Module 1

1. **Differentiate between Data and Information? Explain different types of Data?**

Sol:

Difference between Data and information

Data	Information
Data can be any character, number, images, words, text, means little or nothing to a man. When data is processed	Organized and presented in a context to make it useful is called **information**.
Data alone can never be significant	But information is always important by itself.
Data is based on records, observations etc	**Information** is based on analysis of data
Data is unorganized and does not depends on information	But **Information** always depends on data and organized.

In the era of research and the digital world, there are two different types of data are

1. **Structured Data**

   Structured data is highly specific and is stored in a predefined format. Structured data is commonly stored in data warehouses. Structured data can be used by the average business user.

2. **Unstructured Data**

   Unstructured data is a conglomeration of many varied types of data that are stored in their native formats. Unstructured data is stored in data lakes. Unstructured data requires data science expertise to gain accurate business intelligence.

2. Explain the evolution of storage architecture with neat diagram?

Sol:

Server-centric Storage Architecture

Information-centric Storage Architecture
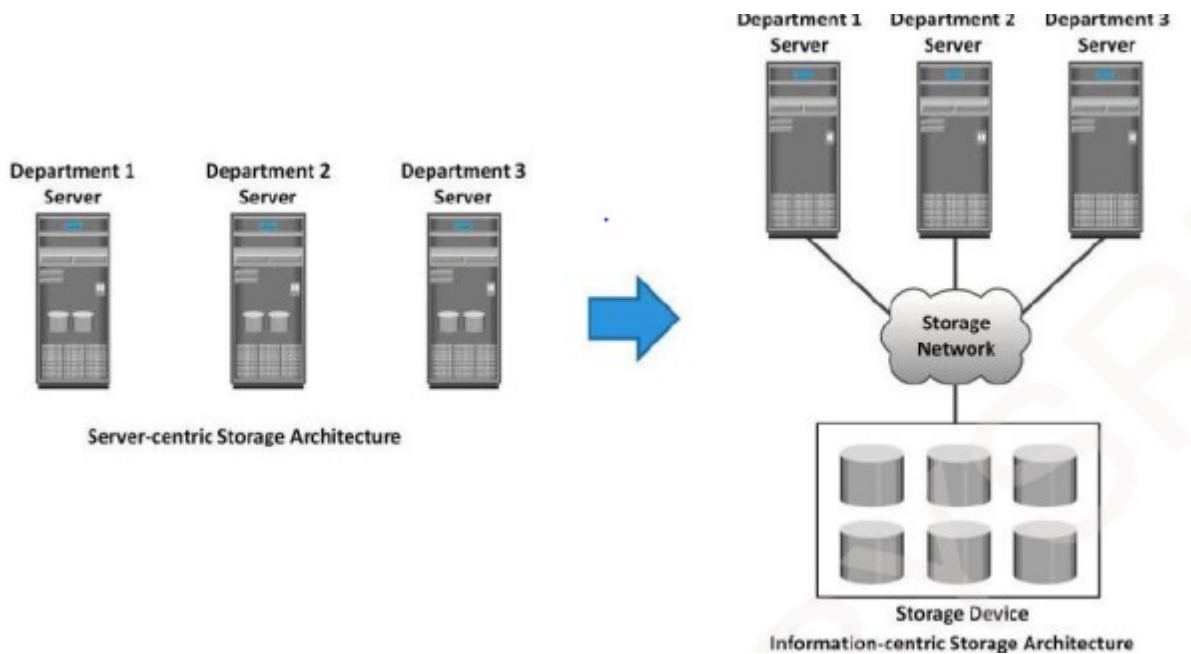
1. Server Centric Storage Architecture

In earlier days, companies had a data-center consisting of

1) Centralized computers (mainframes) and

2) Information storage-devices (such as tape reels and disk packs)

• Each department had their own servers and storage because of following reasons (Figure 1.4):

- ❖ Evolution of open-systems
- ❖ Affordability of open-systems and
- ❖ Easy deployment of open-systems.

• **Disadvantages:**

1) The storage was internal to the server. Hence, the storage cannot be shared with any other servers.

2) Each server had a limited storage-capacity.

3) Any administrative tasks resulted in unavailability of information. The administrative tasks can be maintenance of the server or increasing storage-capacity

---

4) The creation of departmental servers resulted in unprotected, unmanaged, fragmented islands of information and increased capital and operating expenses.

To overcome these challenges, storage evolved from → server-centric architecture to information-centric architecture.

2. Information Centric Architecture

• Storage is managed centrally and independent of servers.

