



ISM Unit 1 part 1 - This is unit 1 of ISM handwritten notes

Information Storage And Management (SRM Institute of Science and Technology)

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 - I 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 user to store and retrieve 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 videotape, bank ledgers, etc. are

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

After the advancement of computer, data can be generated from ebook, 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

- It is that data which can be organised in rows & columns.
- It is easy to access the required data.
- DBMS is an example of structured data.

Unstructured Data

- It is that data which can not be organised in rows & columns.
- It is difficult to access.
- Sticky notes, e-mail, business cards, audio, video, forms, instant message, invoice, 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 consists of :-

- Devices which collect data from multiple locations and also generate new data about the data.
- Data collectors who gather various offline data :-
- Data aggregators that compile the collective data to abstract meaningful information.
- Data users & buyers who benefit from information collected & aggregated by others.

(iv) Information :- Data, whether structured or unstructured, does not fulfil any purpose for individual 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.

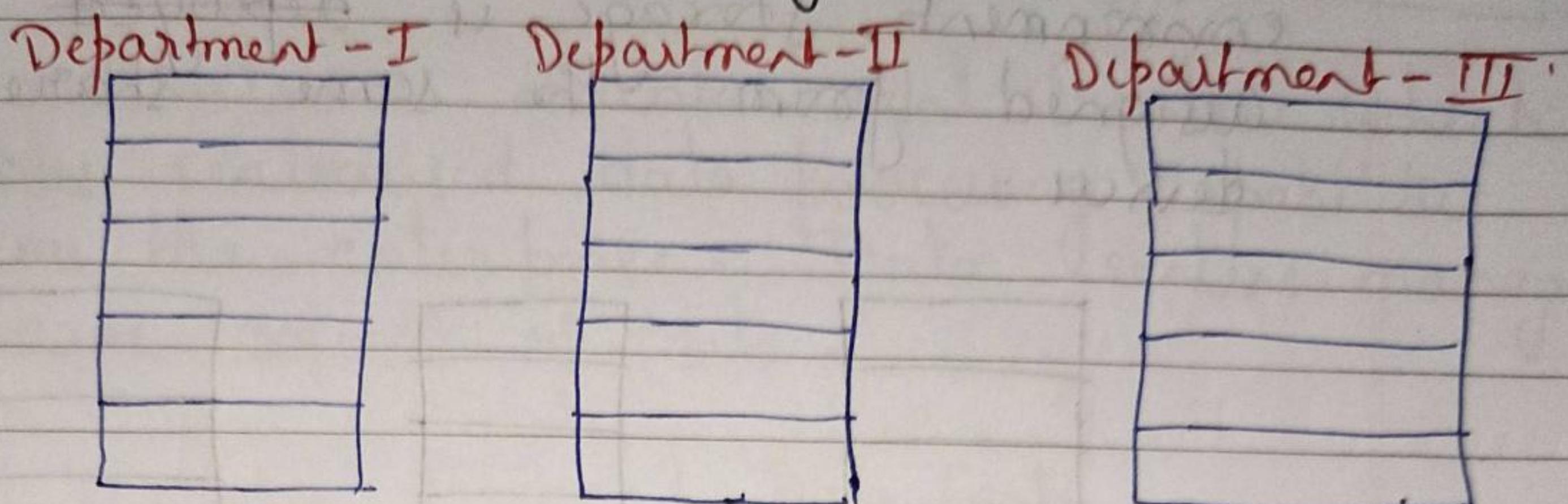
Eg. DVD, CD-ROM, memory in phone, hard disk, etc.

Topic-2 Evolution of Storage Architecture :-

Initially, we used (i) Server-Centric Storage Architecture.

Nowadays, we are using (ii) Information-Centric Storage Architecture.

(ii) 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.

