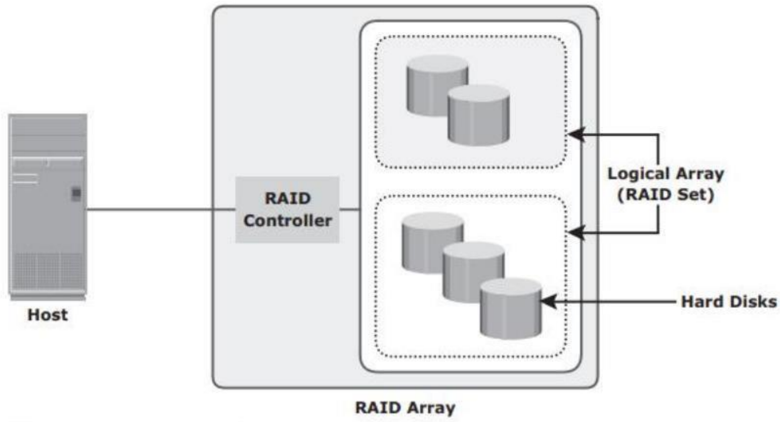
MODULE 2 QB

-PP

1) With the help of neat diagram, explain the components of a RAID array.

Ans) A RAID (Redundant Arrays of Independent Disks) combines two or more disk drives in an array into a RAID set or a RAID group.



Components of a RAID array

1) Host:

- The host is the computer or device that is connected to the RAID array.
- It is responsible for sending data to and receiving data from the array.

2) RAID Array:

- The RAID array is the enclosure that contains the RAID controller, Hard disks and Logical array.
- It can be either an external or internal device, depending on the type of RAID array.
- The RAID array must be created and managed by the RAID controller.

3) RAID Controller:

- The RAID controller is a hardware device that manages the RAID array.
- It is responsible for organizing and distributing data across the hard disks in the array.
- RAID controller is used for the calculation of parity.
- It provides data redundancy, fault tolerance and performance optimization for the disk array.

4) Logical Array:

- The logical array is the virtual disk that is created by the RAID controller and presented to the host.
- It can be partitioned and formatted just like a physical disk.

5) Hard Disks:

- Hard disks are the physical storage devices that make up the RAID array.
- Hard disks must be of the same type, size and speed to be used in a RAID array.
- Hard disks can be replaced or added to the array as needed, without affecting the logical array.

2) With a neat diagram, explain different RAID Techniques.

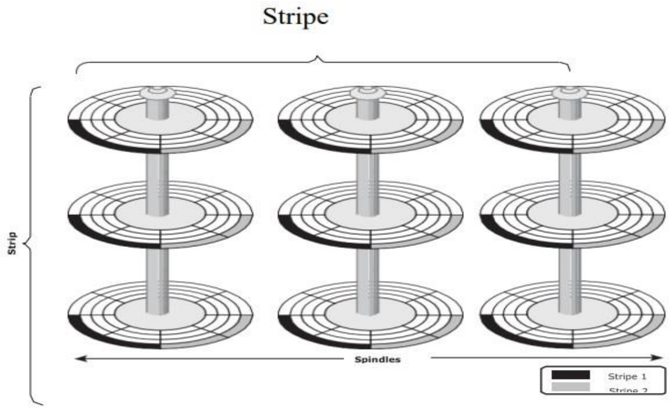
Ans) There are three RAID techniques:

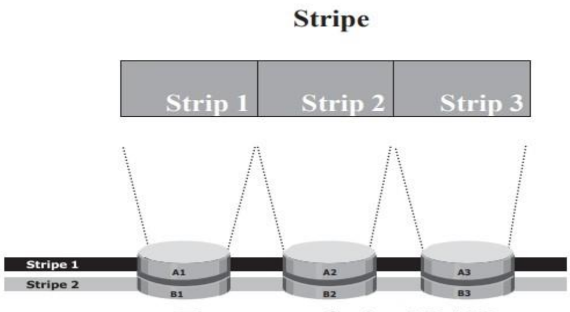
1. Striping
2. Mirroring
3. Parity

1) Striping:

- Striping is a technique to spread data across multiple drives to use the drives in parallel.
- All the read-write heads work simultaneously, allowing more data to be processed in a shorter time and increasing performance, compared to reading and writing from a single disk.

- Within each disk in a RAID set, a predefined number of contiguously addressable disk blocks are defined as a strip.
- The set of aligned strips that spans across all the disks within the RAID set is called a stripe.
- Stripe size (also called Stripe depth) describes the number of blocks in a strip.
- Stripe size is the maximum amount of data that can be written to or read from a single disk in the set.
- Stripe width refers to the number of data strips in a stripe.
- Striped RAID does not provide any data protection unless parity or mirroring is used.





Striped RAID set

2) Mirroring:

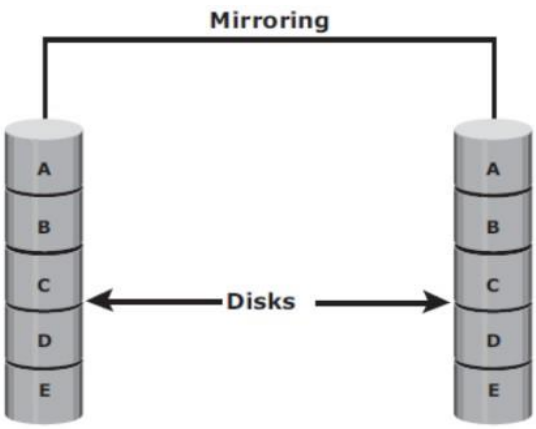
- Mirroring is a technique whereby the same data is stored on two different disk drives, yielding two copies of the data.
- If one of two mirrored disk drives fails, the other drive still has all the data and can continue to work without any interruption.
- When the failed disk is replaced with a new disk, the controller copies the data from the surviving disk of the mirrored pair.
- This activity is transparent to the host.

