Internal Assessment Test 1 – September2019											
Sub:	Storage Area Networks					Sub Code:	15CS754	Branch:	CSE		
Date:	21/09/19	Duration:	90 mins	Max Marks:	50	Sem / Sec:	VII A/B/C			OBE	
	Answer any FIVE FULL Questions								MARK S	CO	RB T
1 (a)	**Explain key characteristics of data center elements with diagram?**								[05]	CO1	L1

Answer:

The following are the key characteristics of Data Center

1. **Availability**: A data center should ensure the availability of information when required. Unavailability of information could cost millions of dollars per hour to businesses, such as financial services, telecommunications, and e-commerce.
2. **Security**: Data centers must establish policies, procedures, and core element integration to prevent unauthorized access to information.
3. **Scalability**: Business growth often requires deploying more servers, new applications, and additional databases. Data center resources should scale based on requirements, without interrupting business operations.
4. **Performance**: All the elements of the data center should provide optimal performance based on the required service levels.
5. **Data integrity**: Data integrity refers to mechanisms, such as error correction codes or parity bits, which ensures that data is stored and retrieved exactly as it was received.
6. **Capacity**: Data center operations require adequate resources to store and process large amounts of data, efficiently. When capacity requirements increase, the data center must provide additional capacity without interrupting availability or with minimal disruption. Capacity may be managed by reallocating the existing resources or by adding new resources.
7. **Manageability**: A data center should provide easy and integrated management of all its elements. Manageability can be achieved through automation and reduction of human (manual) intervention in common tasks.

The figure below shows the key characteristics of a Data Center

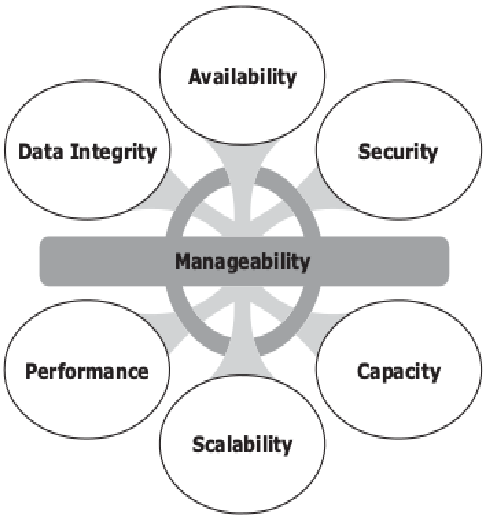


Figure 1-6: Key characteristics of a data center

(b)	**Explain the evolution of Storage Architecture with diagram.**	05]	CO1	L4

Answer:

There are two storage architectures

1. Server-Centric storage architecture
2. Information-centric storage architecture

- The diagrams below show the two types of storage architectures.

1. Server-Centric storage architecture

- Storage is internal to the server.
- These storage devices could not be shared with any other servers.
- limited number of storage devices.
- Any administrative tasks, such as maintenance of the server or increasing storage capacity, might result in unavailability of information.
- An increase in the number of servers resulted in unprotected, unmanaged, fragmented islands of information and increased capital and operating expenses.

2. Information-centric storage architecture

- Storage devices are managed centrally and independent of servers.
- These centrally-managed storage devices are shared with multiple servers.
- The capacity of shared storage can be increased dynamically by adding more storage devices.
- Increasing storage capacity is possible without impacting information availability.
- information management is easier and cost-effective.

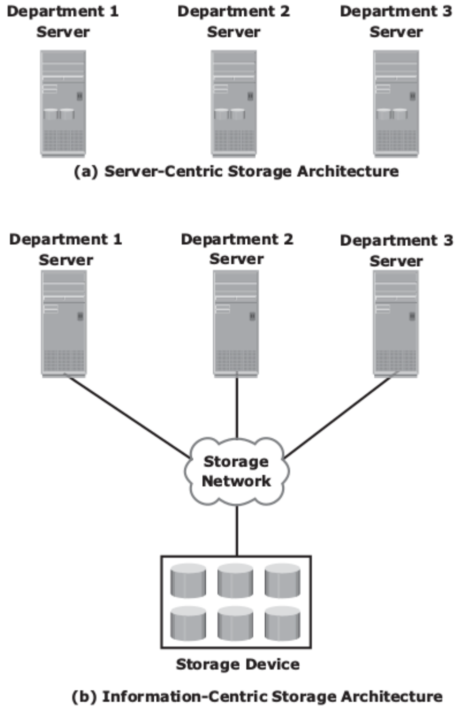


Figure 1-4: Evolution of storage architecture

2 (a)	**What is file system? Explain the process of mapping user files to the disk storage.**	[08]	CO1	L2

Answer:

A file is a collection of related records or data stored as a unit with a name.

- A file system is a hierarchical structure of files. A file system enables easy access to data files residing within a disk drive, a disk partition, or a logical volume.
- A file system consists of logical structures and software routines that control access to files. It provides users with the functionality to create, modify, delete, and access files.
- Access to files on the disks is controlled by the permissions assigned to the file by the owner, which are also maintained by the file system.

A file system organizes data in a structured hierarchical manner via the use of directories, which are containers for storing pointers to multiple files. All file systems maintain a pointer map to the directories, subdirectories, and files that are part of the file system.