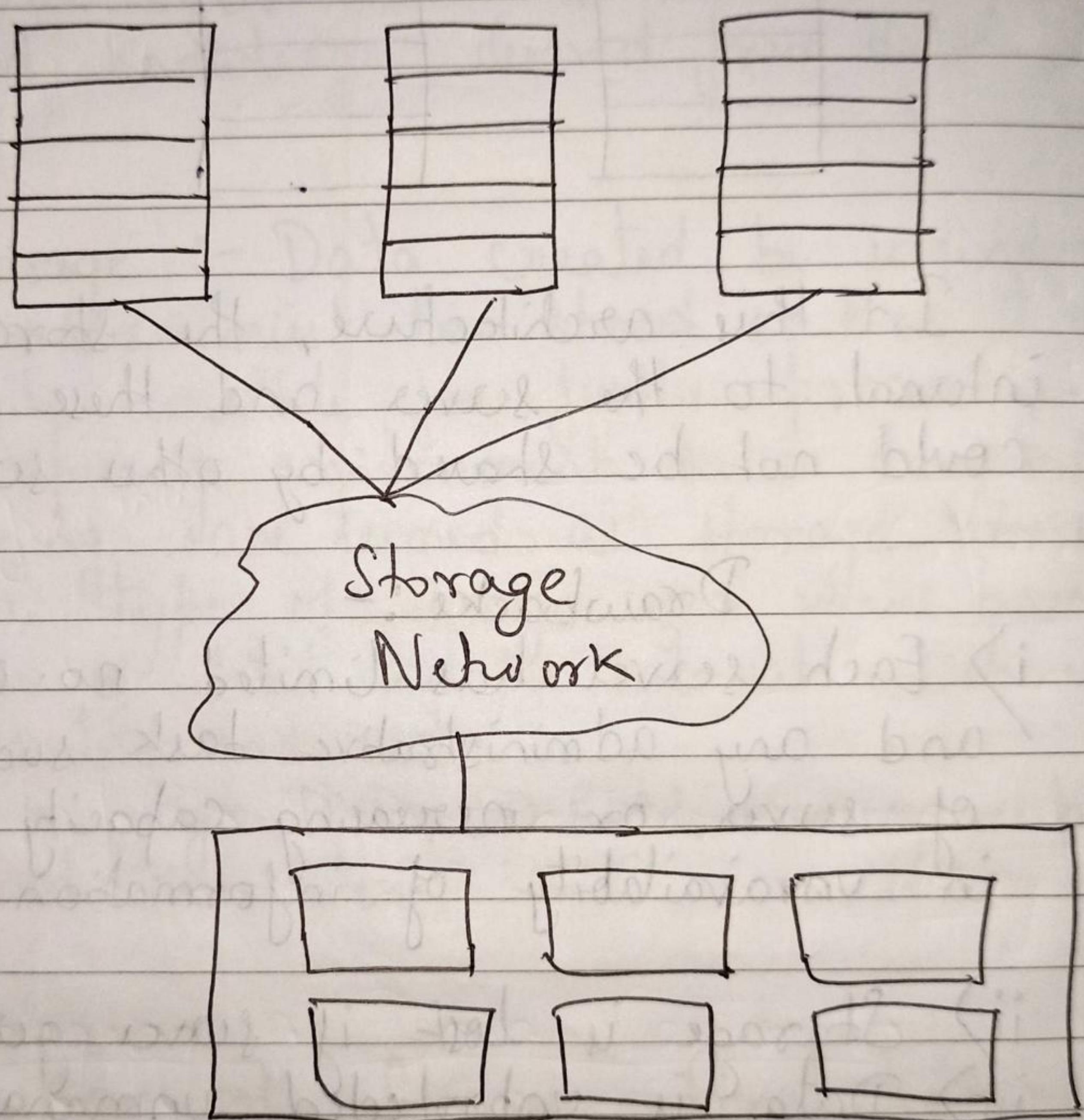
Drawbacks :-

- i) Each server has limited no. of storage devices and any administrative task such as maintenance 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 share same common data storage system. Basically data is managed centrally & independent of server.

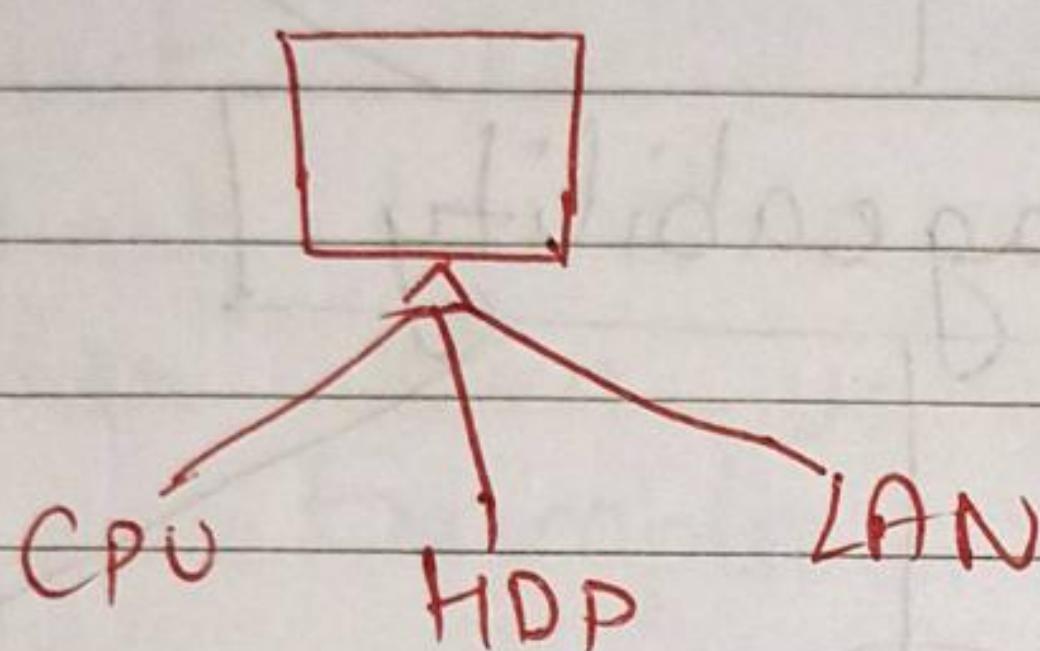
Whenever, new server is deployed in the environment, storage is deployed in the assigned from the same shared storage device.



Adv :- Just reverse of all the disadvantage we studied in Server-Centric.

Topic-3 :- Data Center Infrastructure

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

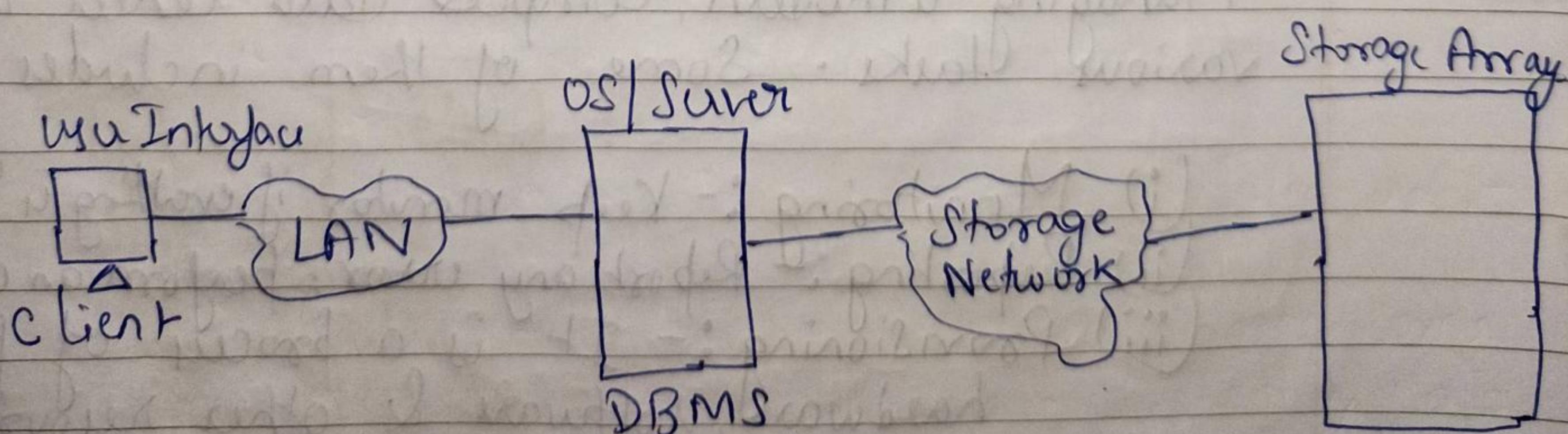


Data Center :-

Server, Storage devices & [H/W Component]
Power backup. [S/W Component]
S/W, App / OS [S/W Component]
Fire supervisor, visitors, A.C, etc [Environmental Condition]