Advantages:

- complete data redundancy
- mirroring enables fast recovery from disk failure.
- data protection

Disadvantages:

- Mirroring involves duplication of data
- Expensive

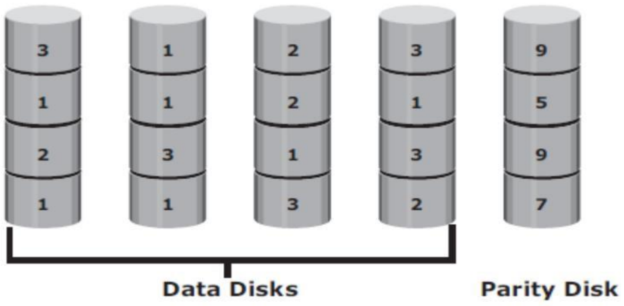


Mirrored disks in an array

3) Parity:

- Parity is a method to protect striped data from disk drive failure without the cost of mirroring.
- An additional disk drive is added to hold parity, a mathematical construct that allows recreation of the missing data.

- Parity is a redundancy technique that ensures protection of data without maintaining a full set of duplicate data.
- Calculation of parity is a function of the RAID controller.
- In the below figure, the first four disks, labelled "Data Disks," contain the data.
- The fifth disk, labelled "Parity Disk," stores the parity information, which, in this case, is the sum of the elements in each row.



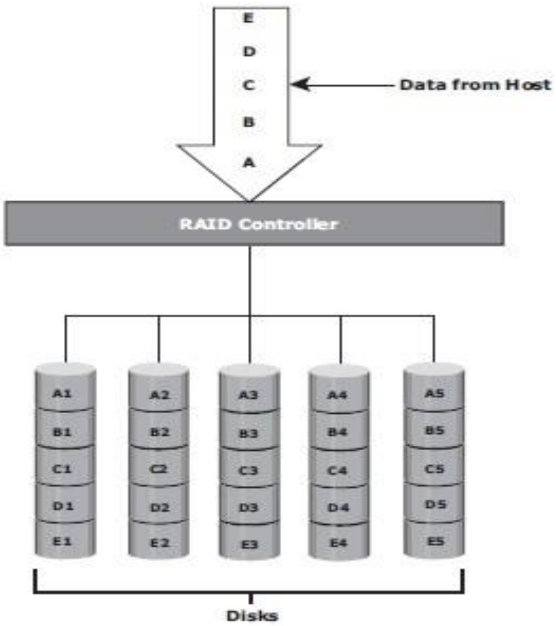
Parity RAID

3) Illustrate the following RAID levels:

i) RAID 0

Ans)

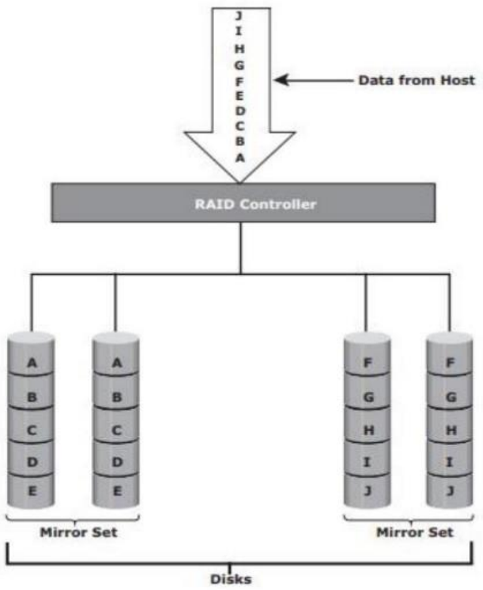
- RAID 0 configuration uses data striping techniques, where data is striped across all the disks within a RAID set.
- To read data, all the strips are put back together by the controller.
- Figure shows RAID 0 in an array in which data is striped across five disks.
- When the number of drives in the RAID set increases, performance improved.
- RAID 0 is a good option for applications that need high I/O throughput.