- Examples of 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

Metadata:

- Apart from the files and directories, the file system also includes a number of other related records, which are collectively called the **metadata**.
- For example, the metadata in a UNIX environment consists of the superblock, inodes, and the list of data blocks free and in use.

Superblock:

- A superblock contains important information about the file system, such as the file system type, creation and modification dates, size, and layout.
- It also contains the count of available resources (such as the number of free blocks, inodes, and so on) and a flag indicating the mount status of the file system.

Inode:

- An inode is associated with every file and directory and contains information such as the file length, ownership, access privileges, time of last access/modification, number of links, and the address of the data.

File System Block:

- A file system block is the smallest "unit" allocated for storing data.
- Each file system block is a contiguous area on the physical disk.

- The block size of a file system is fixed at the time of its creation.
- The file system size depends on the block size and the total number of file system blocks.
- A file can span multiple file system blocks because most files are larger than the predefined block size of the file system.
- File system blocks cease to be contiguous and become fragmented when new blocks are added or deleted. Over time, as files grow larger, the file system becomes increasingly fragmented.

- The following list shows the process of mapping user files to the disk storage subsystem with an LVM:

1. Files are created and managed by users and applications.
2. These files reside in the file systems.
3. The file systems are mapped to file system blocks.
4. The file system blocks are mapped to logical extents of a logical volume.
5. These logical extents in turn are mapped to the disk physical extents either by the operating system or by the LVM.
6. These physical extents are mapped to the disk sectors in a storage subsystem.

- If there is no LVM, then there are no logical extents. Without LVM, file system blocks are directly mapped to disk sectors.

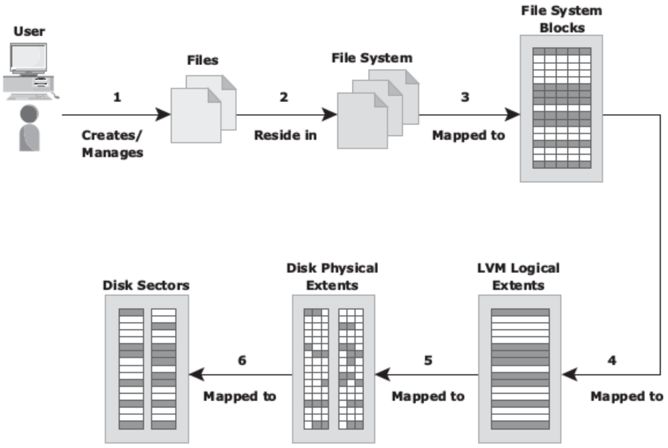


Figure 2-2: Process of mapping user files to disk storage

(b)	**Explain Virtual Memory Manager.** Answer: Memory virtualization enables multiple applications and processes, whose aggregate memory requirement is greater than the available physical memory, to run on a host without impacting each other. • Memory virtualization is an operating system feature that virtualizes the physical memory (RAM) of a host. • It creates virtual memory with an address space larger than the physical memory space present in the compute system. • The virtual memory encompasses the address space of the physical memory and part of the disk storage. • The operating system utility 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 when a process references a virtual address that points to data at the disk storage. • The space used by the VMM on the disk is known as a **swap space**. A swap space (also known as page file or swap file) is a portion of the disk drive that appears to be physical memory to the operating system. • In a virtual memory implementation, the memory of a system is divided into contiguous blocks of fixed-size pages. • A process known as paging moves inactive physical memory pages onto the swap file and brings them back to the physical memory when required. This enables efficient use of the available physical memory among different applications. • The operating system typically moves the least used pages into the swap file so that enough RAM is available for processes that are more active. Access to swap file pages is slower than access to physical memory pages because swap file pages are allocated on the disk drive, which is slower than physical memory.	[02]	CO1	L1
3	**What is RAID? Explain the different RAID levels with neat diagram.** Answer: **RAID:** • **RAID** is an enabling technology that leverages multiple drives as part of a set that provides data protection against drive failures. • RAID implementations also improve the storage system performance by serving I/Os from multiple disks simultaneously. • In 1987, Patterson, Gibson, and Katz at the University of California, Berkeley, published a paper titled "A Case for Redundant Arrays of Inexpensive Disks (RAID)." • This paper described the use of small-capacity, inexpensive disk drives as an alternative to large-capacity drives common on mainframe computers. • The term RAID has been redefined to refer to independent disks to reflect advances in storage technology.	[10]	CO1	L2

RAID LEVELS:

- Application performance, data availability requirements, and cost determine the RAID level selection.
- These RAID levels are defined on the basis of striping, mirroring, and parity techniques.
- Some RAID levels use a single technique, whereas others use a combination of techniques.
- Table below shows the commonly used RAID levels.

Table 3-1: Raid Levels

LEVELS	BRIEF DESCRIPTION
RAID 0	Striped array with no fault tolerance
RAID 1	Disk mirroring
RAID 3	Parallel access array with dedicated parity disk
RAID 4	Striped array with independent disks and a dedicated parity disk
RAID 5	Striped array with independent disks and distributed parity
RAID 6	Striped array with independent disks and dual distributed parity
Nested	Combinations of RAID levels. Example: RAID 1 + RAID 0

RAID 0

- RAID 0 configuration uses data striping techniques, where data is striped across all the disks within a RAID set. Therefore it utilizes the full storage capacity of a RAID set.
- To read data, all the strips are put back together by the controller.
- Figure below shows RAID 0 in an array in which data is striped across five disks.
- When the number of drives in the RAID set increases, performance improves because more data can be read or written simultaneously.
- RAID 0 is a good option for applications that need high I/O throughput. However, if these applications require high availability during drive failures, RAID 0 does not provide data
- protection and availability.