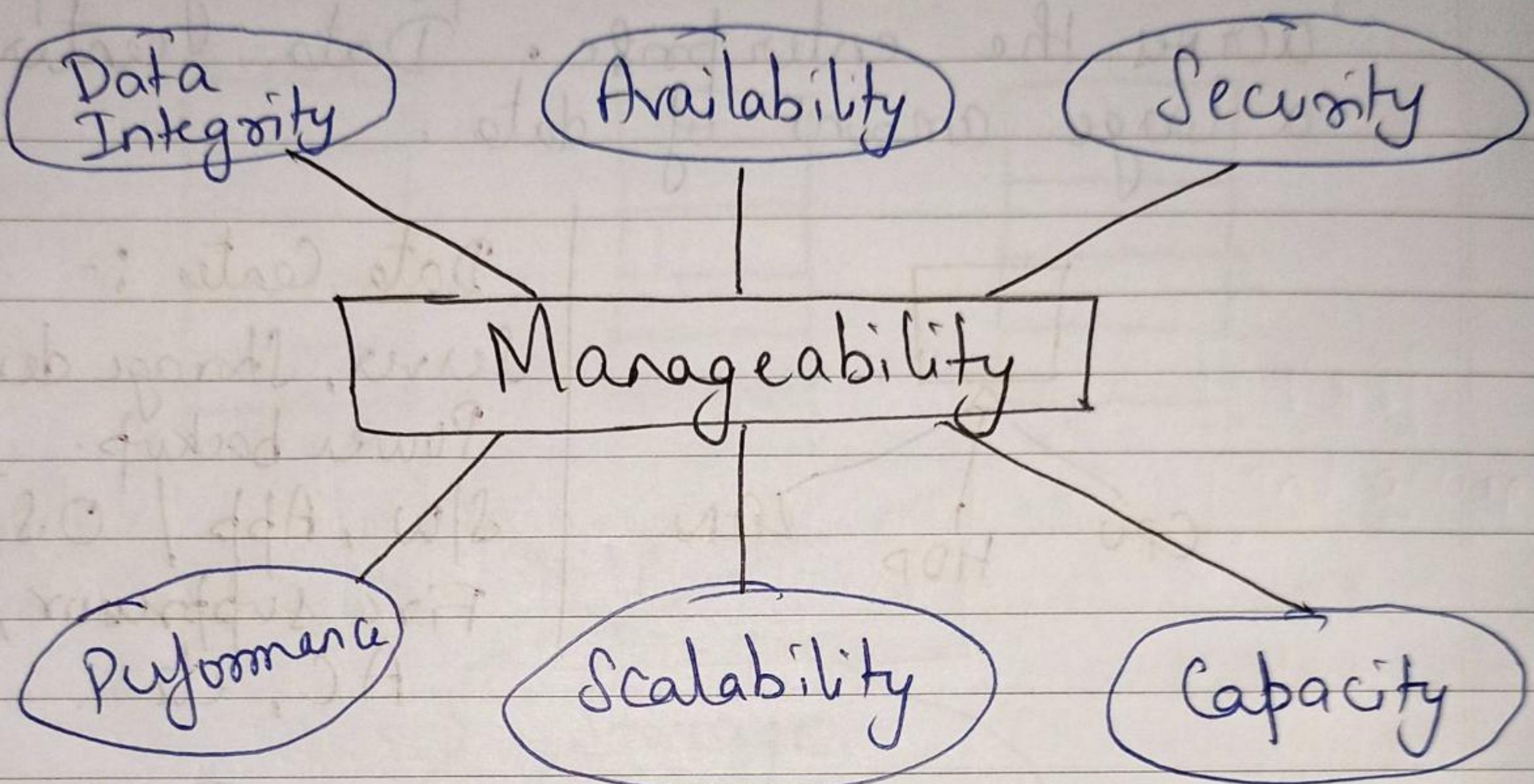
Core-Element of Data Center :-

- (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 service providers includes Google cloud, AWS, IBM cloud, Microsoft Azure, etc.

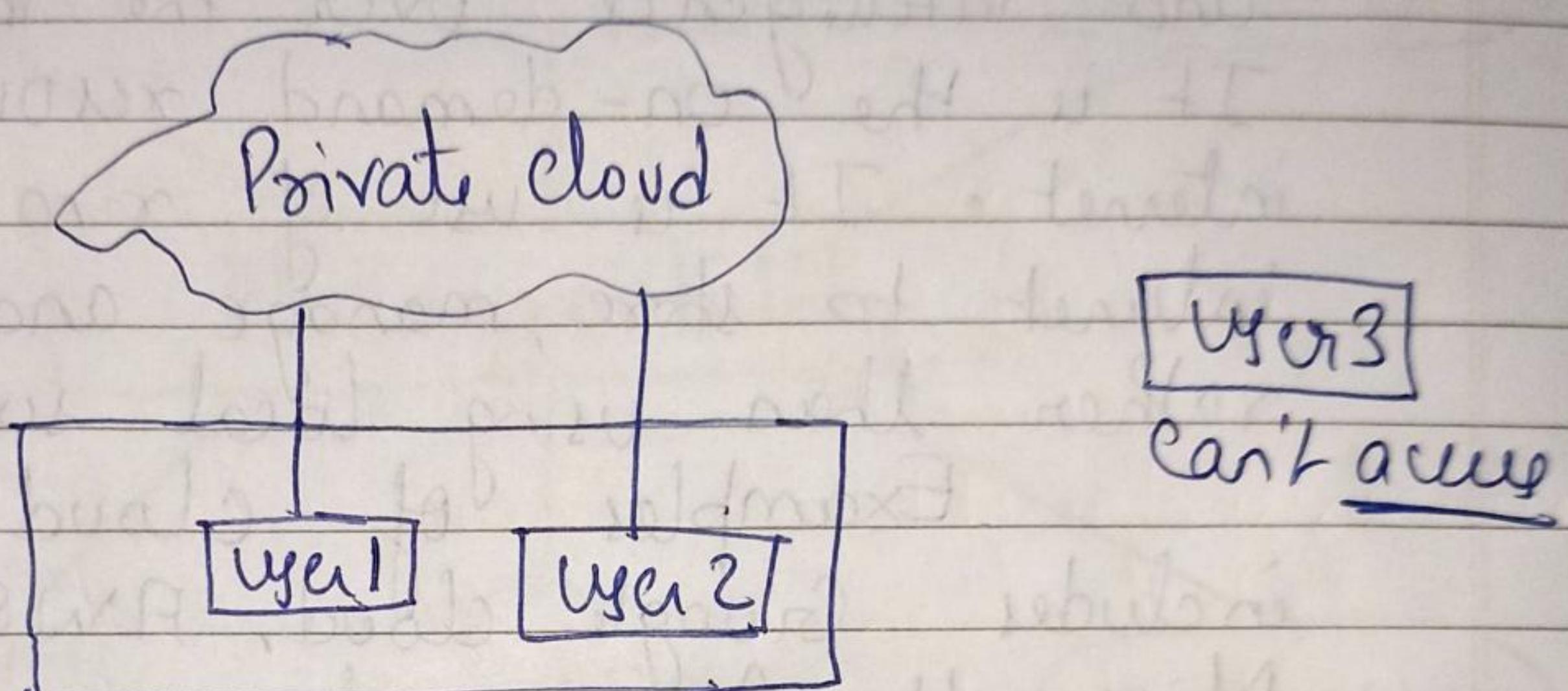
Benefits of cloud Computing :-

- (i) Reduced Investment :- No need to buy private server.
- (ii) Increased scalability :- Rent out or release according to need.
- (iii) Increased availability & Reliability :- fault tolerance

