INTRODUCTION TO STORAGE TECHNOLOGY

① <u>Data Proliferation</u> -

It refers to the prodigious amount of data, structured and unstructured, that businesses and governments continue to generate at an unprecedented rate and usability problems that result from attempting to store and manage the data.

It has become a major problem in primary and secondary data storage on computers.

→ <u>Data</u> is a collection of raw facts from which conclusions may be drawn.

→ Factors that so have contributed to the growth of digital data -

(1) Increase in data processing capabilities

(2) Lower cost of digital storage.

(3) Affordable and faster communication technology.

② <u>Data Categorization</u> (Types of Data) -

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

<u>Structured</u> data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. It is typically stored using a DBMS.

<u>Unstructured</u> data elements cannot be stored in rows and columns, and is therefore difficult to query and retrieve by business applications.
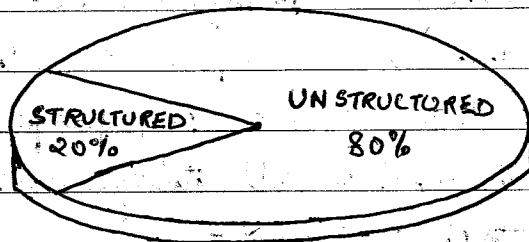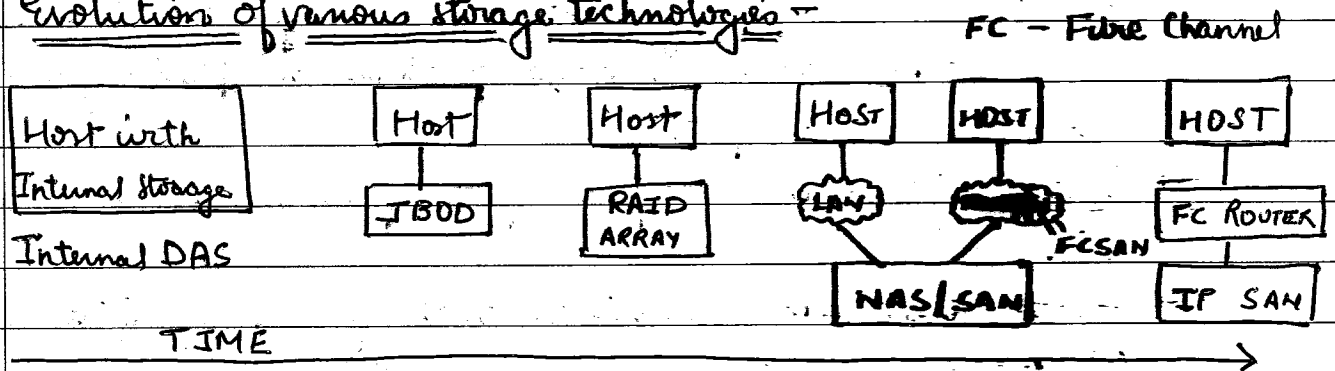
<u>Information</u> is the intelligence and knowledge derived from data.

STRUCTURED 20%   UNSTRUCTURED 80%

<u>Structured data</u> - Rows and Columns.

<u>Unstructured data</u> - Contracts, Forms, Images, Manuals, X-Rays, Checks, E-mail Attachments, PDFs, Instant Messages, Documents, Web Pages, Rich Media, Invoices, Audio, Video.

Storage – In a computing environment, devices designed for storing data are termed storage devices or simply storage.

Eg- CD, ROM's, DVD, Hard Disks etc.

③ **Evolution of various storage technologies** –

FC – Fibre Channel

| Host with Internal Storage | Host | Host | HOST | HOST | HOST |

Internal DAS

JBOD | RAID ARRAY | LAN | FCSAN | FC ROUTER

NAS/SAN | IP SAN

TIME →

DAS (Direct-attached Storage) –

It connects directly to a server (host) or a group of servers in a cluster. Storage can be either internal or external to the server. External DAS alleviated the challenges of limited internal storage capacity.

JBOD (Just a Bunch of disks) –

It is a collection of hard disks that have not been configured to act a ~~redundant array o~~ RAID array. It is accessible either as independent hard disks or as a combined (spanned) single logical volume.

RAID (Redundant Array of Independent Disks) –

It is a data storage virtualization technology that combines multiple physical disk drive components into a single logical unit for the purpose of data redundancy, performance improvement and less cost. It is used in all storage architectures such as DAS, SAN.

NAS (Network attached storage) –

It is a file-level computer data storage server connected to a computer network providing data access to a heterogenous group of clients. It is specialized for serving files either by its hardware,

software, or configuration.

It offers higher scalability, availability, performance and cost benefits compared to general purpose file servers.

SAN (Storage Area Network) -

It is a network that provides access to consolidated, block level data storage. ~~below~~ It has dedicated, high performace Fibre Channel (FC). Storage is partitioned and assigned to a server for accessing its data.