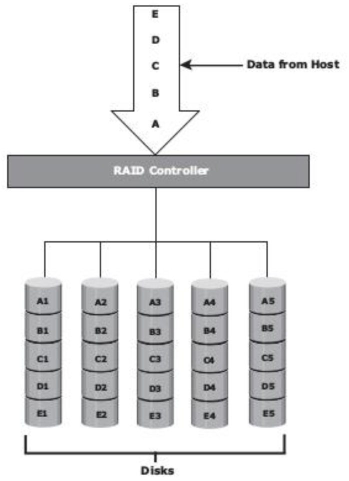


Figure 3-5: RAID 0

RAID 1:

- RAID 1 is based on the mirroring technique.
- In this RAID configuration, data is mirrored to provide fault tolerance as shown in the figure below.
- A RAID 1 set consists of two disk drives and every write is written to both disks.
- During disk failure, the impact on data recovery in RAID 1 is the least among all RAID implementations. This is because the RAID controller uses the mirror drive for data recovery.
- RAID 1 is suitable for applications that require high availability and cost is no constraint.
- The figure shows the difference between RAID 0 and RAID 1

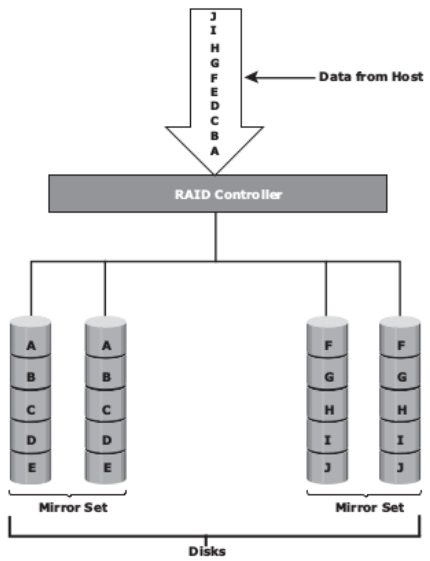
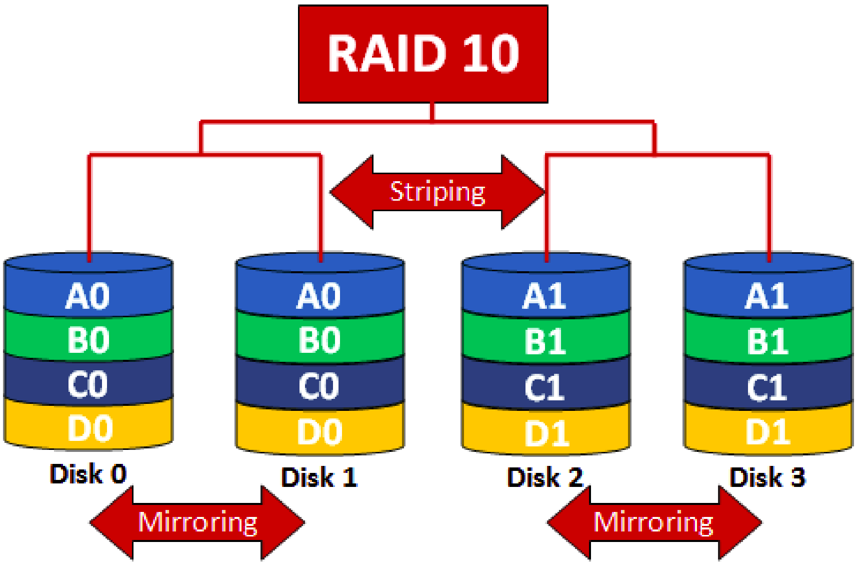
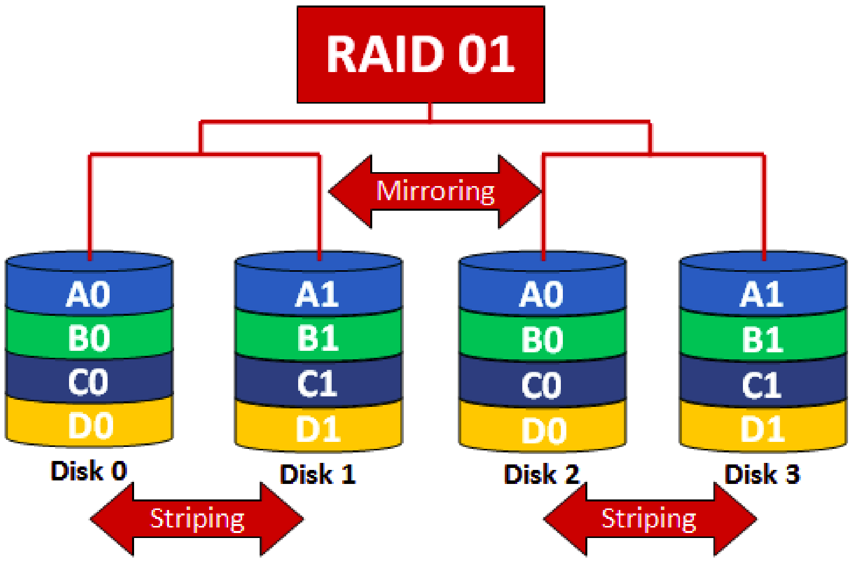