ii) RAID 1

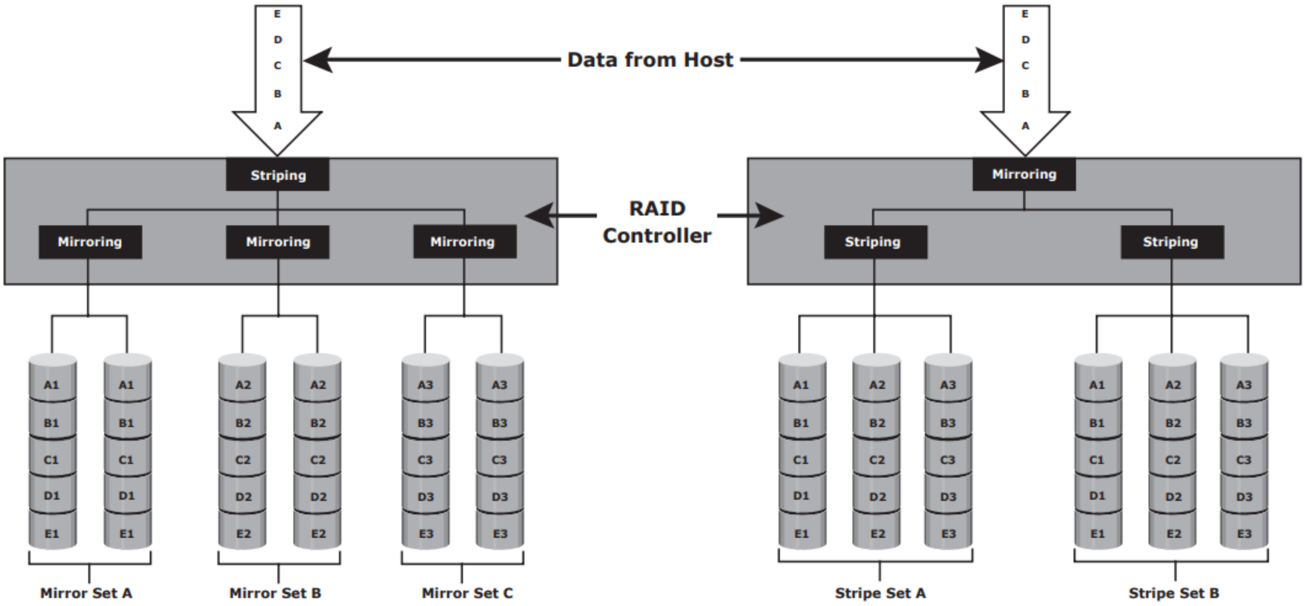
Ans)

- RAID 1 is based on the mirroring technique.
- Here, data is mirrored to provide fault tolerance.
- RAID 1 set consists of two disk drives and every write is written to both disks.
- The mirroring is transparent to the host.
- 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.



iii) Nested RAID

Ans) RAID 1+0 and RAID 0+1 combine together to form a Nested RAID.



(a) RAID 1+0 (b) RAID 0+1

<u>RAID 1+0</u>

- RAID 1+0 is also known as RAID 10 (Ten) or RAID 1/0.
- 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.
- Data from the surviving disk is copied to the replacement disk.
- Consider an example of six disks forming a RAID 1+0.
- These six disks are paired into three sets of two 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)

RAID 0+1:

- RAID 0+1 is also called a mirrored stripe.
- 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.
- Six disks are paired into two sets of three disks each.
- Each of two sets, act as a RAID 0 set and then these two sets are mirrored to form RAID 1.
- Following are the steps performed in RAID 0+1

  Drives 1 + 2 + 3 = RAID 0 (Stripe Set A)

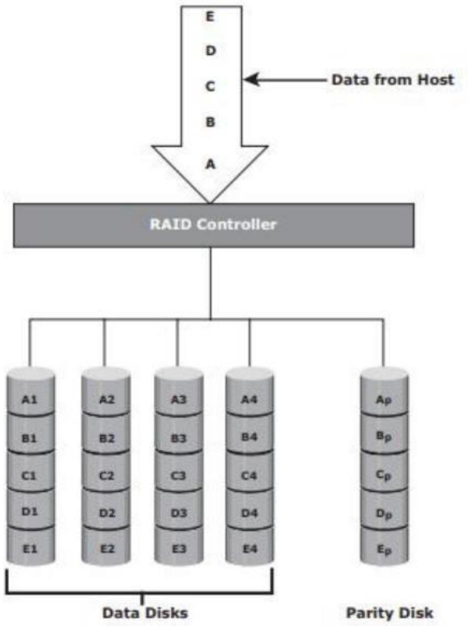
  Drives 4 + 5 + 6 = RAID 0 (Stripe Set B)

Now, these two stripe sets are mirrored.

iv) RAID 3

Ans)

- RAID 3 stripes data for high performance and uses parity for improved fault tolerance.
- Parity information is stored on a dedicated drive so that data can be reconstructed if a drive fails.
- For example, of five disks, four are used for data and one is used for parity.
- 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.
- RAID 3 is used in applications that involve large sequential data access, such as video streaming.



v) RAID 4

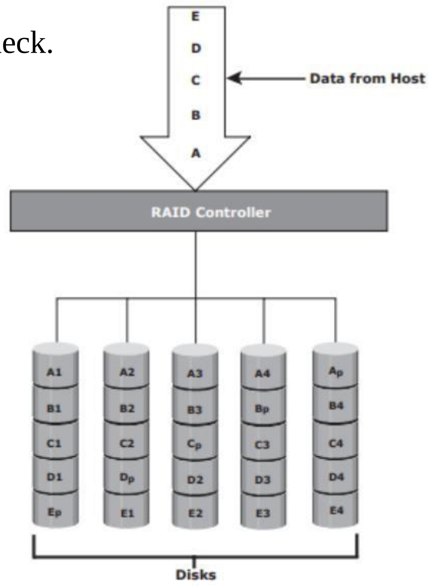
Ans)

- 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.
- Striping is done at the block level.
- Unlike RAID 3, data disks in RAID 4 can be accessed independently so that specific data elements can be read or written on single disk without read or write of an entire stripe.
- RAID 4 provides good read throughput and reasonable write throughput

vi) RAID 5

Ans)

