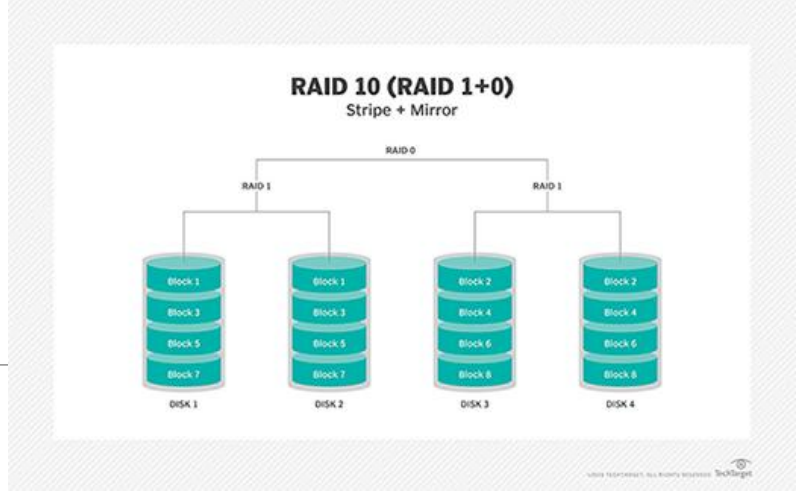
UNIT - 2

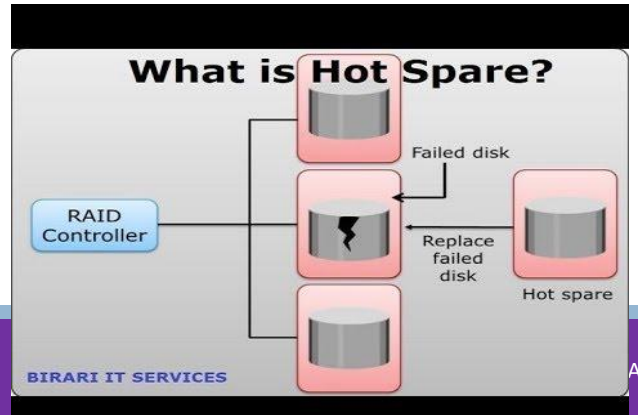
RAMAIAH
Institute of Technology

UNIT 2- Storage System



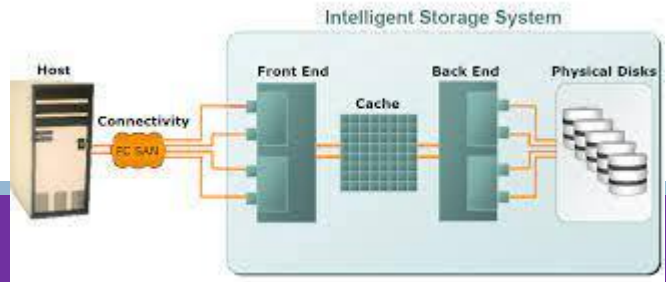
Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- Array Components,
- Techniques,
- Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares



Intelligent Storage System(Chapter 4)

- Components of an Intelligent Storage System
- Storage Provisioning
- Types of Intelligent Storage Systems



RAMAIAH
Institute of Technology

UNIT 2- Storage System

Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- Array Components,
- Techniques,
- Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares

RAID

Imagine you have a large collection of important books, and want to store them safely on bookshelves .

Store them in a way that balances

- **speed** (how quickly can access or store a book)
- **safety** (what happens if a shelf breaks)
- **efficiency** (how much shelf space is used)

RAMAIAH
Institute of Technology

Introduction

- RAID is a storage technology (Redundant Array of Independent Disks) or (Redundant array of Inexpensive disks).
- RAID is a technique to store same data in different places on multiple hard disks or Solid state drives(SSDs) to protect data in the case of disk failure
- RAID is a data storage method **that integrates many disk drives into a single device** to increase performance and offer redundancy.