Types of cloud :-

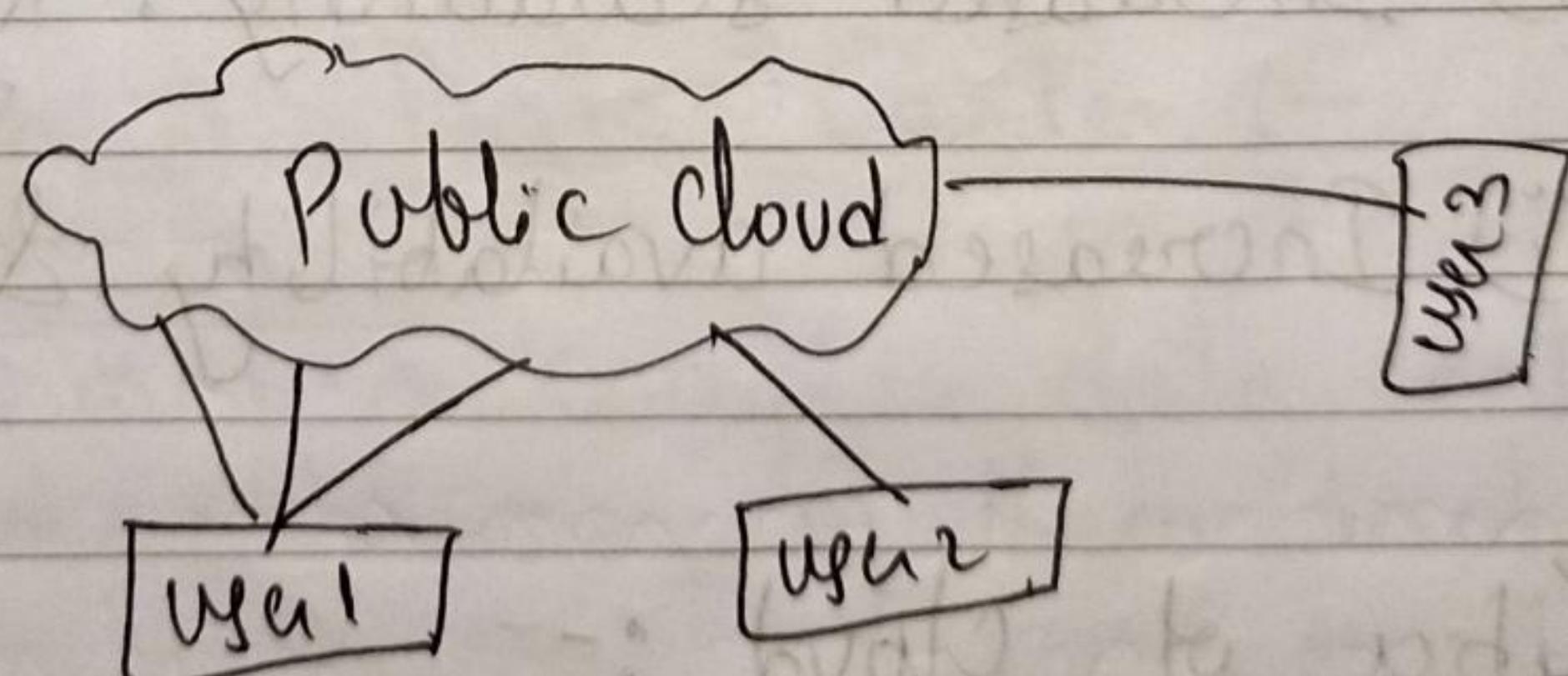
- (i) Private
- (ii) Public
- (iii) Hybrid
- (iv) Community.

(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.



Benefit :- i) Security
ii) More customizable than public.

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

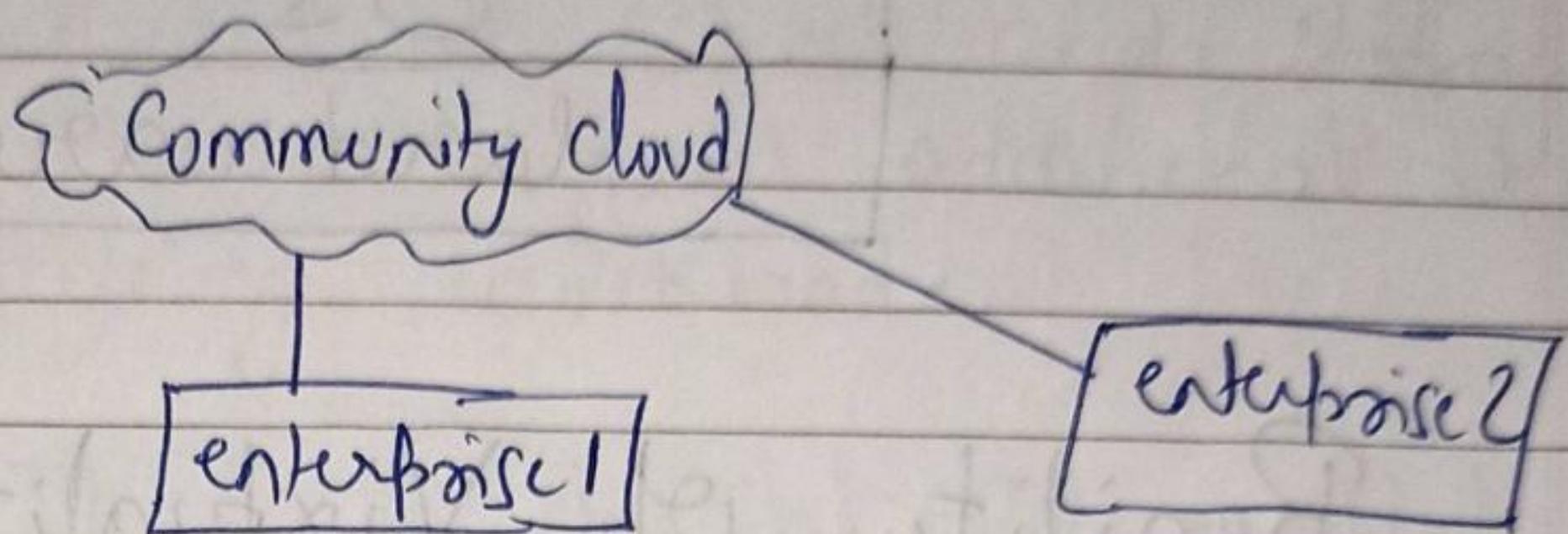


(iii) Hybrid :- Public + Private

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 instance of an application/resource among multiple organization or customers.