Figure 3-6: RAID 1

Nested RAID

- RAID 1+0 and RAID 0+1 combine the performance benefits of RAID 0 with the redundancy benefits of RAID 1.
- They use striping and mirroring techniques and combine their benefits.
- A common misconception is that RAID 1+0 and RAID 0+1 are the same. Under normal conditions, RAID levels 1+0 and 0+1 offer identical benefits. However, rebuild operations in the case of disk failure differ between the two.
- These types of RAID require an even number of disks, the minimum being four as shown in the figure below.





RAID 1+0:

- RAID 1+0 is also known as RAID 10 (Ten) or RAID 1/0.
- RAID 1+0 performs well for workloads with small, random, write-intensive I/Os.
- Some applications that benefit from RAID 1+0 include the following:

1. High transaction rate Online Transaction Processing (OLTP)
2. Large messaging installations
3. Database applications with write intensive random access workloads

- RAID 1+0 is also called **striped mirror**.
- The basic element of RAID 1+0 is a mirrored pair, which means that data is first mirrored and then both copies of the data are striped across multiple disk drive pairs in a RAID set.
- When replacing a failed drive, only the mirror is rebuilt.
- The disk array controller uses the surviving drive in the mirrored pair for data recovery and continuous operation.
- Data from the surviving disk is copied to the replacement disk.

WORKING OF RAID 1+0:

- Consider an example of six disks forming a RAID 1+0 (RAID 1 first and then RAID 0) set.
- These six disks are paired into three sets of two disks, where each set acts as a RAID 1 set (mirrored pair of disks).
- Data is then striped across all the three mirrored sets to form RAID 0.
- Following are the steps performed in RAID 1+0

Drives 1+2 = RAID 1 (Mirror Set A)
Drives 3+4 = RAID 1 (Mirror Set B)
Drives 5+6 = RAID 1 (Mirror Set C)

- Now, RAID 0 striping is performed across sets A through C.
- In this configuration, if drive 5 fails, then the mirror set C alone is affected. It still has drive 6 and continues to function and the entire RAID 1+0 array also keeps functioning.
- Now, suppose drive 3 fails while drive 5 was being replaced.
- In this case the array still continues to function because drive 3 is in a different mirror set. So, in this configuration, up to three drives can fail without affecting the array, as long as they are all in different mirror sets.

RAID 0+1:

- RAID 0+1 is also known as RAID 01 or RAID 0/1.
- RAID 0+1 is also called a **mirrored stripe.**
- The basic element of RAID 0+1 is a **stripe**.
- This means that the process of striping data across disk drives is performed initially, and then the entire stripe is mirrored.
- In this configuration if one drive fails, then the entire stripe is faulted.

WORKING OF RAID 0+1:

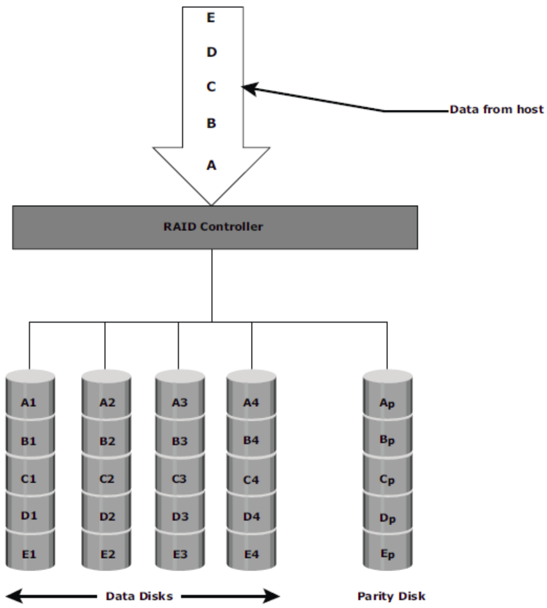
- Consider an example of six disks to understand the working of RAID 0+1 (that is, RAID 0 first and then RAID 1).
- Here, six disks are paired into two sets of three disks each. Each of these sets, in turn, act as a RAID 0 set that contains three disks and then these two sets are mirrored to form RAID 1.
- Following are the steps performed in

RAID 0+1 (see Figure 3-7 [b]):
Drives 1 + 2 + 3 = RAID 0 (Stripe Set A)
Drives 4 + 5 + 6 = RAID 0 (Stripe Set B)

- Now, these two stripe sets are mirrored. If one of the drives, say drive 3, fails, the entire stripe set A fails.
- A rebuild operation copies the entire stripe, copying the data from each disk in the healthy stripe to an equivalent disk in the failed stripe.
- This causes increased and unnecessary I/O load on the surviving disks and makes the RAID set more vulnerable to a second disk failure.