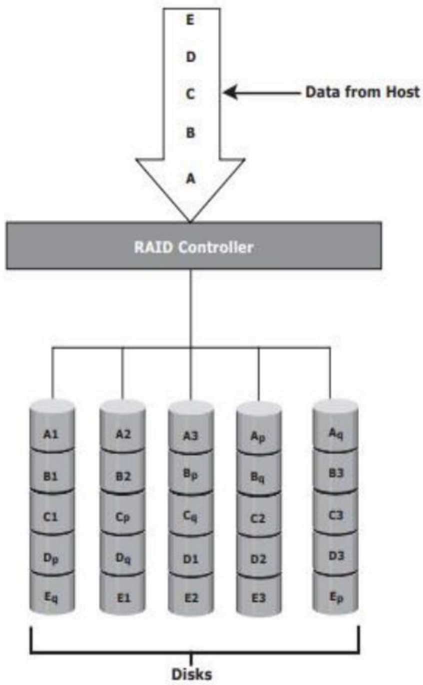
- RAID 5 is a versatile RAID implementation.
- It is similar to RAID 4 because it uses striping.
- The drives (strips) are also independently accessible.
- The difference between RAID 4 and RAID 5 is the parity location.
- In RAID 4, parity is written to a dedicated drive, creating a write bottleneck for the parity disk.
- In RAID 5, parity is distributed across all disks.
- The distribution of parity in RAID 5 overcomes the Write bottleneck.



vii) RAID 6

Ans)

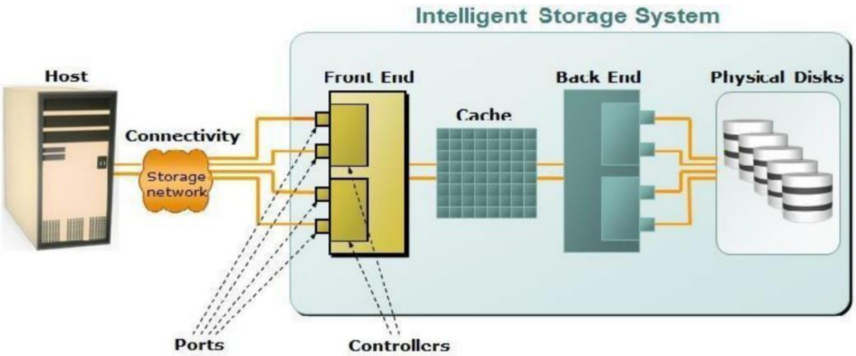
- RAID 6 includes a second parity element to enable survival in the event of the failure of two disks in a RAID group.
- RAID 6 distributes the parity across all the disks.
- The write penalty in RAID 6 is more than that in RAID 5.
- Therefore, RAID 5 writes perform better than RAID 6.
- The rebuild operation in RAID 6 may take longer than that in RAID 5 due to the presence of two parity sets.



4) With a neat block diagram, describe the components of Intelligent storage system.

Ans) An intelligent storage system consists of four key components:

1) Front End
2) Cache
3) Back end
4) Physical Disks



1) Front End:

- The front end provides the interface between the storage system and the host.
- It consists of two components:
  i) Front-End Ports
  ii) Front-End Controllers.
- A front end has redundant controllers for high availability, and each controller contains multiple frontend 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.

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.
- Rotating disks are the slowest component of an intelligent storage system.
- Accessing data from rotating disks usually takes several millisecond 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.

3) Back end:

- The back end provides an interface between cache and the physical disks.
- It consists of two components:
  i) Back-end ports
  ii) 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.
- It also provides additional, but limited, temporary data storage.

4) Physical Disks:

- A physical disk stores data persistently.
- Physical disks are connected to the back-end storage controller.
- Modern intelligent storage systems provide support to a variety of disk drives with different speeds and types, such as FC, SATA, SAS and flash drives.
- They also support the use of a mix of flash, FC, or SATA within the same array.

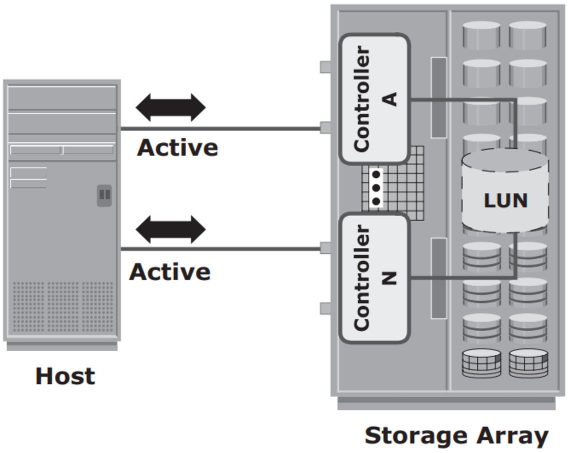
5) Explain types of Intelligent System.

Ans) Intelligent storage systems are of two types:

1) High-end storage systems
2) Midrange storage systems

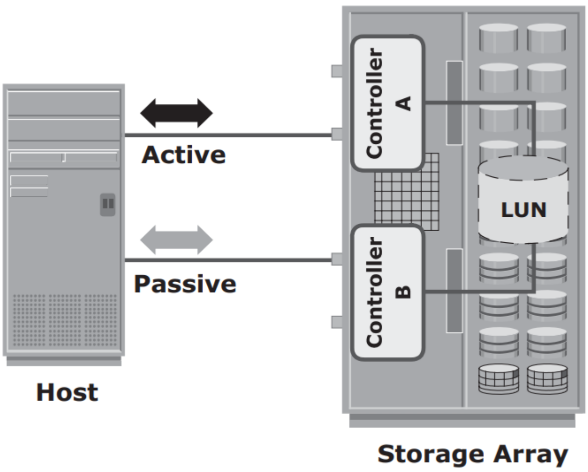
1) High-end storage system:

- High-end storage systems are referred to as active-active arrays.
- They are generally used for large enterprise applications.
- These systems are designed with a large number of controllers and cache memory.
- They provide the following capabilities:
  - ➢ Large storage capacity
  - ➢ Huge cache to service host I/Os
  - ➢ Fault tolerance architecture
  - ➢ Multiple front-end ports and support to interface protocols
  - ➢ High scalability
  - ➢ Ability to handle large amounts of concurrent I/Os
  - ➢ Designed for large enterprises
  - ➢ Connectivity to mainframe computers and open systems hosts



2) Midrange Storage Systems:

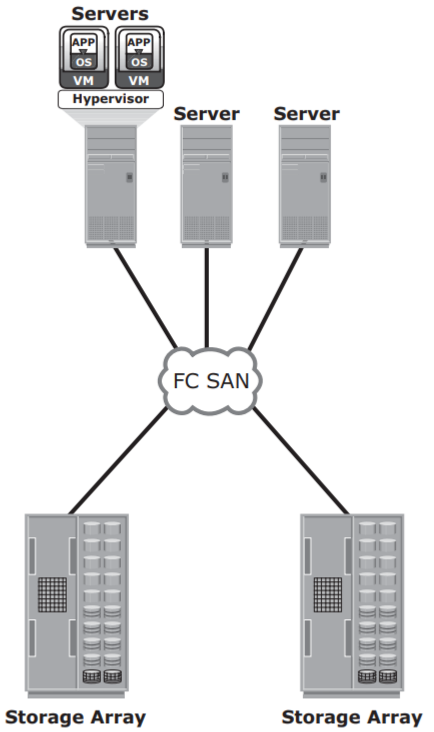
- Midrange storage systems are also referred to as active-passive arrays.
- They are best suited for small-sized and medium-sized enterprise applications.
- They also provide optimal storage solutions at a lower cost.
- Midrange array have two controllers, each with cache, RAID controllers and disk drive interfaces.
- There are less scalable as compared to high-end array.
- They ensure high redundancy and high performance for applications with predictable workloads.



6) With a neat diagram, explain San and it's Evolution.

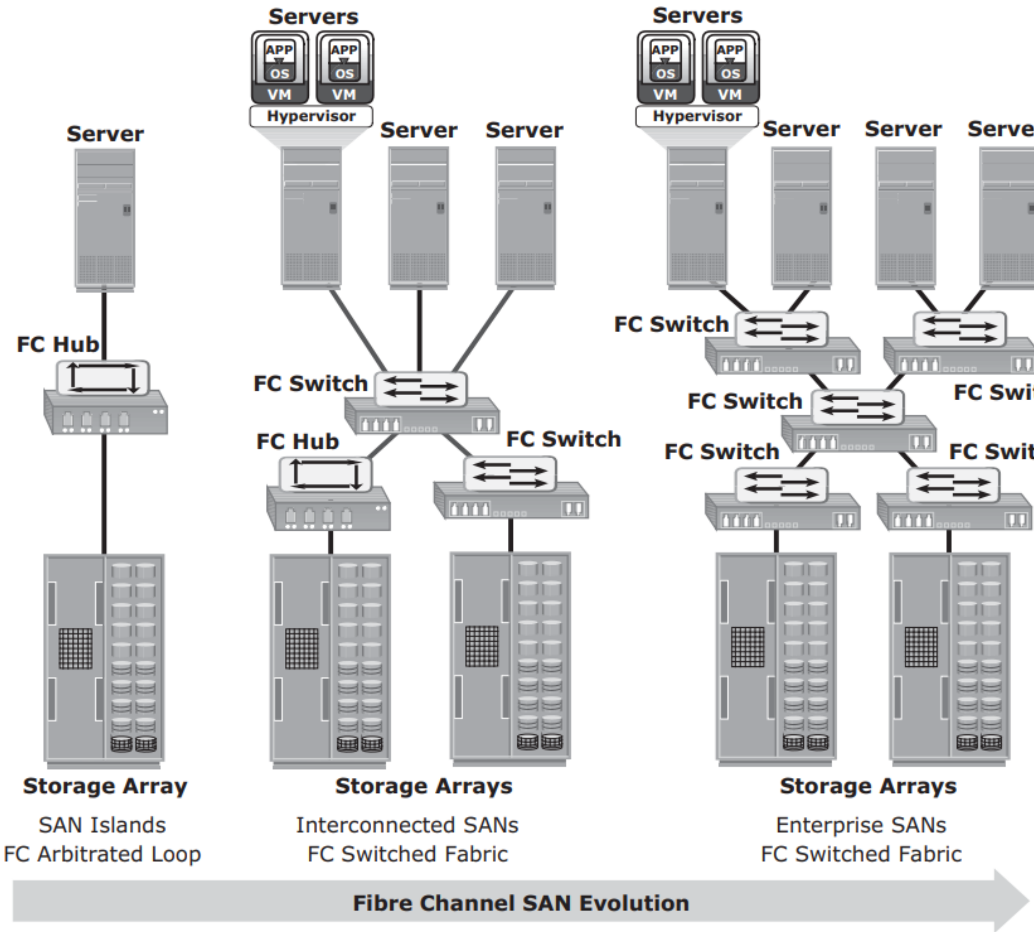
Ans)

- A SAN carries data between servers (or hosts) and storage devices through Fibre Channel network.
- A SAN enables storage consolidation and enables storage to be shared across multiple servers.
- This improves the utilization of storage resources compared to direct-attached storage architecture.
- It reduces the total amount of storage an organization needs to purchase and manage.
- With consolidation, storage management becomes centralized and less complex, which further reduces the cost of managing information.
- SAN also enables organizations to connect geographically dispersed servers and storage.