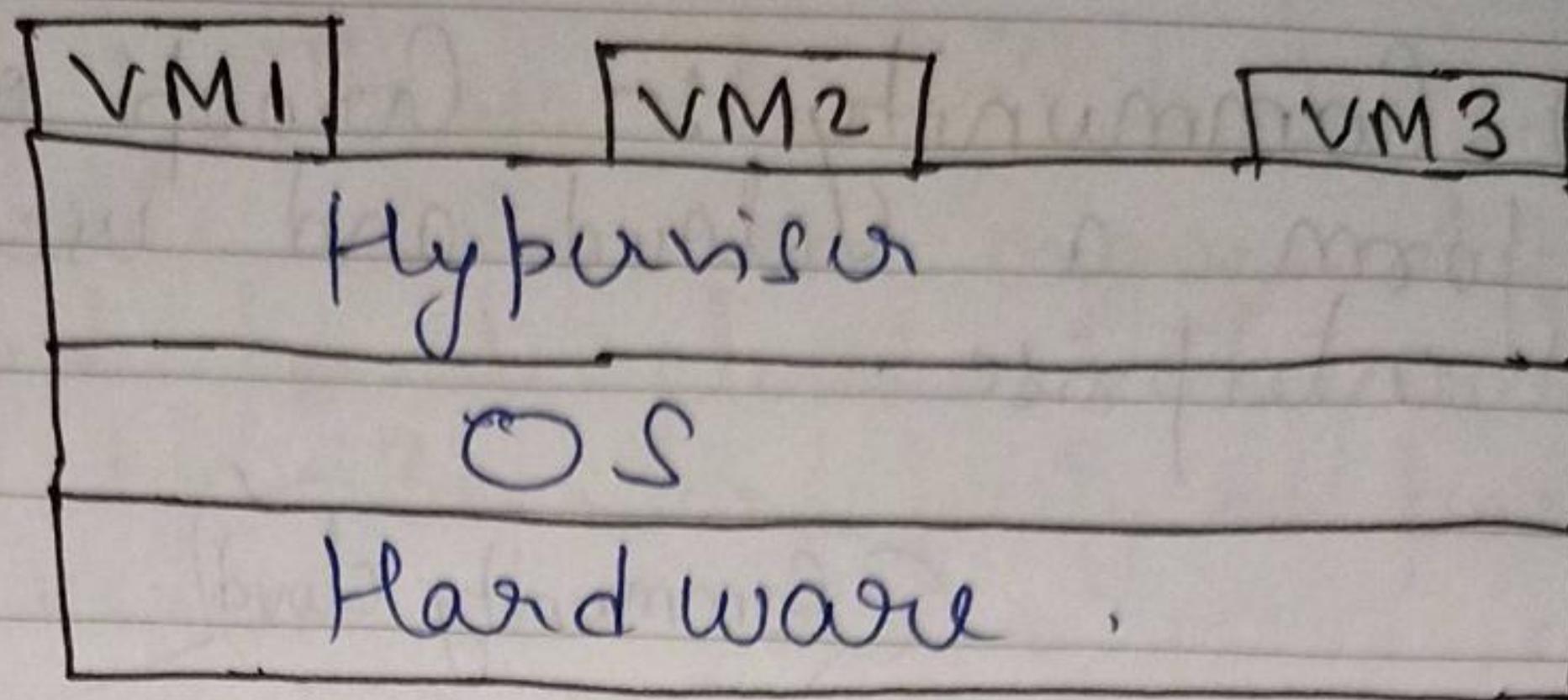
All virtual resources will work independently.

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).



Benefits of Virtualization :-

- (i) Better resource utilization
- (ii) Lower the count of IT infrastructure.
- (iii) Remote access
- (iv) Pay-per-use of IT 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 printers or storage device.

Physical Component of Connectivity :-

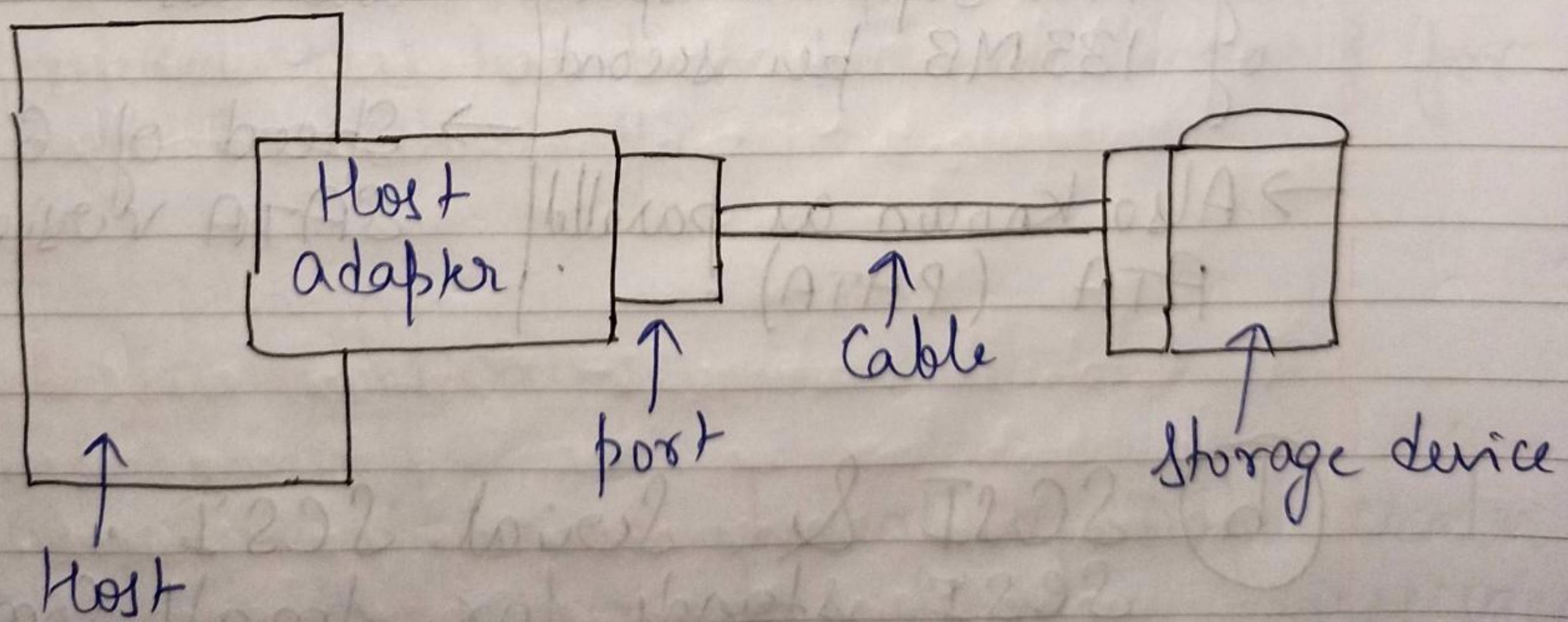
It is the hardware elements that connects the host to storage.