RAID 3:

- RAID 3 stripes data for performance and uses parity for fault tolerance.
- Parity information is stored on a dedicated drive so that the data can be reconstructed if a drive fails in a RAID set.
- For example, in a set of five disks, four are used for data and one for parity.
- Therefore, the total disk space required is 1.25 times the size of the data disks.
- RAID 3 always reads and writes complete stripes of data across all disks because the drives operate in parallel. There are no partial writes that update one out of many strips in a stripe. Figure below illustrates the RAID 3 implementation.
- RAID 3 provides good performance for applications that involve large sequential data access, such as data backup or video streaming.



RAID 4

- Similar to RAID 3, RAID 4 stripes data for high performance and uses parity for improved fault tolerance.
- Data is striped across all disks except the parity disk in the array.
- Parity information is stored on a dedicated disk so that the data can be rebuilt if a drive fails.
- Unlike RAID 3, data disks in RAID 4 can be accessed independently so that specific data elements can be read or written on a single disk without reading or writing an entire stripe. RAID 4 provides good read throughput and reasonable write throughput.

RAID 5

- RAID 5 is a versatile RAID implementation.
- It is similar to RAID 4 because it uses striping.
- The drives (strips) are also independently accessible.
- The difference between RAID 4 and RAID 5 is the parity location.
- In RAID 4, parity is written to a dedicated drive, creating a write bottleneck for the parity disk.
- In RAID 5, parity is distributed across all disks to overcome the write bottleneck of a dedicated parity disk.
- RAID 5 is good for random, read-intensive I/O applications and preferred for messaging, data mining, medium-performance media serving, and relational database management system (RDBMS) implementations, in which database administrators (DBAs) optimize data access.
- Figure illustrates the RAID 5 implementation.

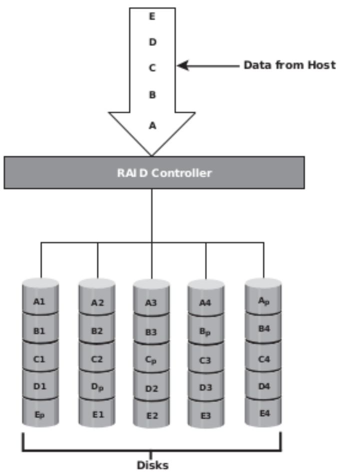
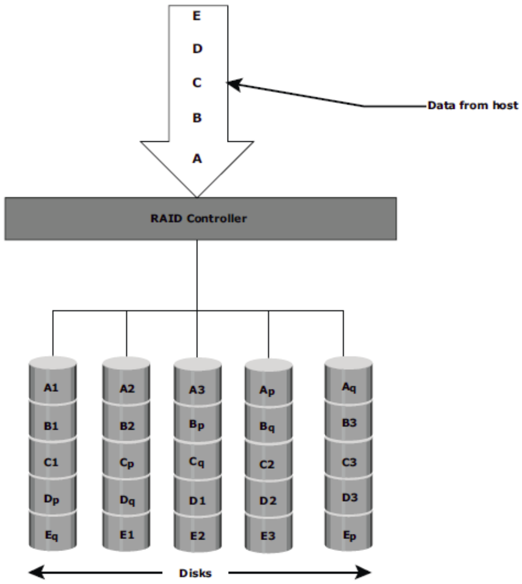


Figure 3-9: RAID 5

RAID 6

- RAID 6 works the same way as RAID 5, except that RAID 6 includes a second parity element to enable survival if two disk failures occur in a RAID set as shown in the figure below.
- Therefore, a RAID 6 implementation requires at least four disks.
- RAID 6 distributes the parity across all the disks.

- The write penalty in RAID 6 is more than that in RAID 5; therefore, RAID 5 writes perform better than RAID 6.
- The rebuild operation in RAID 6 may take longer than that in RAID 5 due to the presence of two parity sets.



4	**With a neat diagram, explain the components of Intelligent Storage System (ISS).**	[10]	CO1	L2

Answer:

COMPONENTS OF AN INTELLIGENT STORAGE SYSTEM

An intelligent storage system consists of four key components:

1. front end
2. cache
3. back end
4. physical disks

- Figure 4-1 illustrates these components and their interconnections.
- 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 the physical disk.
- A read request can be serviced directly from cache if the requested data is found in the cache.
- In modern intelligent storage systems, front end, cache, and back end are typically integrated on a single board (referred to as a storage processor or storage controller).

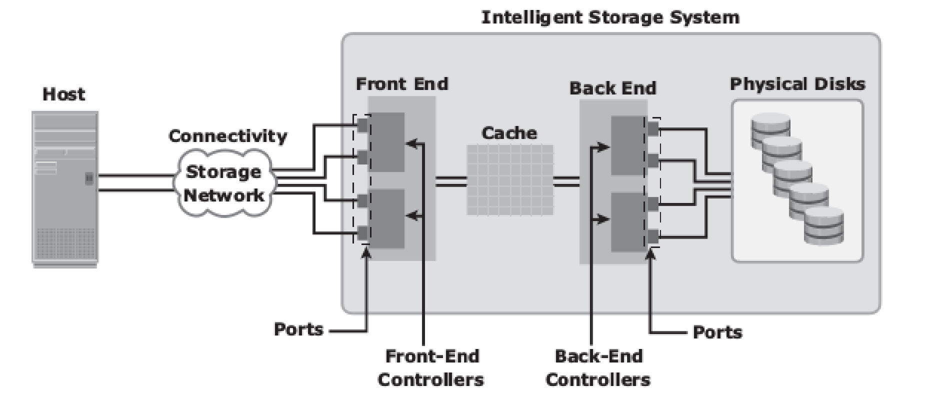


Figure 4-1: Components of an intelligent storage system

4.1.1 FRONT END

- The front end provides the interface between the storage system and the host.
- It consists of two components:
    - front-end ports and
    - front-end controllers.