RAMAIAH
Institute of Technology

Introduction

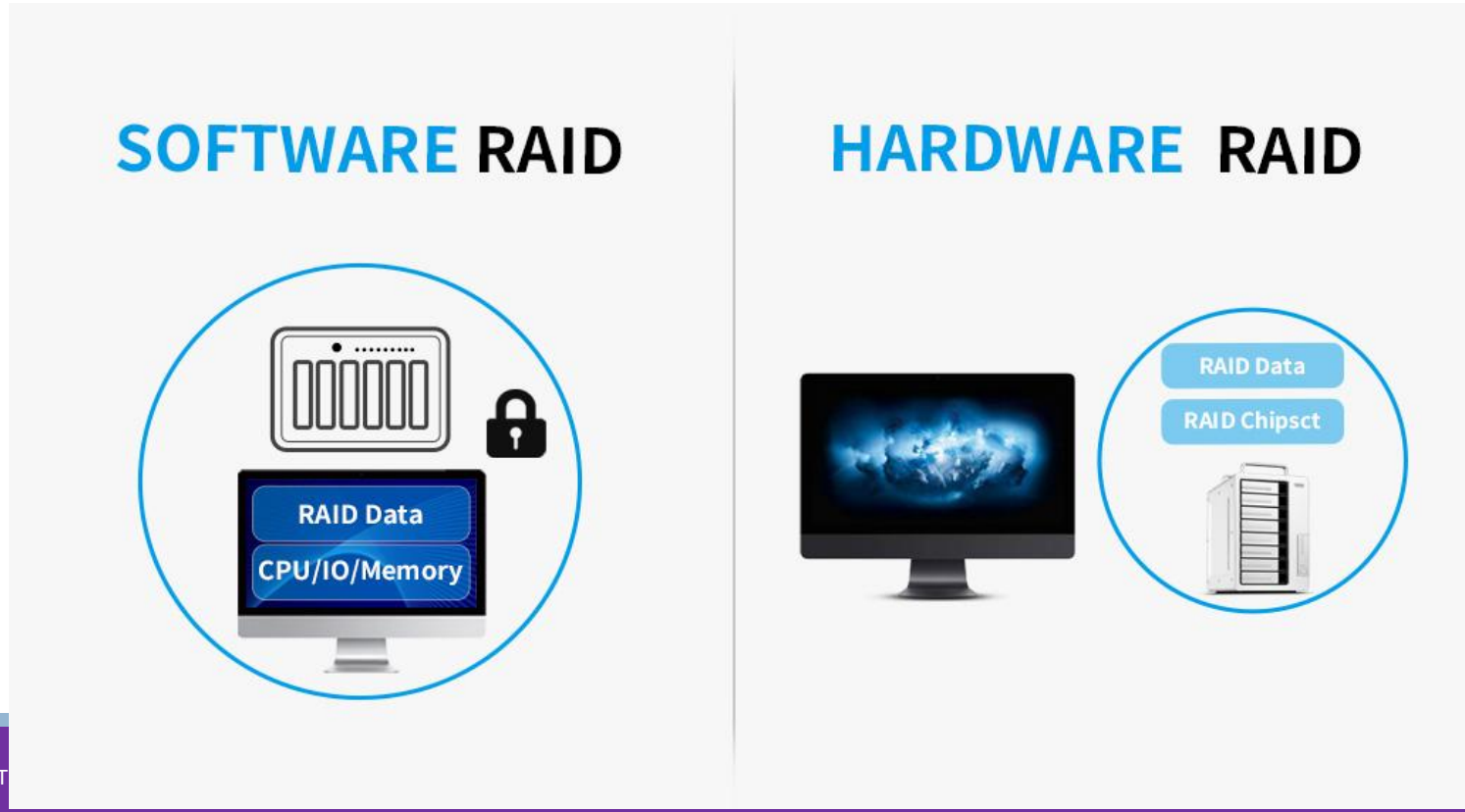
- **RAID** is the way of combining several small disks into a single storage of a large size. The disks included into the array are called **Array Members**.
- It can be used to store important documents, financial information, research data, and more.
- RAID storage is useful for many industries, including healthcare, education, manufacturing, and finance.

RAMAIAH
Institute of Technology

RAID Implementation Methods

The two methods of RAID implementation are hardware and software.

1. **Software RAID**
2. **Hardware RAID**



RAID Implementation Methods

Imagine a company working on a large project, and each team member is assigned to handle parts of the work.

Software RAID is like having the **employees manage their own work** without a dedicated project manager, but they still follow a well-structured plan to ensure that everything runs smoothly.

Hardware RAID is like hiring a dedicated **project manager** (a RAID controller) to manage the tasks, allocate work, and keep everything running smoothly.

RAID Implementation Methods

Software RAID

- Software RAID uses host-based software to provide RAID functions.
- It is implemented at the operating-system level
- Software RAID implementations offer cost and simplicity benefits when compared with hardware RAID.