 • Storage is allocated to the servers "on-demand" from a shared-pool.

• A shared-pool refers to a group of disks.

 • The shared-pool is used by multiple servers.

• When a new server is deployed, storage-capacity is assigned from the shared-pool.

• The capacity of shared-pool can be increased dynamically by → adding more disks without interrupting normal-operations.
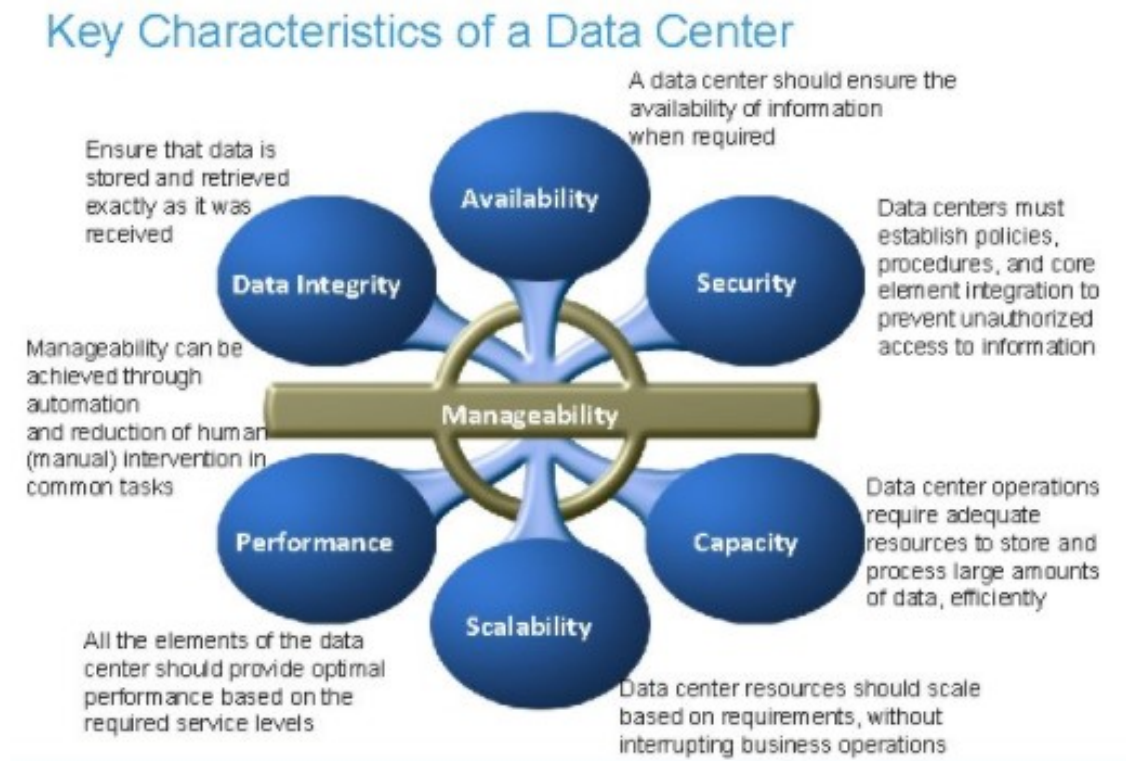
Advantages:

1) Information management is easier and cost-effective.

2) Storage technology even today continues to evolve. This enables companies to consolidate & leverage their data to achieve highest return on info asset

3. **Explain the various key characteristics of Data centres?**

Sol: The various key characteristics of data centers are as follows:

1. Data Integrity
2. Availability
3. Security
4. Capacity
5. Scalability
6. Performance
7. Manageability

Key Characteristics of a Data Center

A data center should ensure the availability of information when required

Ensure that data is stored and retrieved exactly as it was received

Availability

Data Integrity

Security

Data centers must establish policies, procedures, and core element integration to prevent unauthorized access to information

Manageability can be achieved through automation and reduction of human (manual) intervention in common tasks

Manageability

Performance

Capacity

Scalability

Data center operations require adequate resources to store and process large amounts of data, efficiently

All the elements of the data center should provide optimal performance based on the required service levels

Data center resources should scale based on requirements, without interrupting business operations

4. Define Data Structure Infrastructure? Explain different core elements of data centre?

Sol: A data center is a physical facility that organizations use to house their critical applications and data. A data center's design is based on a network of computing and storage resources that enable the delivery of shared applications and data.

Data-center provides centralized data processing capabilities to companies.

Data center design includes routers, switches, firewalls, storage systems, servers, and application delivery controllers. Because these components store and manage business-critical data and applications, data center security is critical in data center design.