- Typically a front end has redundant controllers for high availability, and each controller contains multiple ports that enable large numbers of hosts to connect to the intelligent storage system.
- Each front-end controller has processing logic that executes the appropriate transport protocol, such as Fibre Channel, iSCSI, FICON, or FCoE for storage connections.

- Front-end controllers route data to and from cache via the internal data bus. When the cache receives the write data, the controller sends an acknowledgment message back to the host.
- **4.1.2 CACHE**
- Cache is semiconductor memory where data is placed temporarily to reduce the time required to service I/O requests from the host.
- Cache improves storage system performance by isolating hosts from the mechanical delays associated with rotating disks or hard disk drives (HDD).
- Rotating disks are the slowest component of an intelligent storage system. Data access on rotating disks usually takes several milliseconds because of seek time and rotational latency.
- Accessing data from cache is fast and typically takes less than a millisecond. On intelligent arrays, write data is first placed in cache and then written to disk.

STRUCTURE OF CACHE

- Cache is organized into **pages**.
- Page is the smallest unit of cache allocation.
- The size of a cache page is configured according to the application I/O size.
- Cache consists of the **data store and tag RAM**.

Data Store:

- The data store holds the data.

Tag RAM:

- The tag RAM tracks the location of the data in the data store (see Figure 4-2) and in the disk.

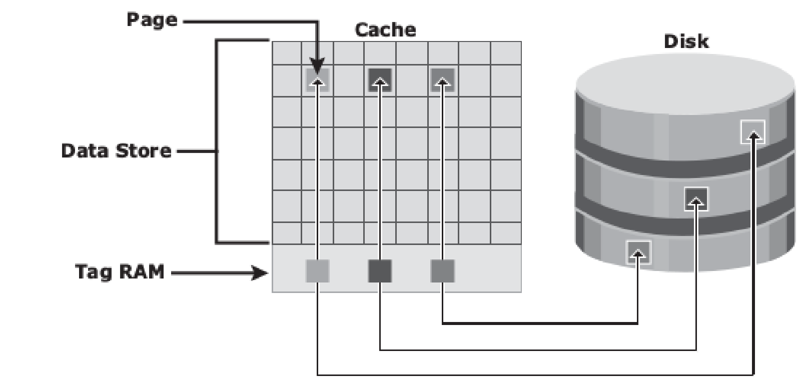


Figure 4-2: Structure of cache

- Entries in tag RAM indicate where data is found in cache and where the data belongs on the disk.
- Tag RAM includes a **dirty bit** flag, which indicates whether the data in cache has been committed to the disk. It also contains time-based information, such as the time of last access, which is used to identify cached information that has not been accessed for a long period and may be freed up.

BACK END

- The back end provides an interface between cache and the physical disks.
- It consists of two components:
    - **back-end ports and**
    - **back-end controllers.**
- The back-end controls data transfers between cache and the physical disks.
- From cache, data is sent to the back end and then routed to the destination disk.
- Physical disks are connected to ports on the back end.

- The back-end controller communicates with the disks when performing reads and writes and also provides additional, but limited, temporary data storage.
- The algorithms implemented on back-end controllers provide error detection and correction, along with RAID functionality.
- For high data protection and high availability, storage systems are configured with dual controllers with multiple ports. Such configurations provide an alternative path to physical disks if a controller or port failure occurs.
- This reliability is further enhanced if the disks are also dual-ported. In that case, each disk port can connect to a separate controller.
- Multiple controllers also facilitate load balancing.

PHYSICAL DISK

- Physical disks are connected to the back-end storage controller and provide persistent data storage.
- Modern intelligent storage systems provide support to a variety of disk drives with different speeds and types, such as FC (Fibre Channel), SATA (Serial Advanced Technology Attachment), SAS (Serial Attached SCSI), and flash drives.
- They also support the use of a mix of flash, FC, or SATA within the same array.

5	**Explain FC connectivity options with relevant diagram.**	[05]	CO2	L2

Answer:
The FC architecture supports three basic interconnectivity options:

1. point-to-point
2. arbitrated loop
3. Fibre Channel switched fabric.

1. Point-to-Point

Two devices are connected directly to each other (Figure 2-5).

Advantage:

1) Provides a dedicated-connection for data-transmission between nodes.

Disadvantages:

1) Provides limited connectivity, '.' only 2 devices can communicate with each other at given time

2) Not Scalable: Cannot be scaled to accommodate a large number of network-devices.

Standard DAS uses point-to-point connectivity.