Software RAID

1. It is PC bundled, only the PC with its RAID data can use this RAID.
2. It uses PC's computing resources.
3. It cannot survive from the PC failure; RAID data is on the PC.
4. It can only be used as secondary drive.
5. DRIVE/DATA 50% redundant and secured.
6. From cost aspect, it is cheaper.

RAID Implementation Methods

Software RAID - limitations:

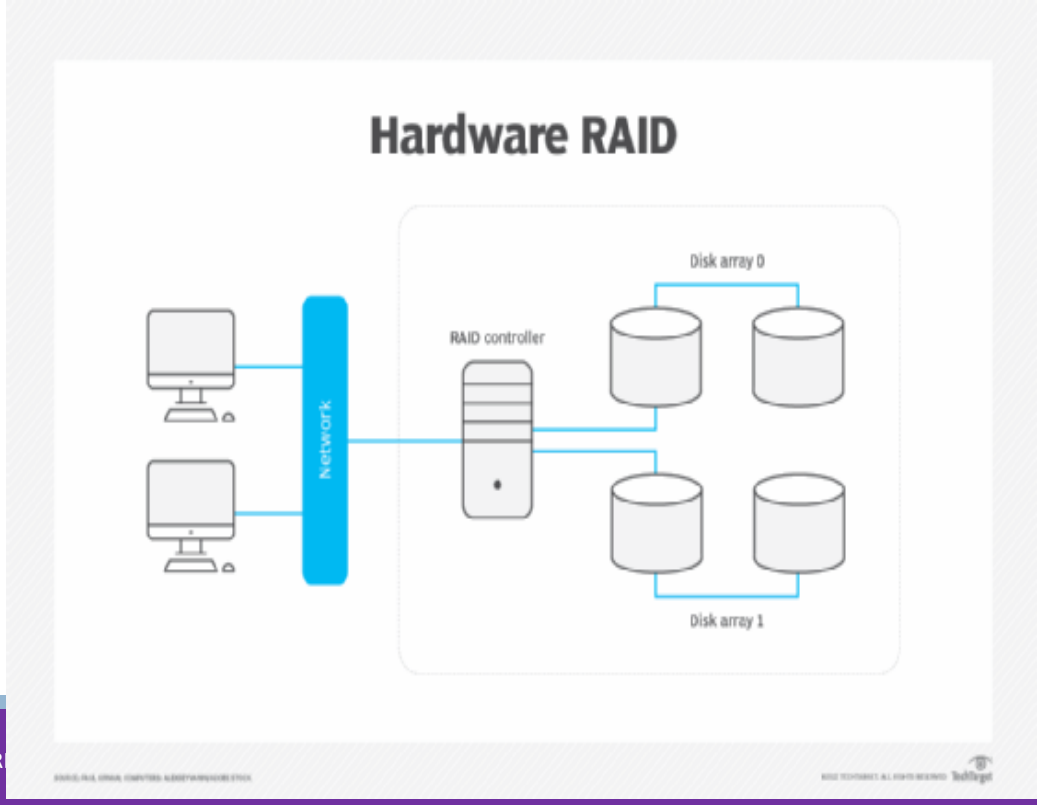
1. Performance: Software RAID affects overall system performance. This is due to additional CPU cycles required to perform RAID calculations.
2. Supported features: Software RAID does not support all RAID levels.
3. Operating system compatibility: Software RAID is tied to the host operating system; hence, upgrades to software RAID or to the operating system should be validated for compatibility. This leads to inflexibility in the data-processing environment.