(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 outlet 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 fiber 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:-

① IDE/ATA & Serial ATA

IDE/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 GBS in SATA Version 3.0.

② 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 supports 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.

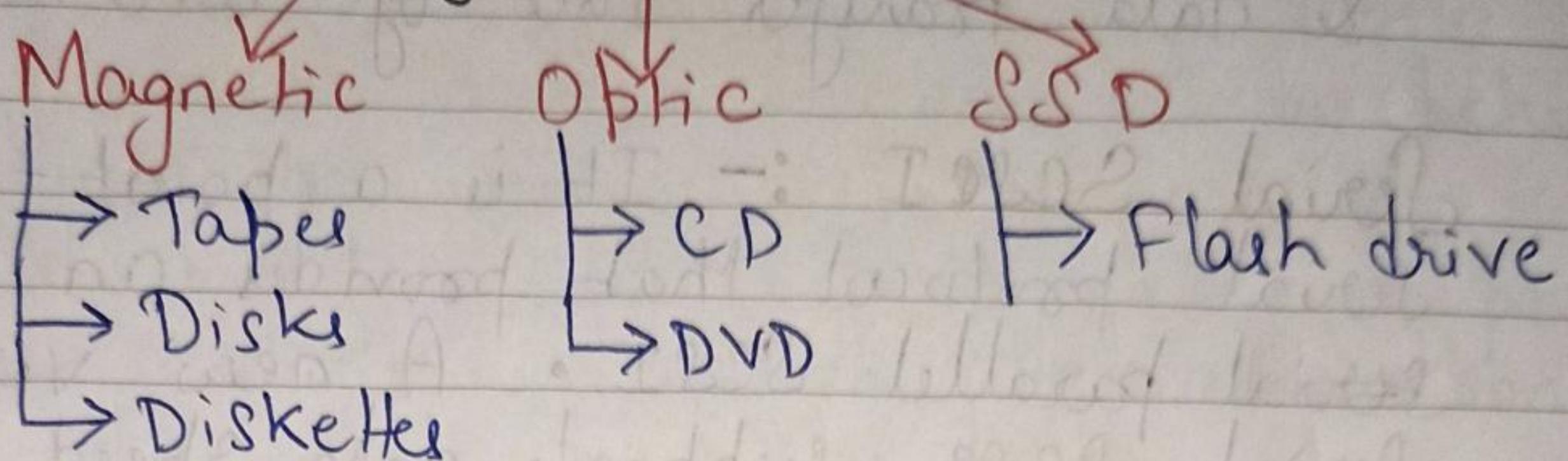
* Latest version of the Fiber channel interface allows transmission upto 16 Gbps.

(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 retrieval.
- (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 ~~Gadgets~~ :- Optical disc storage is popular in small, single user computing environments like games. It works on Write 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-write head, actuator arm assembly & controller board.

(i) Platter :-

- * A typical HDD consists of one or more flat circular disk called platter. Data is recorded on those platters in binary form.
- * Set of rotating platters ~~on these~~ 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 surfaces of the platter.
- * The no. of platters and storage capacity of each platter determine total capacity of drive.

(ii) Spindle :-

* It connects all the platters and is connected to a motor. The motor of the spindle rotates with a constant speed.

- * The disk platters spins at the speed of several thousand ~~per~~ revolution per second 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 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 sits on a special area on the platter near the spindle. This area is called **landing zone**.

* The landing zone is coated with lubricant to avoid friction between the head & platter.

* If the drive malfunctions the read/write head accidentally touches the surface of the platter ~~the outside~~ the landing zone, a head crash occurs.

* 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 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 consists of a microprocessor, internal memory, Circuity and firmware.

The firmware controls 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 controls 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 surfaces of each drive platters. The location of R/W head is referred to as by cylinder number not by the track number.