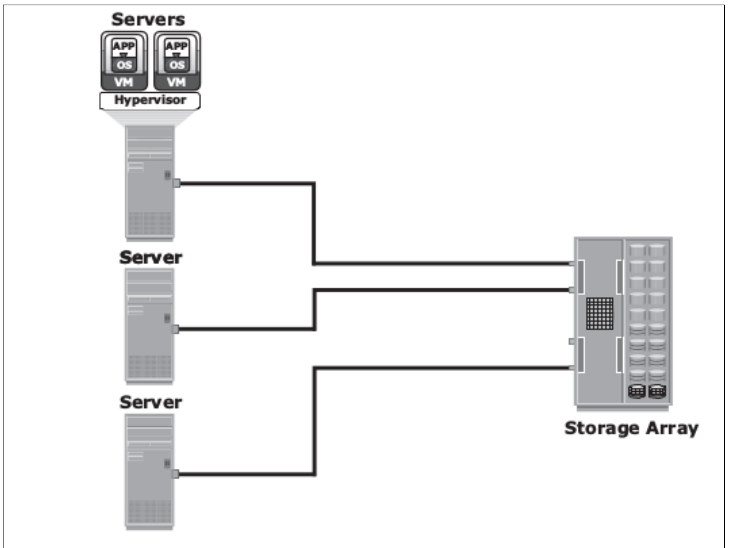


Figure 5-6: Point-to-point connectivity

2. Fibre Channel Arbitrated Loop

In the FC-AL confi guration, devices are attached to a shared loop. FC-AL has the characteristics of a token ring topology and a physical star topology. In FC-AL, each device contends with other devices to perform I/O operations. Devices on the loop must “arbitrate” to gain control of the loop. At any given time, only one device can perform I/O operations on the loop.

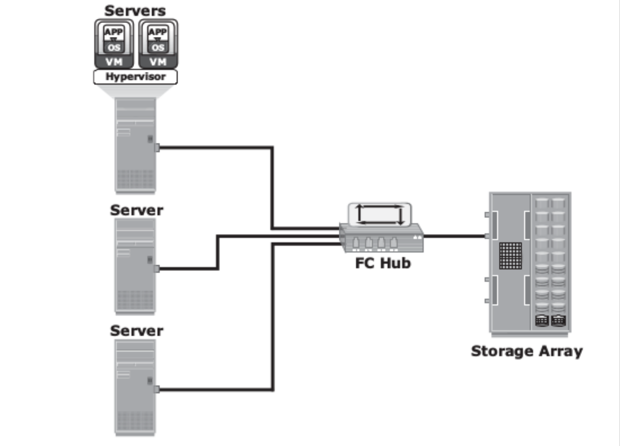


Figure 5-7: Fibre Channel Arbitrated Loop

As a loop configuration, FC-AL can be implemented without any interconnecting devices by directly connecting one device to another two devices in a ring through cables.
However, FC-AL implementations may also use hubs whereby the arbitrated loop is physically connected in a star topology.
The FC-AL configuration has the following limitations in terms of scalability:

- FC-AL shares the loop and only one device can perform I/O operations at a time. Because each device in a loop must wait for its turn to process an I/O request, the overall performance in FC-AL environments is low.
- FC-AL uses only 8-bits of 24-bit Fibre Channel addressing (the remaining 16-bits are masked) and enables the assignment of 127 valid addresses to the ports. Hence, it can support up to 127 devices on a loop. One address is reserved for optionally connecting the loop to an FC switch port. Therefore, up to 126 nodes can be connected to the loop.
- Adding or removing a device results in loop re-initialization, which can cause a momentary pause in loop traffic.

3. Fibre Channel Switched Fabric

Unlike a loop configuration, a Fibre Channel switched fabric (FC-SW) network provides dedicated data path and scalability. The addition or removal of a device in a switched fabric is minimally disruptive; it does not affect the ongoing traffic between other devices.
FC-SW is also referred to as fabric connect. A fabric is a logical space in which all nodes communicate with one another in a network. This virtual space can be created with a switch or a network of switches. Each switch in a fabric contains a unique domain identifier, which is part of the fabric's addressing scheme. In FC-SW, nodes do not share a loop; instead, data is transferred through a dedicated path between the nodes. Each port in a fabric has a unique 24-bit Fibre Channel address for communication. Figure 5-8 shows an example of the FC-SW fabric.
In a switched fabric, the link between any two switches is called an Interswitch link (ISL). ISLs enable switches to be connected together to form a single, larger fabric. ISLs are used to transfer host-to-storage data and fabric management traffic from one switch to another. By using ISLs, a switched fabric can be expanded to connect a large number of nodes.

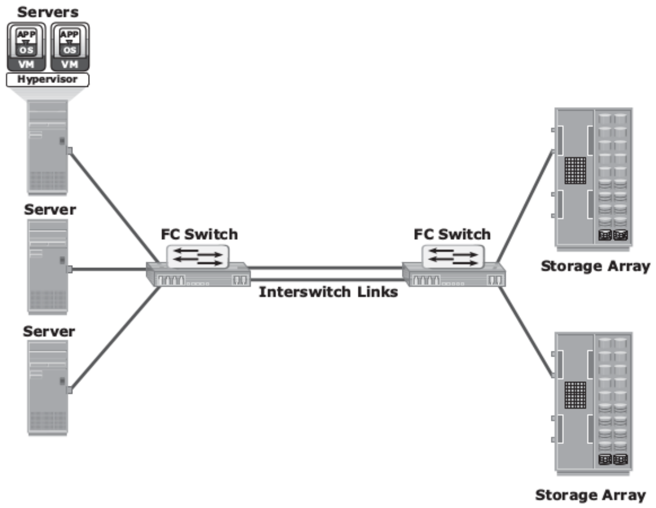


Figure 5-8: Fibre Channel switched fabric

6	**Define zoning and explain type of zoning.**	[10]	CO2	L1

Answer:
Zoning is an FC switch function that enables node ports within the fabric to be logically segmented into groups and to communicate with each other within the group.