RAMAIAH
Institute of Technology

RAID Implementation Methods

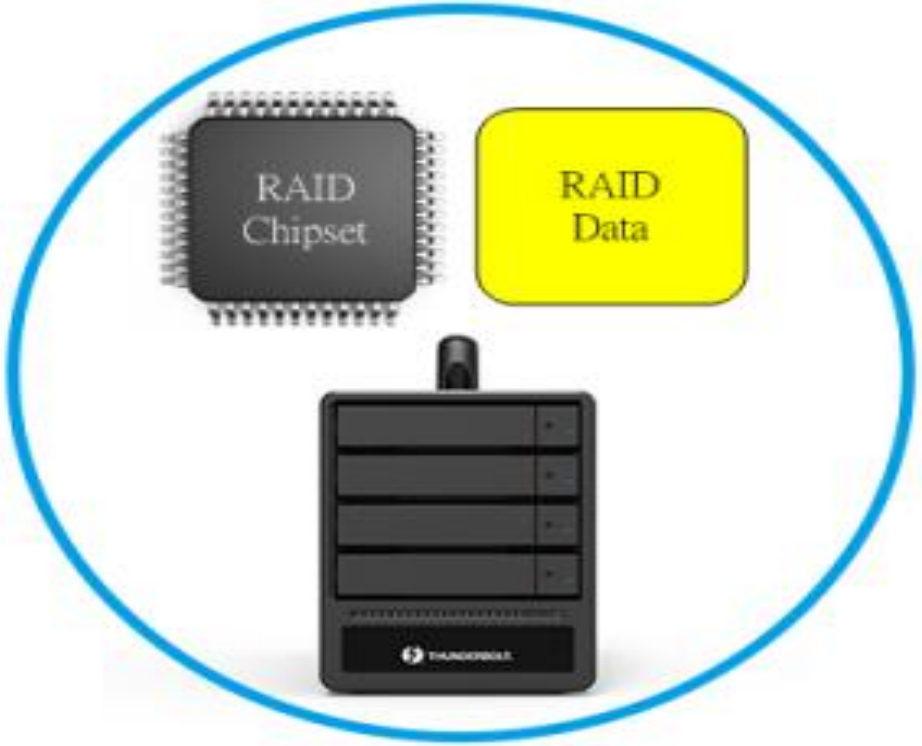
Hardware RAID

- Specialized hardware controller is implemented either on the host or on the array.
- Controller card RAID is a host-based hardware RAID implementation in which a specialized RAID controller is installed in the host, and disk drives are connected to it.
- Manufacturers also integrate RAID controllers on motherboards.



Hardware RAID

1. It works independently, can be connected to any PC.
2. It does not consume PC's computing resources.
3. It can survive from the PC failure; RAID data is on the hardware RAID.
4. It can be used as the primary booting drive.
5. DRIVE/DATA fully redundant and secured.
6. From cost aspect, it costs more



RAMAIAH
Institute of Technology

RAID Implementation Methods

Hardware RAID

- A host-based RAID controller is not an efficient solution in a data center environment with a large number of hosts.
- The external RAID controller is an array-based hardware RAID.
- It acts as an interface between the host and disks.
- It presents storage volumes to the host, and the host manages these volumes as physical drives.

RAMAIAH
Institute of Technology

RAID Implementation Methods

Hardware RAID

The key functions of the RAID controllers are as follows:

- Management and control of disk aggregations
- Translation of I/O requests between logical disks and physical disks
- Data regeneration in the event of disk failures

RAID Implementation Methods

Hardware RAID

- A host-based RAID controller is not an efficient solution in a data center environment with a large number of hosts.
- The external RAID controller is an array-based hardware RAID.
- It acts as an interface between the host and disks.
- It presents storage volumes to the host, and the host manages these volumes as physical drives.