FC SAN implementation

- In its earliest implementation, the FC SAN was a simple grouping of hosts and storage devices connected to a network using an FC hub as a connectivity device.
- This configuration of an FC SAN is known as a Fibre Channel Arbitrated Loop (FC-AL).
- Use of hubs resulted in isolated FC-AL SAN islands.
- This is because the hubs provide limited connectivity and bandwidth.
- The inherent limitations associated with hubs gave way to high-performance FC switches.
- Use of switches in SAN improved connectivity and performance and enabled FC SANs to be highly scalable.
- This enhanced data accessibility to applications across the enterprise.
- Now, FC-AL has been almost abandoned for FC SANs due to its limitations but still survives as a back-end connectivity option to disk drives.



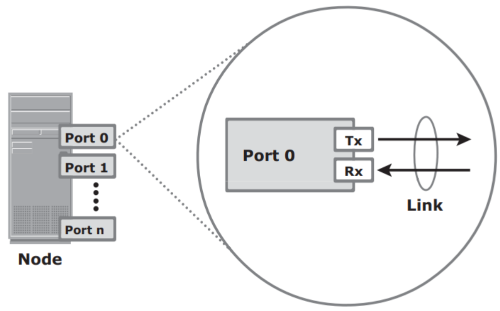
FC SAN evolution

7) Explain Components of FC San.

Ans) FC SAN infrastructure consists of node ports, cables, connectors and interconnecting devices, along with SAN management software.

1) Node Ports:

- In a Fibre Channel network, the end devices, such as hosts, storage arrays and tape libraries, are all referred to as nodes.
- Each node is a source or destination of information.
- Each node requires one or more ports to provide a physical interface for communicating with other nodes.
- Examples of nodes: Hosts, storage and tape library
- Ports are available on:
  - ➢ HBA in host
  - ➢ Front-end adapters in storage
  - ➢ Each port has transmit (Tx) link and receive (Rx) link.



2) Cables:

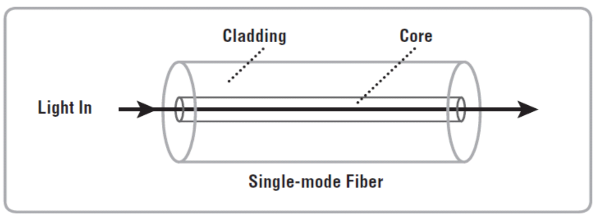
SAN implementation uses:

- Copper cables for short distance
- Optical fiber cables for long distance

Two types of optical cables are:

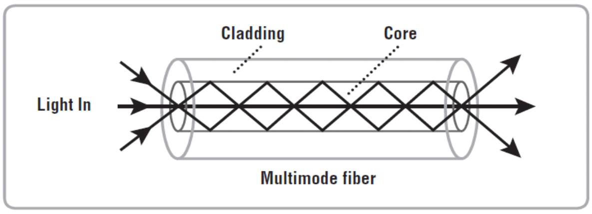
1) Single-mode:

- Can carry single beams of light
- Distance up to 10 KM



2) Multi-mode:

- Can carry multiple beams of light simultaneously
- Distance up to 500 meters



3) Connectors

A connector is attached at the end of a cable to enable swift connection and disconnection of the cable to and from a port.

1) Node Connectors:

- SC Duplex Connectors
- LC Duplex Connectors

2) Patch panel Connectors

- ST Simplex Connectors

4) Interconnect Devices:

Hubs, switches and directors are the interconnect devices commonly used in FC SAN.

i) Hubs:

- Hubs are used as communication devices in FC-AL implementations.
- Hubs physically connect nodes in a logical loop or a physical star topology.
- All the nodes must share the loop because data travels through all the connection points.

ii) Switches:

- Switches are more intelligent than hubs and directly route data from one physical port to another.
- Nodes do not share the bandwidth.
- Instead, each node has a dedicated communication path.

iii) Directors:

- Directors are high-end switches with a higher port count and better fault-tolerance capabilities.

5) SAN Management Software:

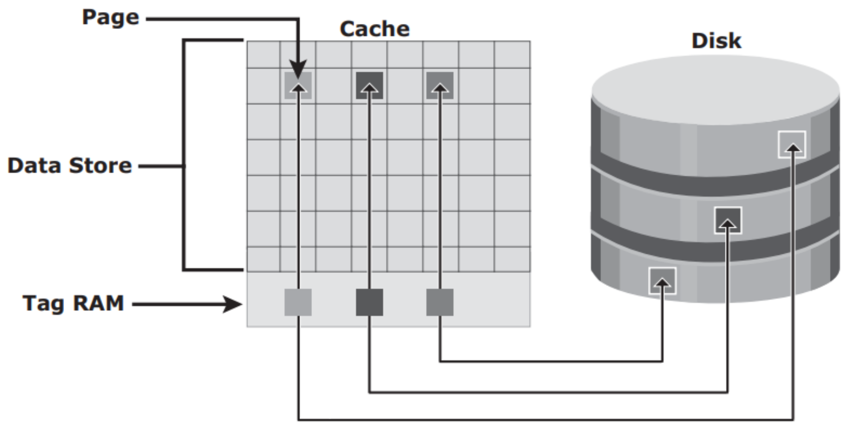
- SAN management software manages the interfaces between hosts, interconnect devices and storage arrays.
- The software provides a view of the SAN environment.
- The software also enables management of various resources from one central console.
- It provides key management functions, including mapping of storage devices, switches and servers.
- This helps in monitoring and generating alerts for discovered devices and zoning.