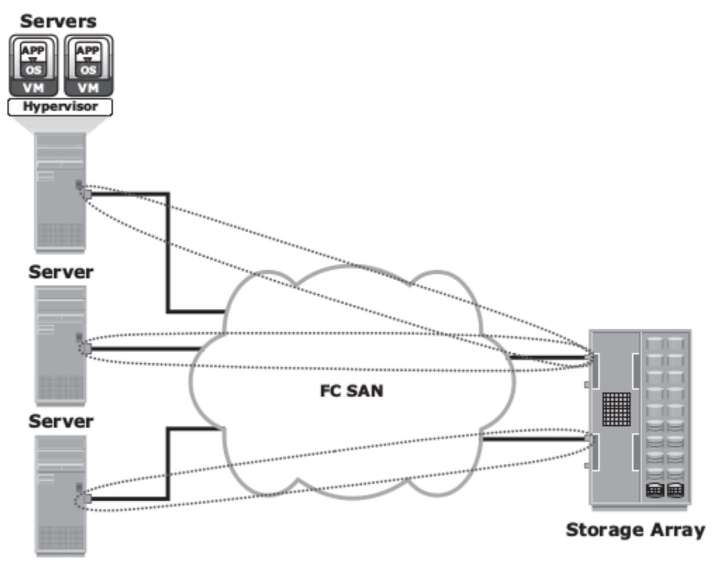


Figure 5-17: Zoning

Zone members, zones, and zone sets form the hierarchy defined in the zoning process. A zone set is composed of a group of zones that can be activated or deactivated as a single entity in a fabric. Multiple zone sets may be defined in a fabric, but only one zone set can be active at a time. Members are nodes within the SAN that can be included in a zone. Switch ports, HBA ports, and storage device ports can be members of a zone. A port or node can be a member of multiple zones. Nodes distributed across multiple switches in a switched fabric may also be grouped into the same zone. Zone sets are also referred to as zone configurations. Zoning provides control by allowing only the members in the same zone to establish communication with each other.

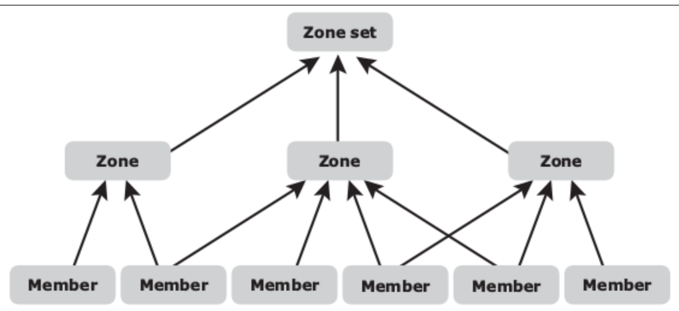


Figure 5-18: Members, zones, and zone sets

Types of Zoning
Zoning can be categorized into three types:

Port zoning: Uses the physical address of switch ports to defi ne zones. In port zoning, access to node is determined by the physical switch port to which a node is connected. The zone members are the port identifi er (switch domain ID and port number) to which HBA and its targets (storage devices) are connected. If a node is moved to another switch port in the fabric, then zoning must be modifi ed to allow the node, in its new port, to participate in its original zone. However, if an HBA or storage device port fails, an administrator just has to replace the failed device without changing the zoning confi guration.

WWN zoning: Uses World Wide Names to defi ne zones. The zone members are the unique WWN addresses of the HBA and its targets (storage devices). A major advantage of WWN zoning is its fl exibility. WWN zoning allows nodes to be moved to another switch port in the fabric and maintain connectivity to its zone partners without having to modify the zone confi guration. This is possible because the WWN is static to the node port.

Mixed zoning: Combines the qualities of both WWN zoning and port zoning. Using mixed zoning enables a specifi c node port to be tied to the WWN of another node.

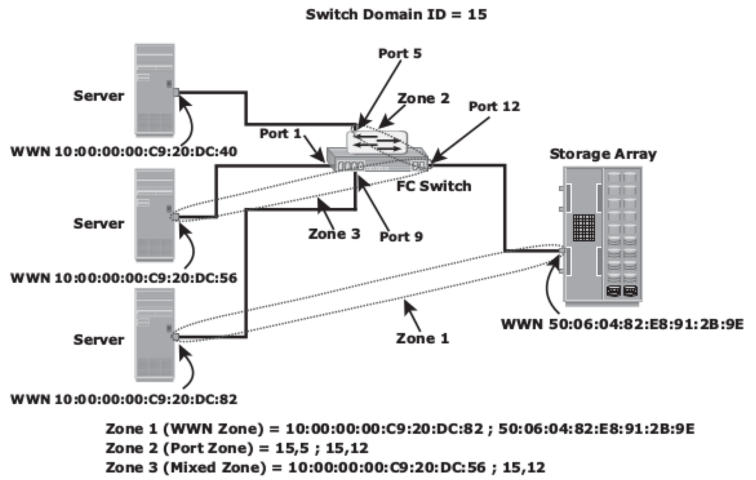


Figure 5-19: Types of zoning