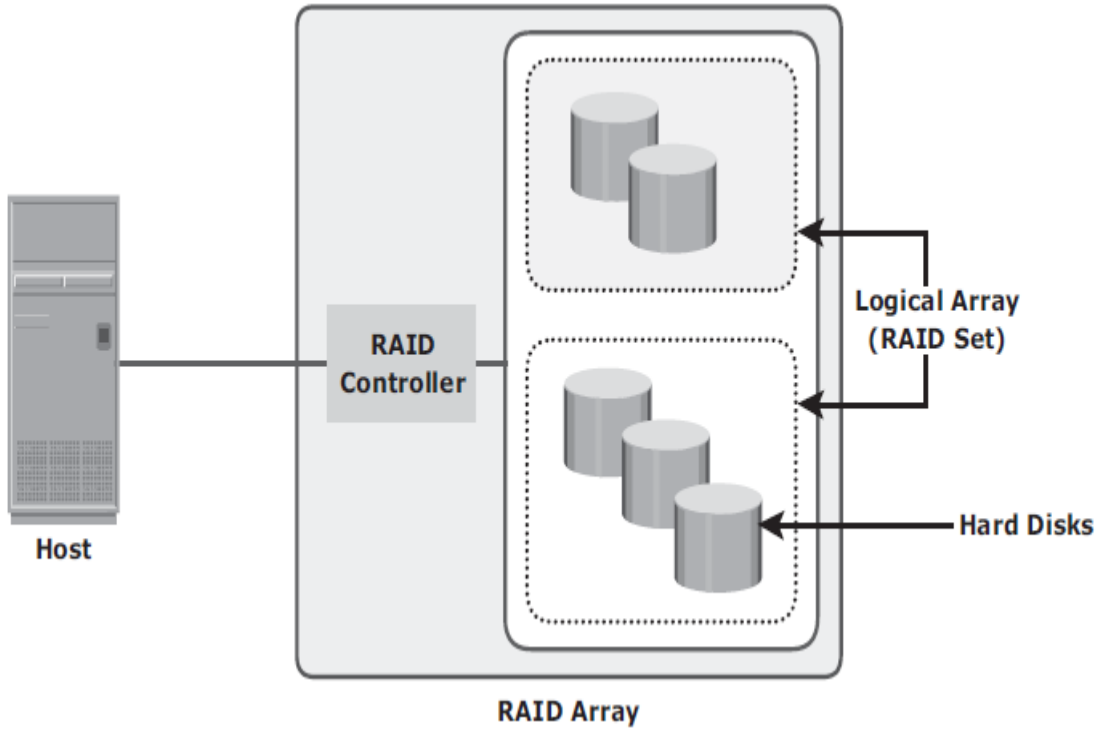
UNIT 2- Storage System

Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- RAID Array Components,
- Techniques,
- Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares

RAID Array Components

- A RAID array is an enclosure that contains a number of disk drives and supporting hardware to implement RAID.
- A subset of disks within a RAID array can be grouped to form logical associations called logical arrays(RAID set)(RAID group)



UNIT 2- Storage System

Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- RAID Array Components,