There are five core elements which are essentially needed for managing the data centers they are

1. **Applications**

   An application is a program that provides the logic for computing-operations.

   For example: Order-processing-application. Here, an Order-processing-application can be placed on a database. Then, the database can use OS-services to perform R/W-operations on storage

2. **Database**

   DBMS is a structured way to store data in logically organized tables that are interrelated.

   Advantages:

   1) Helps to optimize the storage and retrieval of data.

   2) Controls the creation, maintenance and use of a database.

3. **Server and Operating System**

   A computing-platform that runs 1) applications and 2) databases.

4. **Network**

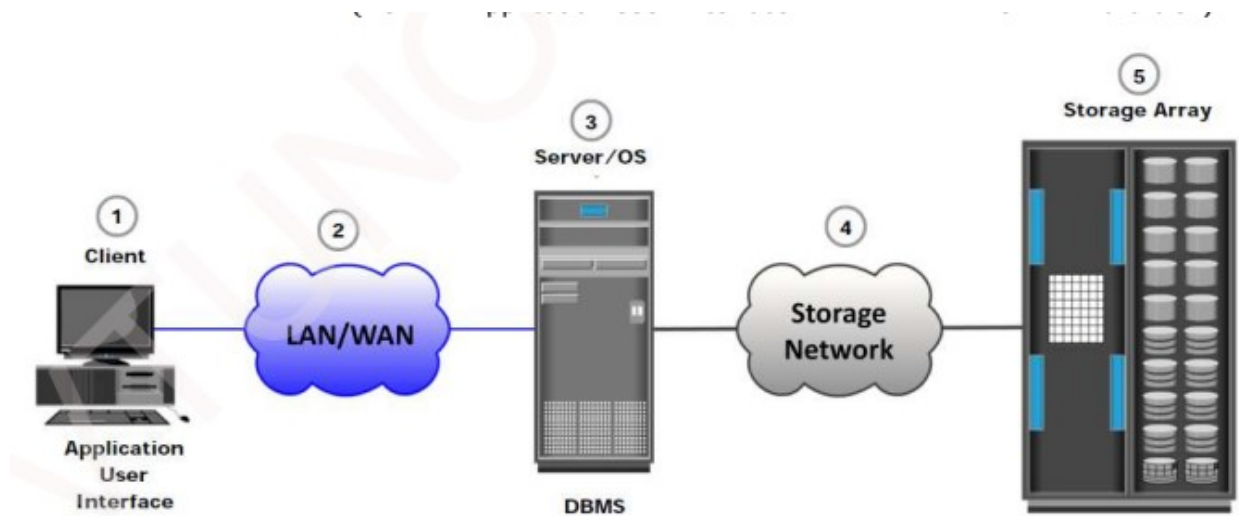A data-path that facilitates communication

1) Between clients and servers or

2) Between servers and storage.

5. **Storage Array**

A device that stores data permanently for future-use

Example: Figure 1.5 represents the order processing application system

Figure 1.5: Order processing application system

Step 1: A customer places an order through the AUI on the client-computer.

Step 2: The client accesses the DBMS located on the server to provide order-related information. (Order-related information includes customer-name, address, payment-method & product-ordered).

**Step 3:** The DBMS uses the server to write this data to the disks in the storage.

Step 4: The storage-network → provides the communication-link between server and storage and → transports the write-command from server to storage

Step 5: After receiving the write command the storage saves the data on disk

5. **What is file system? Explain the process of mapping user files to the disk storage?**

Sol: A file is a collection of related-records stored as a unit with a name. (say employee.lst)