It offers scalability, availability, performance and cost benefits compared to DAS.

IP-SAN (Internet Protocol SAN) -

It is the convergence of technologies used in SAN and NAS. It provides block level communication across a ~~local~~ LAN or WAN, resulting in greater consolidation and availability of data.

④ Overview of Storage Infrastructure Components -

Data centers (Storage Infrastructure) provides Centralized data processing capabilities across the enterprise.

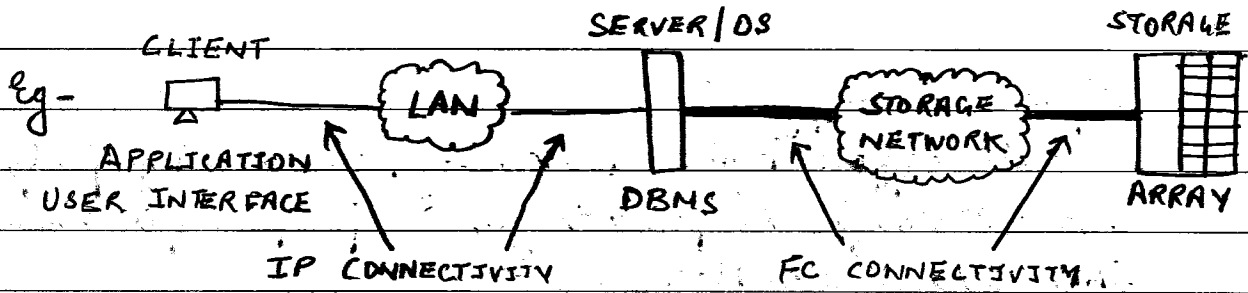→ Five core elements (Storage Infrastructure components) are -

(1) Application - It is computer program that provides the logic for computing operations.

(2) Database - It provides a structured way to store data in logically organized that are interrelated. It optimizes the storage and retrieval of data.
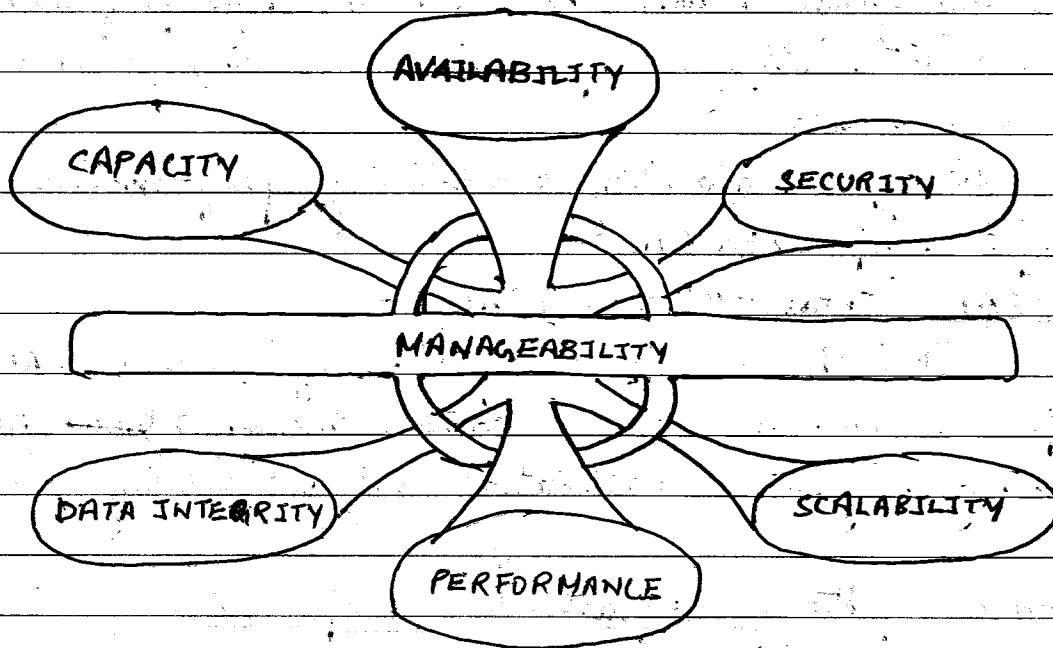
(3) Server and operating system - A computing platform that runs applications and databases.

(4) Network - A data path that facilitates communication between clients and servers or between servers and storage.

(5) Storage array - A device that stores data persistently for subsequent use.

eg -

CLIENT   SERVER/OS   STORAGE

LAN   STORAGE NETWORK

APPLICATION

USER INTERFACE   DBMS   ARRAY

IP CONNECTIVITY   FC CONNECTIVITY.

→ <u>Key characteristics of Data center elements</u> -

AVAILABILITY

CAPACITY   SECURITY

MANAGEABILITY

DATA INTEGRITY   SCALABILITY

PERFORMANCE

→ <u>Key management storage infrastructure activities are</u> -

(1) <u>Monitoring</u> data center includes information, security, performance, accessibility and capacity.