- RAID Techniques,
- Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares

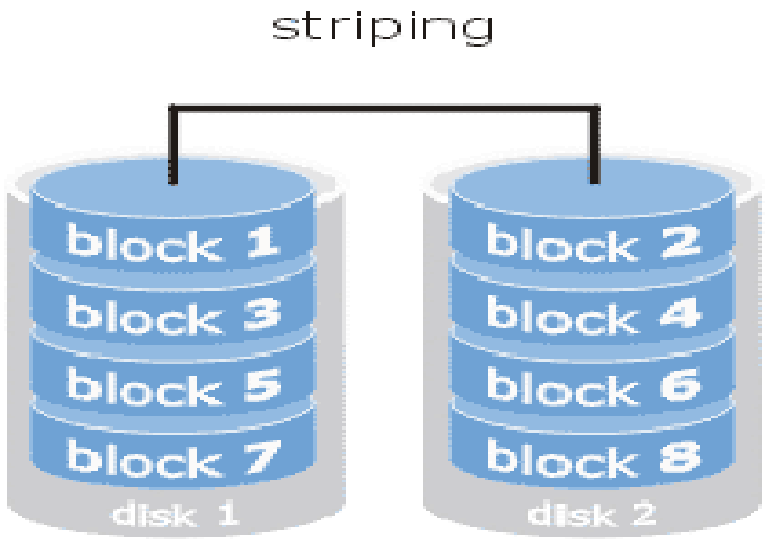
RAID Techniques

RAID techniques determine the data availability and performance characteristics of a RAID set.

1. Striping
2. Mirroring
3. Parity

RAID Techniques - Striping

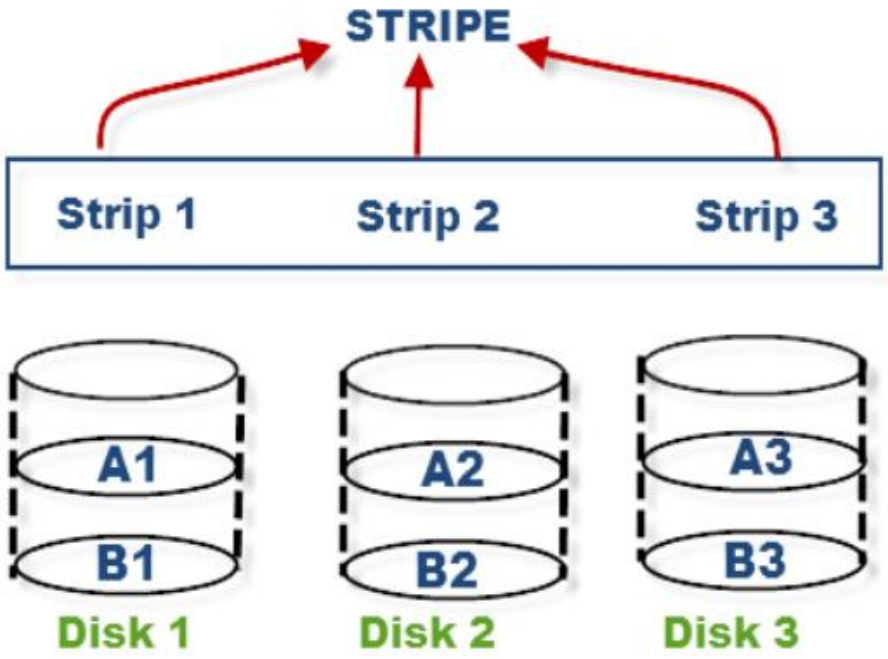
- **Striping** is the technique of segmenting logically sequential data, so that consecutive segments are stored on different physical storage devices.
- Striping is a technique to spread data across multiple drives (more than one) to use the drives in parallel.