- A file-system is a structured way of storing and organizing data in the form of files.
- File-systems enable easy access to data-files residing within → disk-drive → disk-partition or → logical-volume.
- A file-system needs host-based software-routines (API) that control access to files.
- It provides users with the functionality to create, modify, delete and access files.
- A file-system organizes data in a structured hierarchical manner via the use of directories (i.e. folder)
- A **directory** refers to a container used for storing pointers to multiple files.
- All file-systems maintain a pointer-map to the directories and files.
- Some common file-systems are: → FAT 32 (File Allocation Table) for Microsoft Windows → NT File-system (NTFS) for Microsoft Windows → UNIX File-system (UFS) for UNIX → Extended File-system (EXT2/3) for Linux
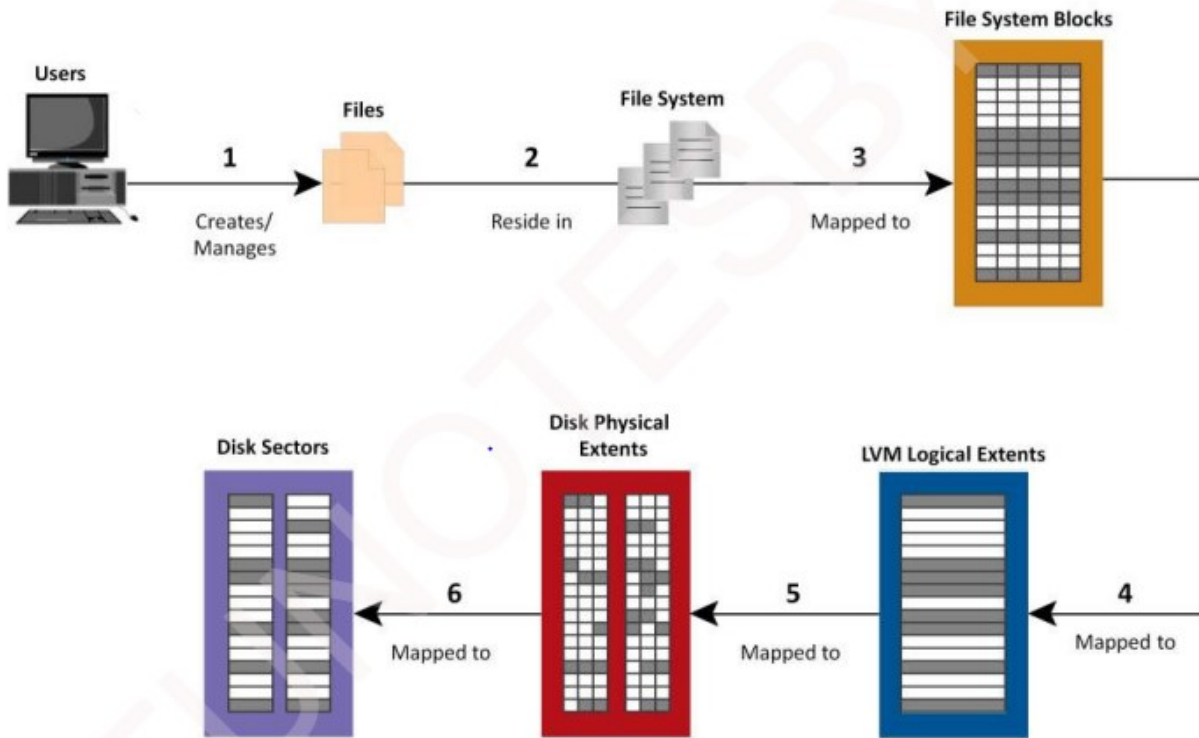
Figure: Process of mapping user file to disk storage

The above figure shows process of mapping user-files to the disk-storage with an LVM:

1) Files are created and managed by users and applications.
2) These files reside in the file-system.
3) The file-system are mapped to file-system blocks.
4) The file-system blocks are mapped to logical-extents.
5) The logical-extents are mapped to disk physical-extents by OS or LVM.
6) Finally these physical extents are mapped to the disk storage

6. Define Host? List and explain the various logical components of host?

Sol: A host is **any hardware device that has the capability of permitting access to a network via a user interface, specialized software, network address, protocol stack, or any other means**.

Examples of host devices include **computers, personal devices, thin clients, and multi-function devices**.

The Various logical components of Hosts are :

1. **Operating System (OS)**
   - An OS is a program that acts as an intermediary between → application and → hardware-components.
   - The OS controls all aspects of the computing-environment.
   - Data-access is one of the main service provided by OS to the application
   - Tasks of OS:
     - Monitor and respond to user actions and the environment.
     - Organize and control hardware-components.
     - Manage the allocation of hardware-resource (simply the resource).
     - Provide security for the access and usage of all managed resources.
     - Perform storage-management tasks. 6) Manage components such as file-system, LVM & device drivers.

2. **Memory Virtualization**

   Memory-virtualization is used to virtualize the physical-memory (RAM) of a host. It creates a VM with an address-space larger than the physical-memory space present in computer. The virtual-memory consists of:

address-space of the physical-memory and → part of address-space of the disk-storage.