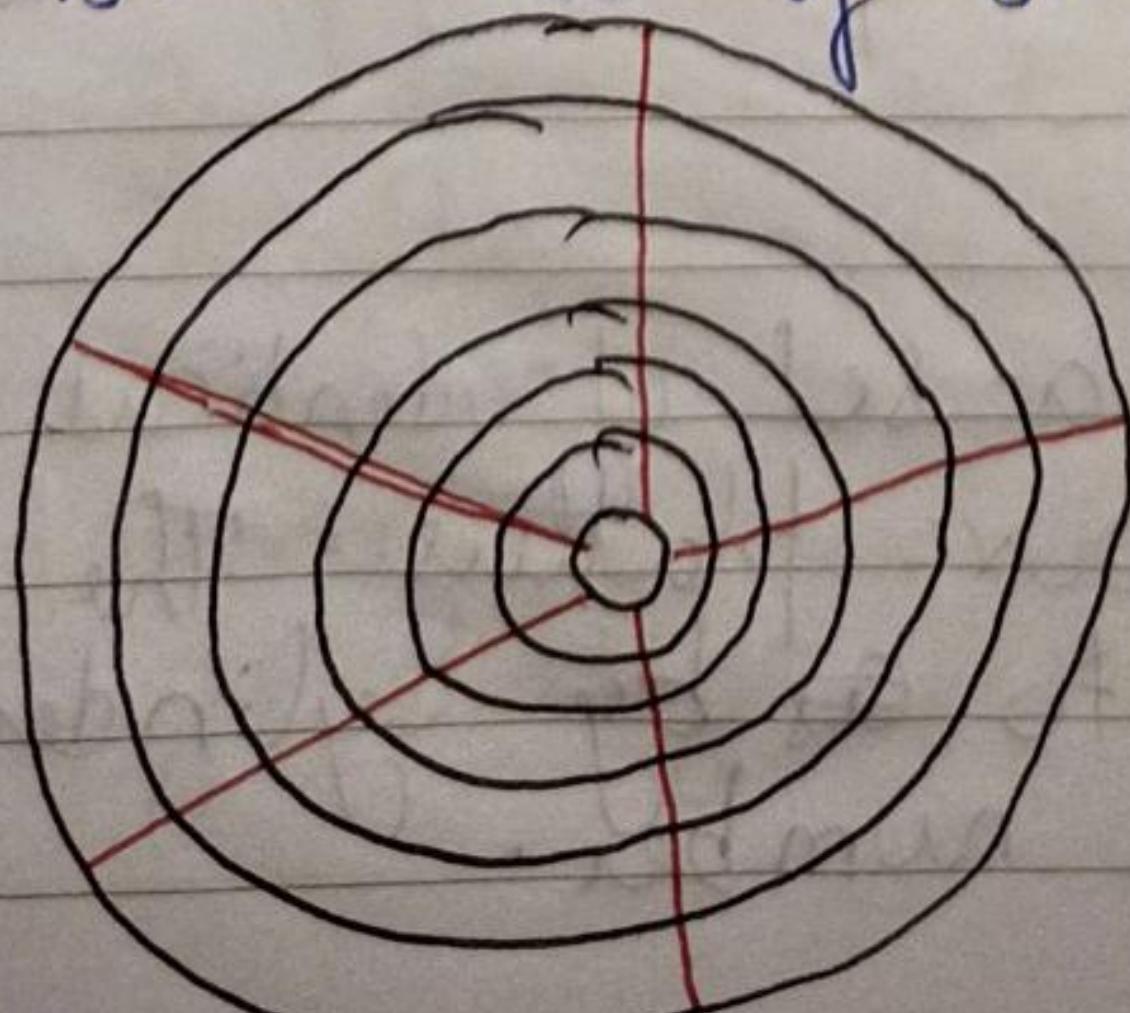
Zoned Bit Recording :-

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

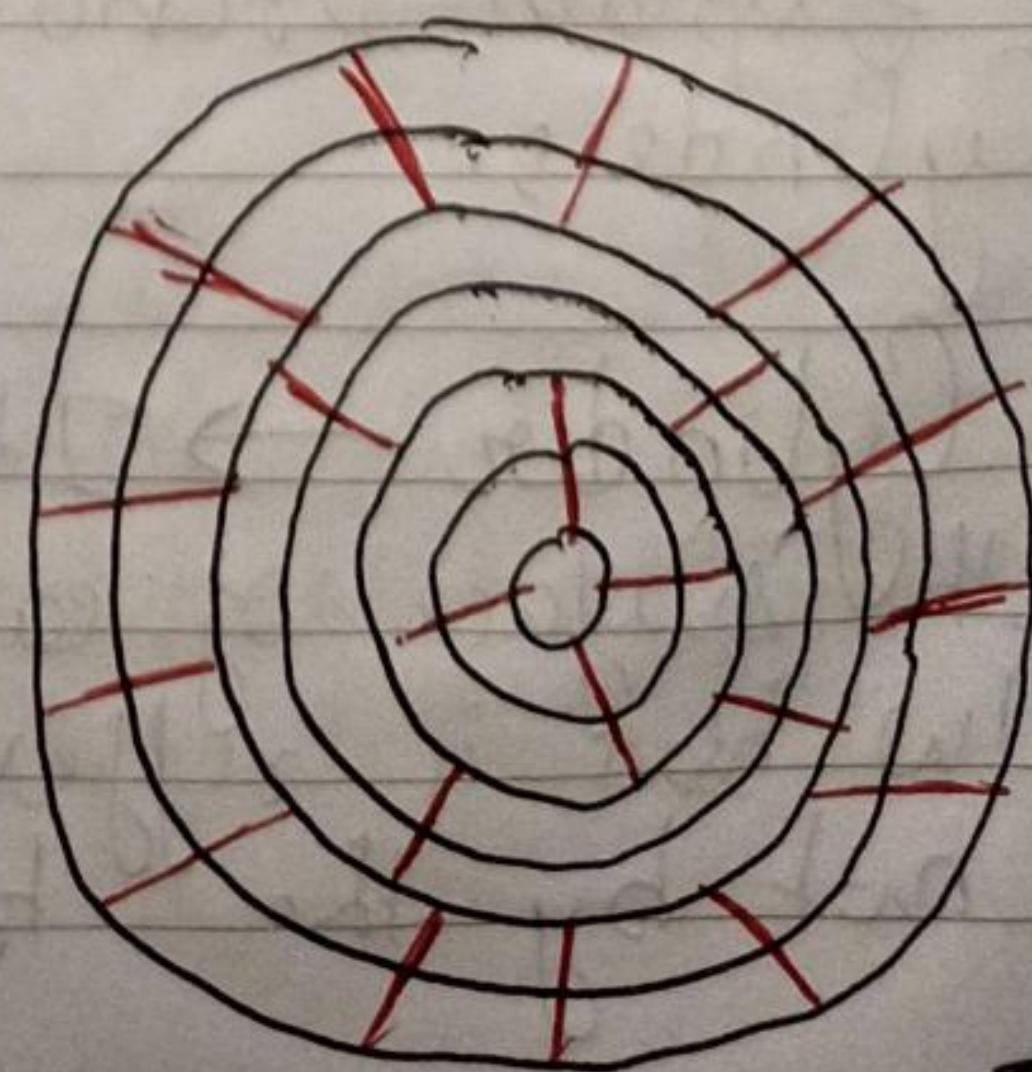
On older disk, the outer track had the same no. of sectors than the inner track so the data density was low on outer tracks. This was an inefficient use of available space.