8) With a neat diagram, explain the structure of cache with read and write operations.

Ans) Cache is semiconductor memory where data is placed temporarily to reduce the time required to service I/O requests from the host.

Structure of Cache:

- Cache is organized into pages, which 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.
- The data store holds the data.
- The tag RAM tracks the location of the data in the data store and in the disk.

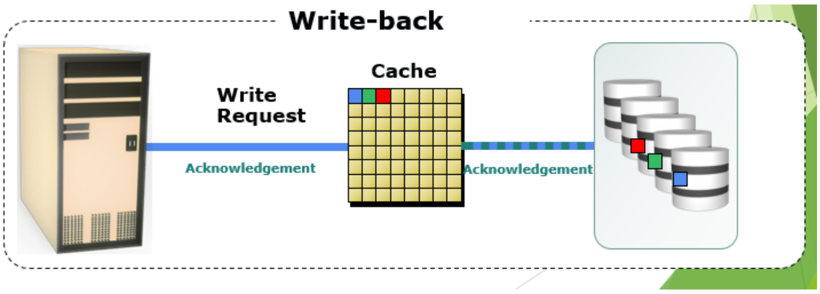


Write Operation:

A write operation with cache is implemented in the following ways:

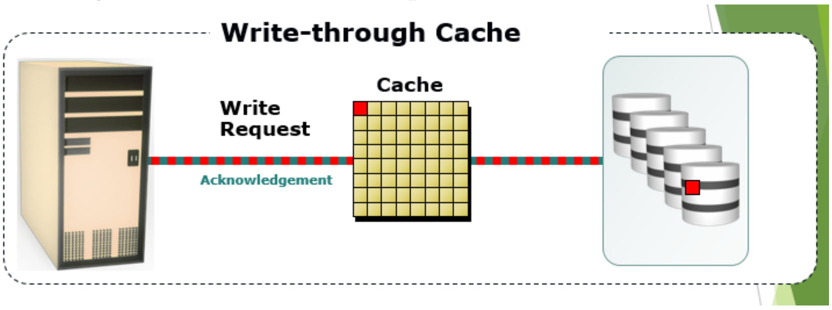
1) Write-back cache:

- Data is placed in cache and an acknowledgment is sent to the host immediately.
- Later, data from several writes are committed to the disk.
- Write response times are much faster because the write operations are isolated from the mechanical delays of the disk.
- However, uncommitted data is at risk of loss if cache failures occur.



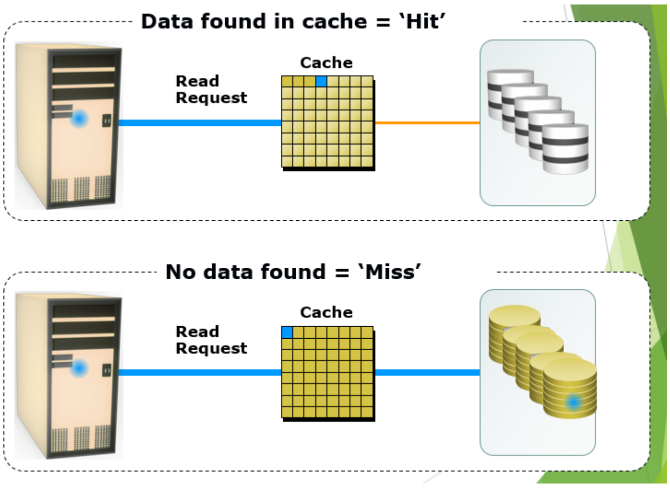
2) Write-through cache:

- Data is placed in the cache and immediately written to the disk and an acknowledgment is sent to the host.
- Because the data is committed to disk as it arrives, the risks of data loss are low.
- Write-response time is longer because of the disk operations.



Read Operation:

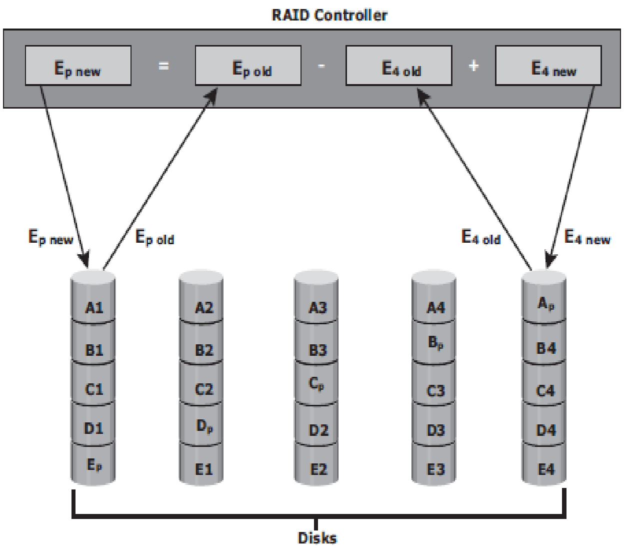
- When a host issues a read request, the storage controller reads the tag RAM to determine whether the required data is available in cache.
- If the requested data is found in the cache, it is called a read cache hit or read hit and data is sent directly to the host, without any disk operation.
- If the requested data is not found in cache, it is called a cache miss and the data must be read from the disk.
- The back end accesses the appropriate disk and retrieves the requested data.
- Data is then placed in cache and finally sent to the host through the front end.
- Cache misses increase the I/O response time.