(2) <u>Reporting</u> is done periodically on resource performance, capacity and utilization.

(3) <u>Provisioning</u> is the process of providing hardware, software and other resources needed to run a data center. It includes capacity and resource planning -

→ <u>Key challenges in managing information</u> -

(1) Exploding digital universe.

(2) Increasing dependency on information

(3) Changing value of information

⑤ <u>Information Lifecycle Management (ILM) -</u>

     <u>Information lifecycle</u> is the "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.

     <u>Information Lifecycle Management (ILM)</u> is a proactive strategy that enables an IT organization to effectively manage the data throughout its lifecycle.

→ <u>An ILM strategy</u> include following characteristics —

(1) <u>Business-centric -</u>

     It should be integrated with key processes, applications and initiatives of the business to meet both current and future growth in information.

(2) <u>Centrally Managed -</u>

     All the information assets of a business should be under the scope of the ILM strategy.

(3) <u>Policy-based -</u>

     The implementation of ILM should not be restricted to a few departments. ILM should be implemented as a policy and encompass all business applications, processes and resources.

(4) <u>Heterogeneous -</u>

     An ILM strategy should take into account all types of storage platforms and operating systems.

(5) <u>Optimized -</u>

     Because the value of information varies, an ILM strategy should consider the different storage requirements and allocate storage resources based on the information's value to the business.

Tiered storage is an approach to define different storage levels in order to reduce total storage cost.

→ **ILM Implementation** -

It includes four activities -

(1) Classifying data and applications on the basis of business rules and policies to enable differentiated treatment of information.

(2) Implementing policies by using information management tools, starting from the creation of data and ending with its disposal.

(3) Managing the environment by using integrated tools to reduce operational complexity.

(4) Organizing storage resources in tiers to align the resources with data classes, and storing information in the right type of infrastructure based on the information's current value.

Eg -

TIER 1
TIER 2
TIER 3

NETWORK TIERED STORAGE — APPLICATION SPECIFIC ILM — ENTERPRISE-WIDE ILM
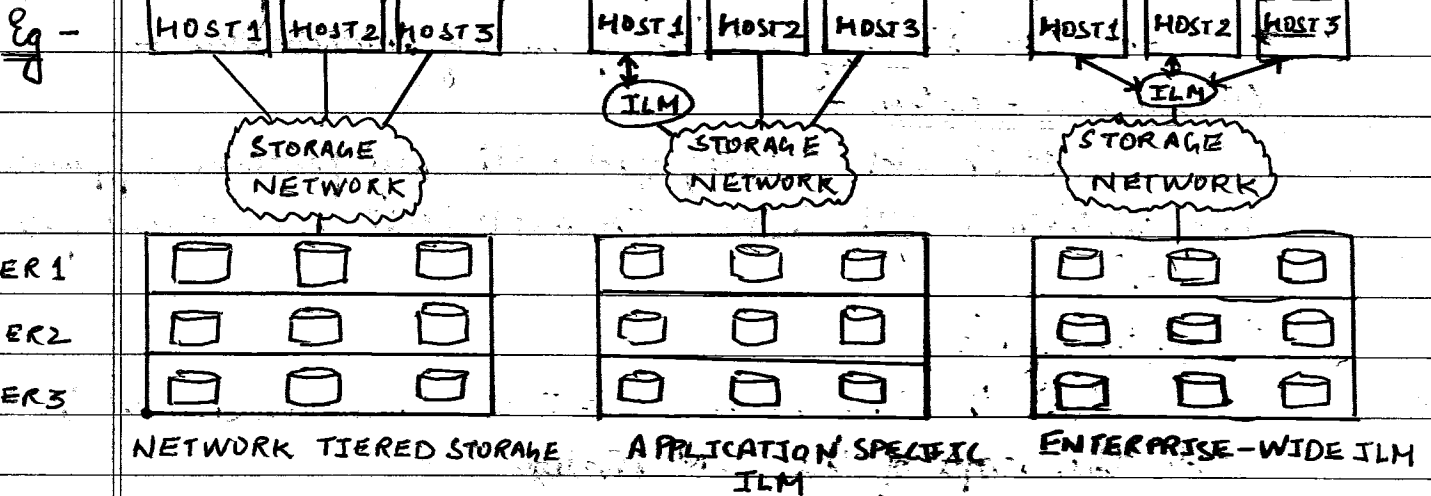
→ **ILM Benefits** -

(1) Improved Utilization by using tiered storage platforms and increased visibility of all enterprise information.

(2) Simplified management by integrating process steps and interfaces with individual tools and by increasing automation.

(3) A wider range of options for backup, and recovery to balance the need for business continuity.

(4) Maintaining compliance by knowing what data needs to be protected for what length of time.

(5) Lower Total Cost of Ownership (TCO) by aligning the infrastructure and management costs with information value.