- ❖ The entity that manages the virtual-memory is known as the virtual-memory manager (VMM).
- ❖ The VMM → manages the virtual-to-physical-memory mapping and → fetches data from the disk-storage
- ❖ The space used by the VMM on the disk is known as a swap-space.
- ❖ A swap-space is a portion of the disk that appears like physical-memory to the OS.
- ❖ The memory is divided into contiguous blocks of fixed-size pages. (VM --> virtual-memory) Paging ¬ A paging → moves inactive-pages onto the swap-file and → brings inactive-pages back to the physical-memory when required.
- ❖ **Advantages:**
  - o Enables efficient use of the available physical-memory among different applications.
  - o Normally, the OS moves the least used pages into the swap-file.
  - o Thus, sufficient RAM is provided for processes that are more active.
- ❖ **Disadvantages**
  - o Access to swap-file pages is slower than physical-memory pages. This is because → swap-file pages are allocated on the disk which is slower than physical-memory.

3. Device Driver

It is a special software that permits the OS & hardware-component to interact with each other. The hardware-component includes printer, a mouse and a hard-drive.

A device-driver enables the OS to → recognize the device and → use a standard interface to access and control devices.Device-drivers are hardware-dependent and OS-specific.

7. Define Connectivity? Explain different types of Components used for Connectivity process between the host and peripheral devices?

Sol: Connectivity refers to interconnection between host and peripheral-devices such as storage-devices

• **Components of connectivity are classified as:**

1. Physical-Components

Physisal-components refer to hardware-components used for connection between host & storage. Three components of connectivity are:

1) Host interface device

Host Interface Device is used to connect a host to other hosts and storage-devices.

**Example:** HBA (host bus adapter) NIC (network interface card).

- HBA is an ASIC board that performs I/O-operations between host and storage.
- **Advantage:**
  o HBA relieves the CPU from additional I/O-processing workload.
  o A host typically contains multiple HBAs. (ASIC --> application-specific integrated circuit).

2) Port

Port refers to a physical connecting-point to which a device can be attached. An HBA may contain one or more ports to connect the host to the storage-device.

3) Cable

Cable is used to connect hosts to internal/external devices using copper-wire or optical-fiber.
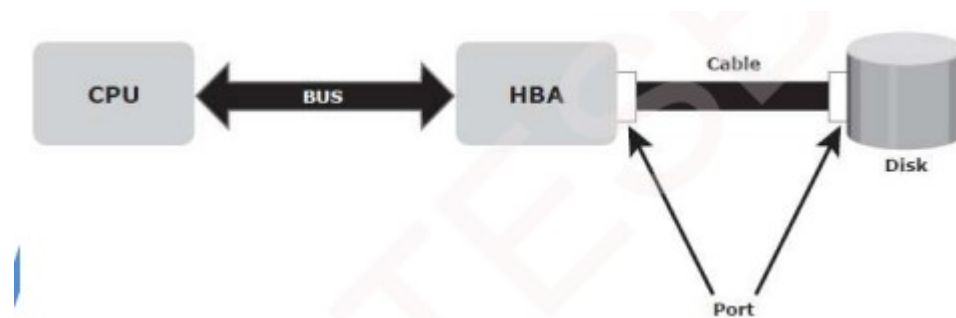
Figure: Physical Components of Connectivity

2. Interface Protocols

Interface-Protocol enables communication between host and storage. • Protocols are implemented using interface-devices (or controllers) at both source and destination. The popular protocols are:

1. IDE/ATA (Integrated Device Electronics/Advanced Technology Attachment)

It is a standard interface for connecting storage-devices inside PCs (Personal Computers). The storage-devices can be disk-drives or CD-ROM drives. It supports parallel-transmission. Therefore, it is also known as Parallel ATA (PATA).

2. SCSI (Small Computer System Interface)

It has emerged as a preferred protocol in high-end computers. • Compared to ATA, SCSI → supports parallel-transmission and → provides improved performance, scalability, and compatibility.

3. FC (Fibre Channel) and

It is a widely used protocol for high-speed communication to the storage-device.

4. IP (Internet Protocol).

It is a protocol used for communicating data across a packet-switched network. It has been traditionally used for host-to-host traffic. '.' of new technologies, IP network has become a feasible solution for host-to-storage communication.

Module 2

8. Define RAID? Explain the different types of RAID implementation?

9. List and Explain different techniques of RAID?

10.   Describe RAID levels with reference to RAID 0, RAID 1 and nested RAID, with neat diagram?

11.   With neat diagram explain the structure of read and write operation with cache?

12.   With a neat diagram explain different RAID techniques?

13.   Explain two different types of intelligent storage system?

14.   Define Cache? Explain read and write operation of Cache?

15.   Define SAN? List and explain different components of SAN?

16.   Define Fibre Channel? Explain three basic interconnectivity of FC?

*******************************GOOD LUCK********************************