed in cache, acknowledgement is sent to host and then later it is written to the disk. ↑ Risk of data loss
Write through: Data is placed in cache & immediately written to the disk, then acknowledgement is sent to the host. ↓ Risk of data loss.

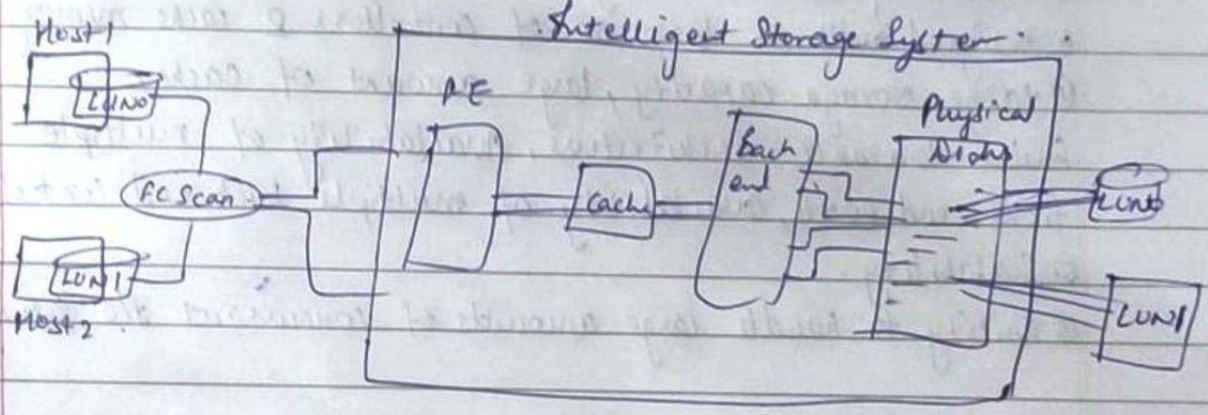
The risk of losing uncommited data in cache can be mitigated using:
→ Cache mirroring: ① Each write two cache is held in 2 different memory locations on 2 independent memory cards. If cache failure occurs, then the data will still be safe in mirrored location.
→ Cache vaulting: Powering the memory with a battery. Using a battery to write cache content to the disk.

→ Back end: ① Provides an interface b/w cache & physical disks.
2 components: back end ports & back end controllers.
② Back end controller communicates with the disks when performing reads and writes and also provides additional, but limited, temporary data storage, provide error detection and correction along with RAID functionality.

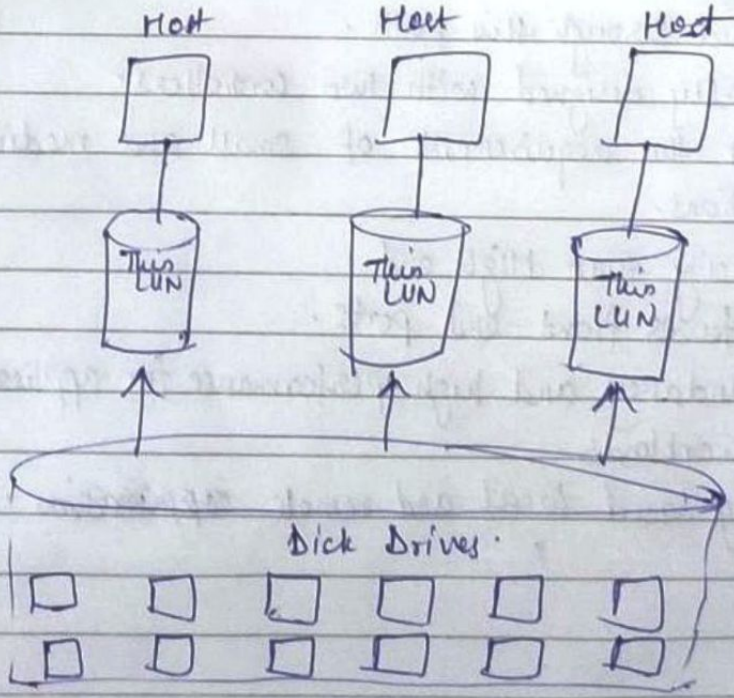
→ Physical Disks: ① Connected to back end Storage Controller & provide persistent data storage.