9) Discuss RAID impact on disk performance.

Ans)

- In both mirrored and parity RAID configurations, every write operation translates into more I/O overhead for the disks, which is referred to as a write penalty.
- In a RAID 1 implementation, every write operation must be performed on two disks configured as a mirrored pair.
- In a RAID 5 implementation, a write operation may manifest as four I/O operations.
- In RAID 5 configuration, when performing I/Os to a disk, the controller has to read, recalculate and write a parity segment for every data write operation.
- Figure illustrates a single write operation on RAID 5 that contains a group of five disks.



- The new parity ($E_{p\ new}$) is computed as follows:

$$\mathbf{E_{p\ new} = E_{p\ old} - E_{4\ old} + E_{4\ new}} \text{ (XOR operations)}$$

- In RAID 5, for every write operation the controller performs two disk reads and two disk writes and the write penalty is 4.
- In RAID 6, the controller performs six I/O operations for each write I/O, and the write penalty is 6.

10) Explain cache management and data protection methods.

Ans)

<u>Cache Management:</u>

- Cache is a finite and expensive resource that needs proper management.
- When all cache pages are filled, some pages have to be freed up to accommodate new data and avoid performance degradation.

- Cache management algorithms are:
  - Least Recently Used (LRU)
  - Most Recently Used (MRU)

Cache Management using Watermarking:

- Flushing is the process that commits data from cache to the disk.
- On the basis of the I/O access rate and pattern, high and low levels called watermarks are set in cache to manage the flushing process.

High watermark (HWM): It is the cache utilization level at which the storage system starts high-speed flushing of cache data.

Low watermark (LWM): It is the point at which the storage system stops flushing data to the disks.

Watermark manages peak I/O requests "bursts" through three important techniques - **Idle flushing, High Watermark flushing and Forced flushing**.

Cache Data Protection:

Protecting cache data against failure can be done in two ways:

1) Cache mirroring:

Each write to the cache is held in two different memory locations on two independent memory cards.

2) Cache vaulting:

Cache is exposed to the risk of uncommitted data loss due to power failure.

In the event of power failure, uncommitted data is dumped to a dedicated set of drives called vault drives.

Problems:

RAID	Write Penalty
0	1
1	2
3	4
5	4
6	6
DP	2
10	2

(You need to know these above values for the calculation of Disk Load).

1) Consider an application that generates 7,200 IOPS, with 60 percent of them being reads.
i) Show the disk load in RAID 1, RAID 5 and RAID 6.
ii) If a HDD with the specification of maximum 180 IOPS for the application needs to be used, show the number of disks required to meet the workload for RAID 1, RAID 5 and RAID 6.

Ans) i)

Disk Load for RAID 1 = 7200*0.6 + (**2***0.4*7200) = 10080 IOPS

Disk Load for RAID 5 = 7200*0.6 + (**4***0.4*7200) = 15840 IOPS

Disk Load for RAID 6 = 7200*0.6 + (**6***0.4*7200) = 21600 IOPS

ii) Number of drives required to support the application in different RAID environments, drives with a rating of 180 IOPS.

For RAID 1 = 10080 /180 = 56 disks
For RAID 5 = 15840 /180 = 88 disks
For RAID 6 = 21600 /180 = 120 disks

2) Consider an application that generates 11,200 IOPS, with 60 percent of them being reads.
i) Show the disk load in RAID 1, RAID 5 and RAID 6.
ii) If a HDD with the specification of maximum 130 IOPS for the application needs to be used, show the number of disks required to meet the workload for RAID 1, RAID 5 and RAID 6.
Ans) i)

Disk Load for RAID 1 = 11200*0.6 + (**2***0.4*11200) = 15680 IOPS

Disk Load for RAID 5 = 11200*0.6 + (**4***0.4*11200) = 24640 IOPS

Disk Load for RAID 6 = 11200*0.6 + (**6***0.4*11200) = 33600 IOPS

ii) Number of drives required to support the application in different RAID environments, drives with a rating of 130 IOPS.

For RAID 1 = 15680 /130 = 121 disks
For RAID 5 = 24640 /130 = 190 disks
For RAID 6 = 33600 /130 = 258 disks

3) Consider an application that generates 5,200 IOPS, with 60 percent of them being reads.
i) Show the disk load in RAID 1, RAID 5 and RAID 6.
ii) If a HDD with the specification of maximum 180 IOPS for the application needs to be used, show the number of disks required to meet the workload for RAID 1, RAID 5 and RAID 6.
Ans) i)

Disk Load for RAID 1 = 5200*0.6 + (**2***0.4*5200) = 7280 IOPS

Disk Load for RAID 5 = 5200*0.6 + (**4***0.4*5200) = 11440 IOPS

Disk Load for RAID 6 = 5200*0.6 + (**6***0.4*5200) = 15600 IOPS

ii) Number of drives required to support the application in different RAID environments, drives with a rating of 180 IOPS.

For RAID 1 = 7280 /180 = 40 disks
For RAID 5 = 11440 /180 = 64 disks
For RAID 6 = 15600 /180 = 87 disks