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

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



Without
Zone



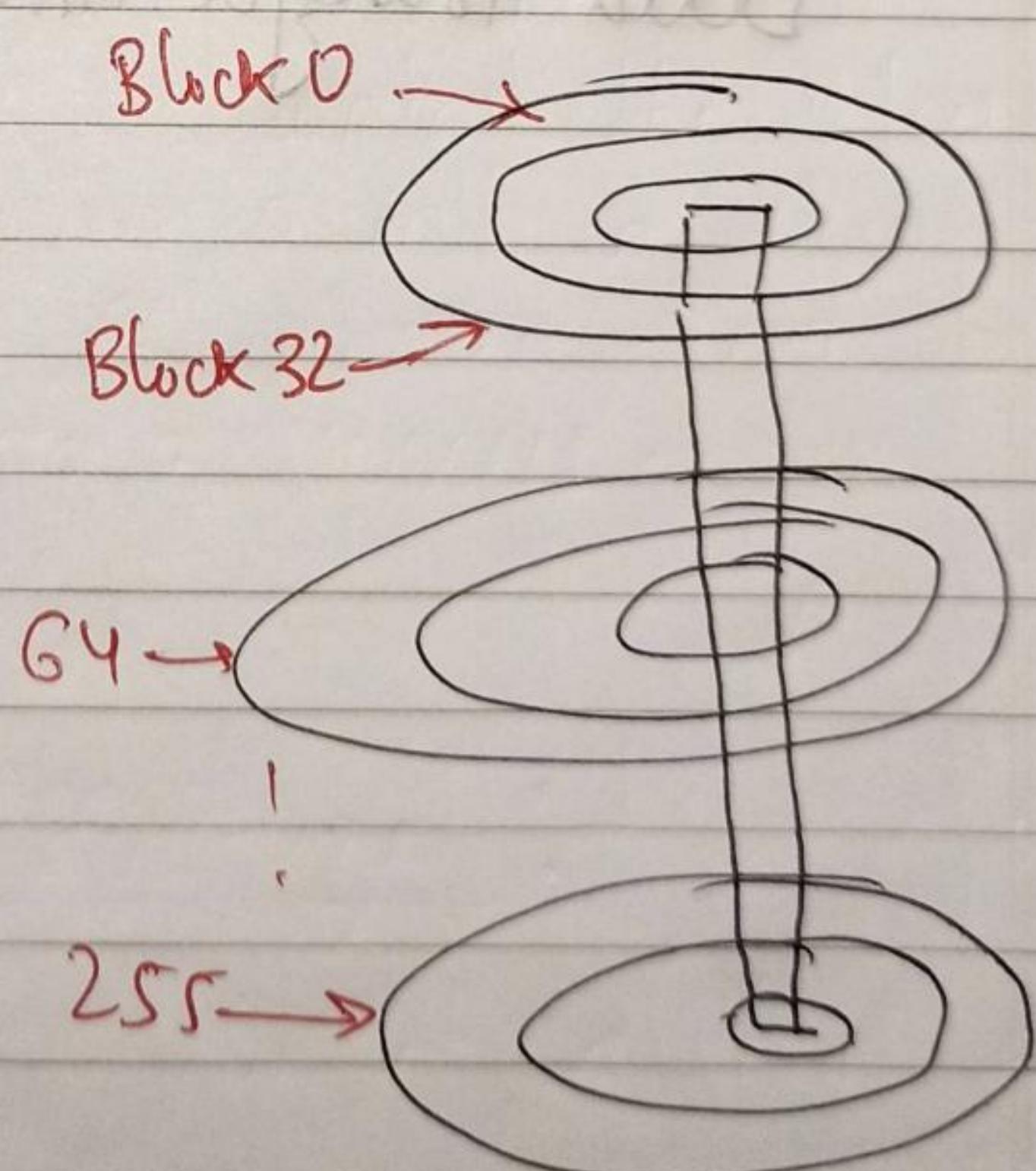
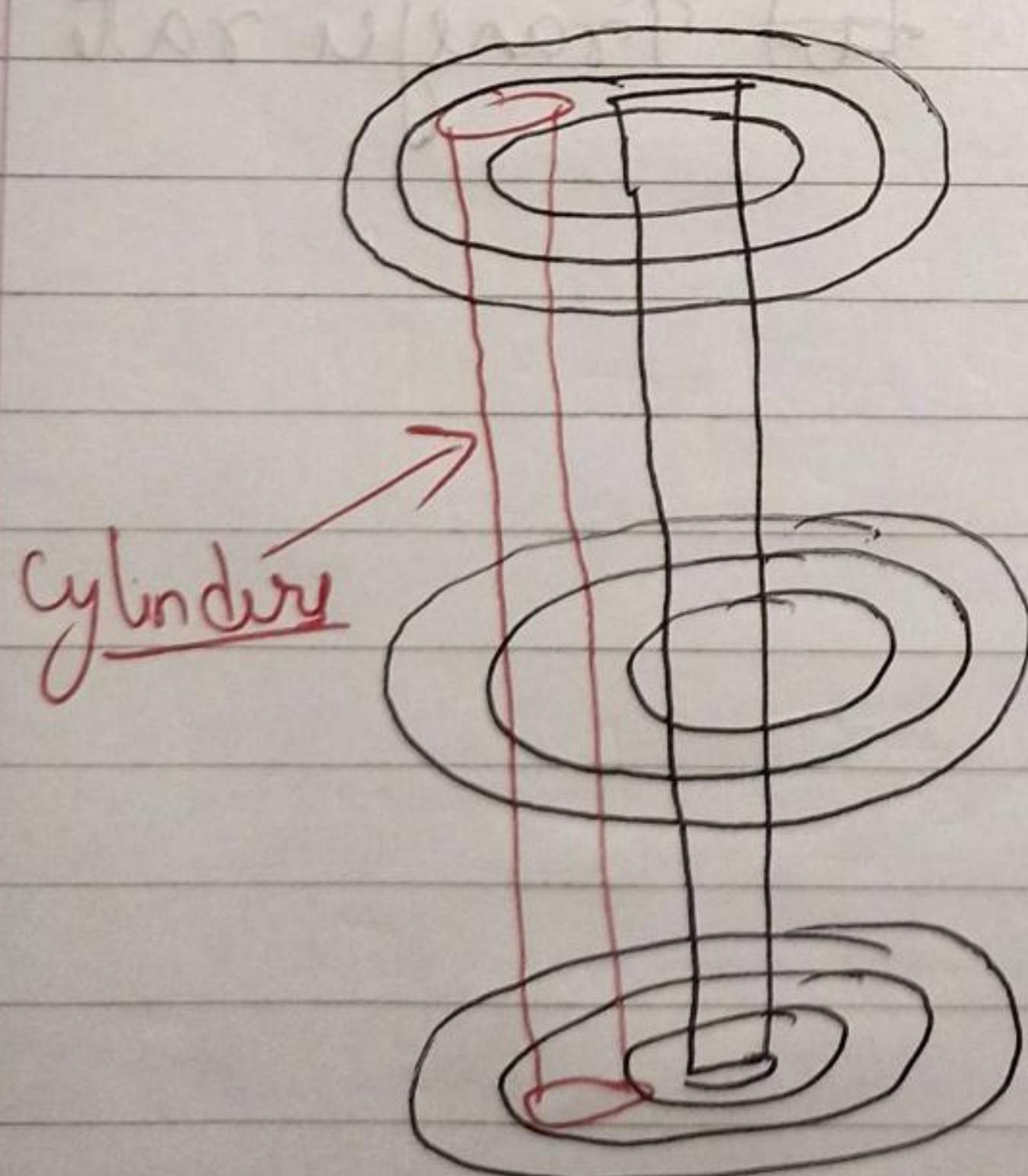
With
Zone

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 terms 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 = $\frac{\text{Data to be transferred}}{\text{Transfer rate}}$