* Storage Provisioning: The process of assigning storage resources to hosts based on capacity, availability & performance Requirements.
Done in 2 ways: ⓐ Traditional
ⓑ Virtual.

Traditional Storage Provisioning: ① Physical disks are logically grouped together and a required RAID level is applied to form a set (RAID set).
② The no. of drives in the RAID set and the RAID level determines the availability, capacity, performance, etc.
③ Logical units are created from RAID sets.
④ Each logical unit created is assigned a unique ID called logical Unit Number (LUN). Then this is known as thick LUN.




Virtual Storage Provisioning: ① Enables creating and presenting a LUN with more capacity than is physically allotted to storage array.
② This LUN distinguish it from traditional LUN.
③ More efficient allocation.



## * Types of Intelligent Storage Systems:

① High-end storage systems
② Midrange storage systems.

High-end SS → ① Refessed to as active-active arrays
② Generally aimed at large enterprise applications.
③ Designed with a large no. of controllers & cache memory.
④ Large storage capacity, large amount of cache, fault tolerance architecture, Availability of multiple front-end ports, availability of multiple back end controllers
⑤ Scalability.
⑥ Ability to handle large amounts of concurrent I/O from host

Mid Range SS → ① Active-Passive arrays.
② Provide optimal storage solutions at lower cost.
③ The hosts can perform reads/writes to the LUN only through the path through controller A b/c controller A is the owner of that LUN. The path to controller B remains passive & no I/O activity is performed through this path.
④ These are typically designed with two controllers.
⑤ Designed to meet the requirements of small and medium enterprise applications.
⑥ Less storage capacity than High end.
⑦ There are also fewer front-end ports.
⑧ Ensure high redundancy and high performance for applications with predictable workloads.
⑨ Also support array based local and remote replication.

(*) Data → Structured
→ Unstructured

Structured → ① · Easily Readable.
② · In form of rows and columns
③ · Example → In DBMS.

Unstructured → ① · Not Easily Readable
② · Not organized.
③ · Example → In sticky Notes or in emails.

(*) · Big Data → Data whose size is larger than the capacity of the storage system

Types of Virtualization:

① Application Virtualization
② Network Virtualization
③ Desktop Virtualization
④ Storage Virtualization.

* Data Center Environment: Application.

① Application is a computer program that provides logic for computing operations.
② Application sends request to the OS to read/write operations on storage devices.
③ Applications can be layered on OS which in turn uses OS services to perform R/W operations.
④ Commonly categorized as: business applications, Infrastructure management applications, data protection applications, security applications.
examples: email, resource management backup, authentication, antivirus applications, etc.

* Data base Management System:

① Database is a structured way to store data in logically organized tables that are interrelated.
② Helps to optimize storage and retrieval of data
③ Controls the creation, maintenance and use of database.
④ It processes an application's request for data and instructs the OS to transfer the appropriate data from the storage.

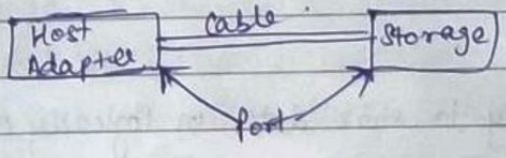
* Host

① Users store and retrieve data through applications. The computers on which these applications run are known as hosts / compute systems.
② Hosts can be physical or virtual machines.
③ Examples of physical hosts include: desktop computers, servers or a cluster of servers, laptops, mobile devices, etc.

① A host consists of CPU, memory, I/O devices and a collection of software to perform operations.
② This software includes OS, file systems, logical volume manager, device drivers, etc.

Host : ① Operating System
② Device Driver
③ Volume Manager
④ File System.

* Connectivity : ① Refers to the interconnection b/w hosts or b/w hosts and peripheral devices such as printers & other storage devices.
→ Physical components (Hardware elements) that connect host to storage).
Components are : ① Host Interface device.
② port
③ cable




* Storage : ① Core component in data center.
② Storage device uses magnetic, optic or solid state media.
③ Disks, tapes, diskettes → use magnetic media.
CD/DVD → use optical media for storage.
④ Removable flash memory or flash drives are examples of solid state media.

* Disk Drives : ① Most popular storage medium used in modern computers for storing and accessing data.
② Disks support rapid access to random data locations. Data can be written or retrieved quickly for a large no. of users / applications simultaneously.