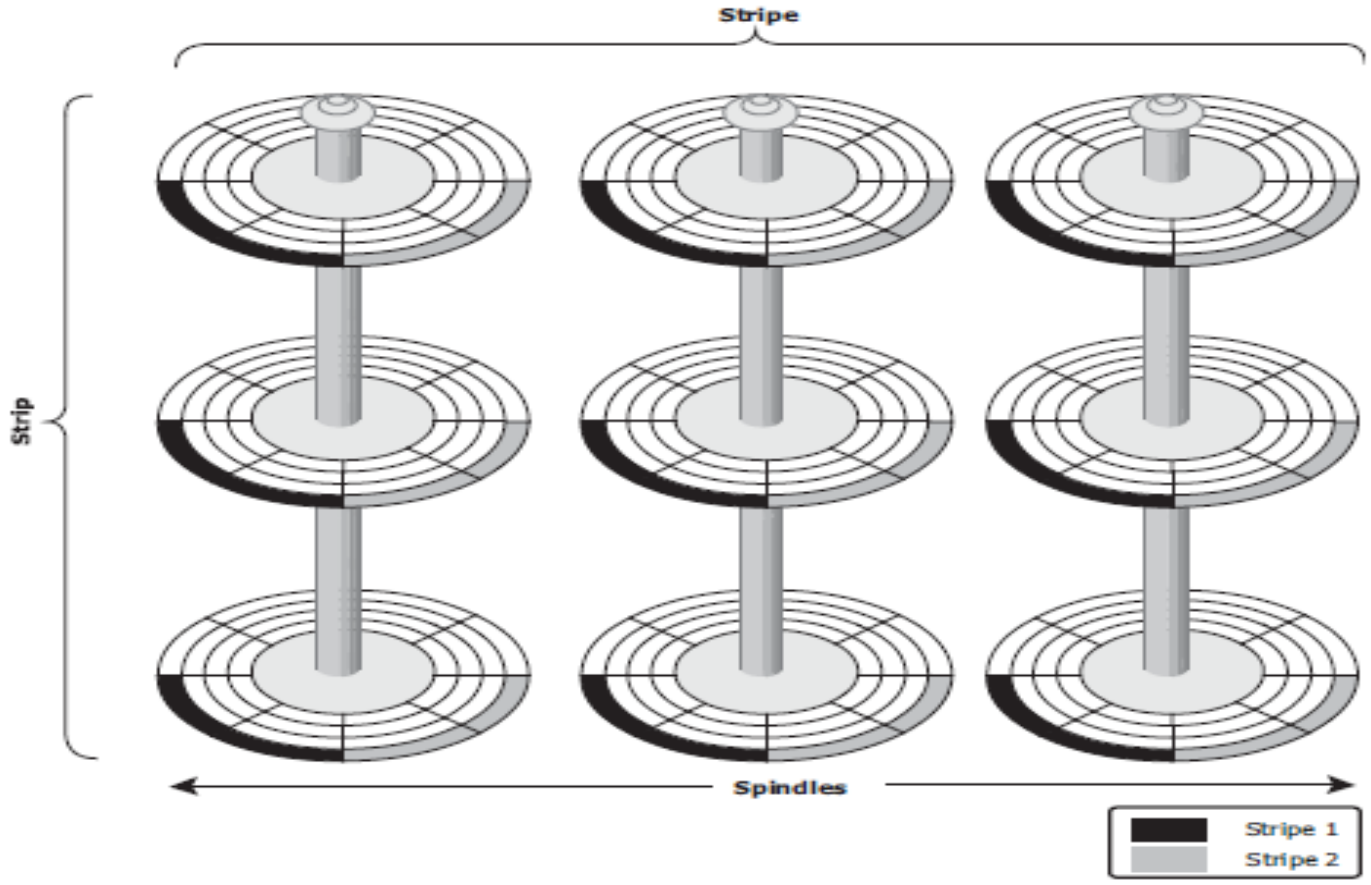


RAMAIAH
Institute of Technology

RAID Techniques - Striping

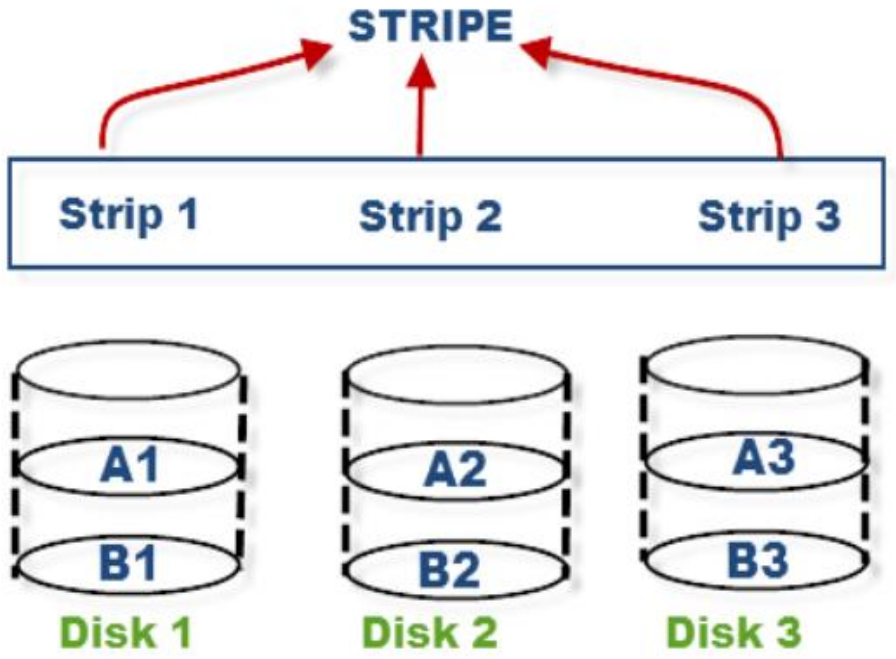
Figure shows physical and logical representations of a striped RAID set.





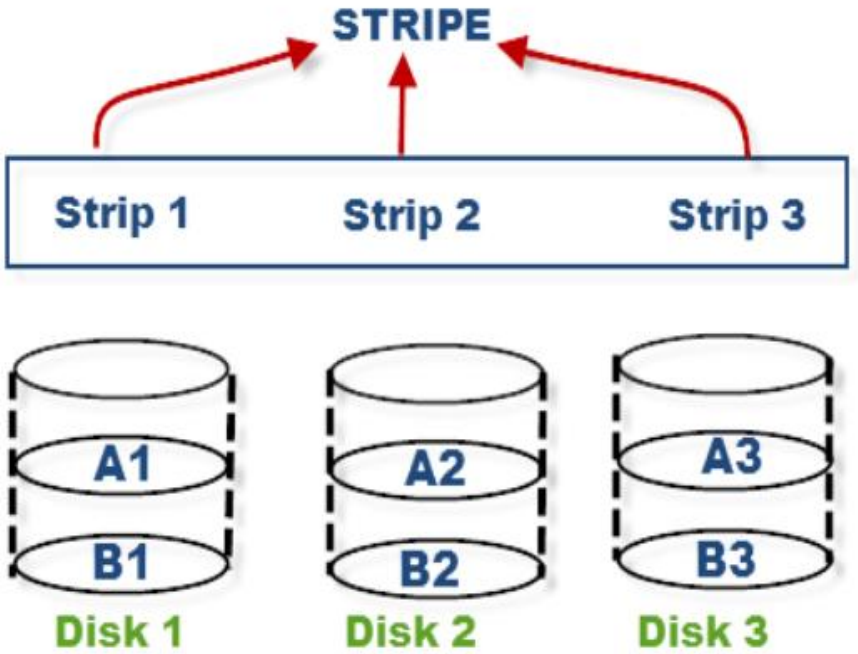
RAID Techniques - Striping

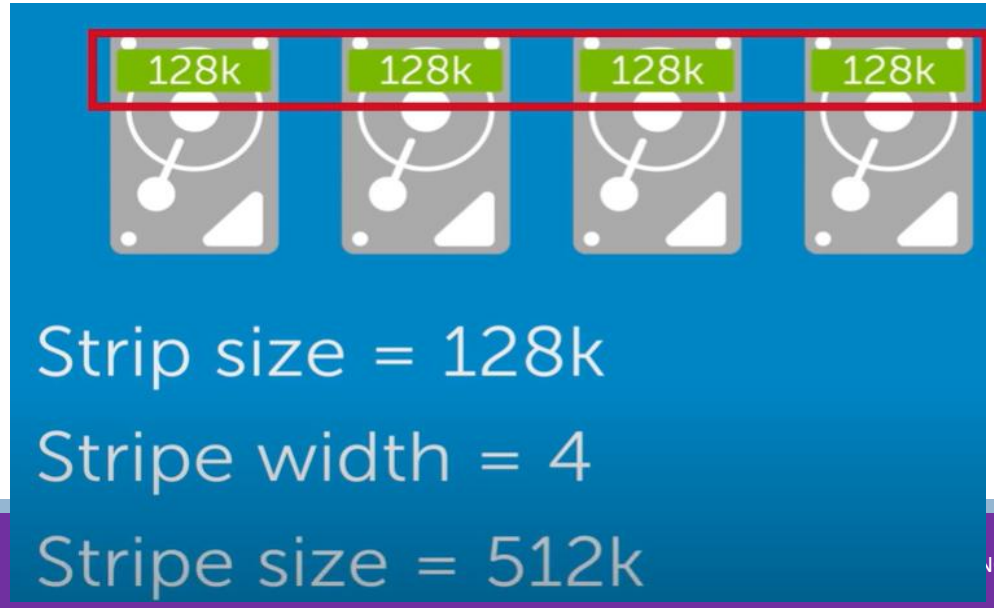
- Within each disk in a RAID set, a predefined number of contiguously addressable disk blocks are defined as a strip.
- The set of aligned strips that spans across all the disks within the RAID set is called a stripe.



RAID Techniques - Striping

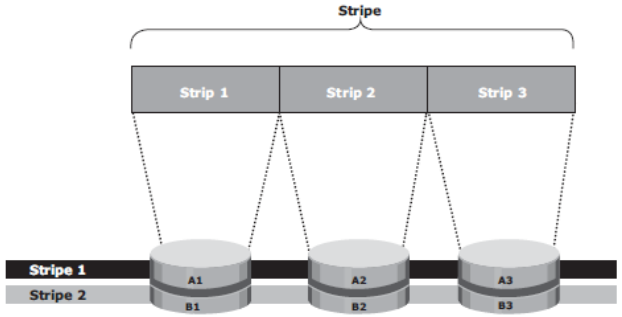
Strip size (also called stripe depth) describes the number of blocks in a strip and is the maximum amount of data that can be written to or read from a single disk in the set, assuming that the accessed data starts at the beginning of the strip.





RAMAIAH
Institute of Technology

RAID Techniques - Striping



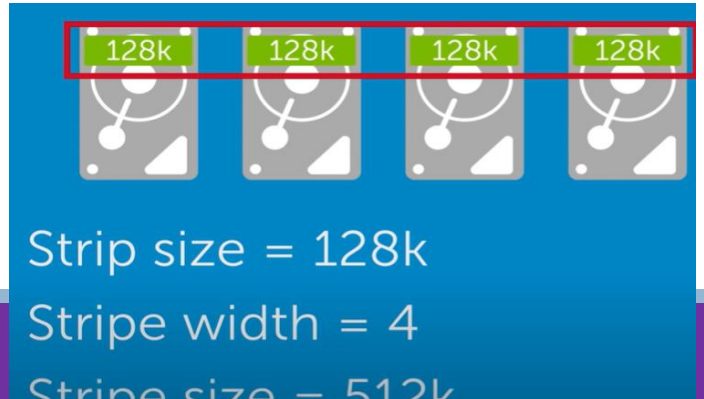
- Stripe size is a multiple of strip size by the number of data disks in the RAID set.
- Stripe width refers to the number of data strips in a stripe.

Example:

Five disk striped RAID with a strip size of 64 KB, the stripe size is

= 64KB * 5

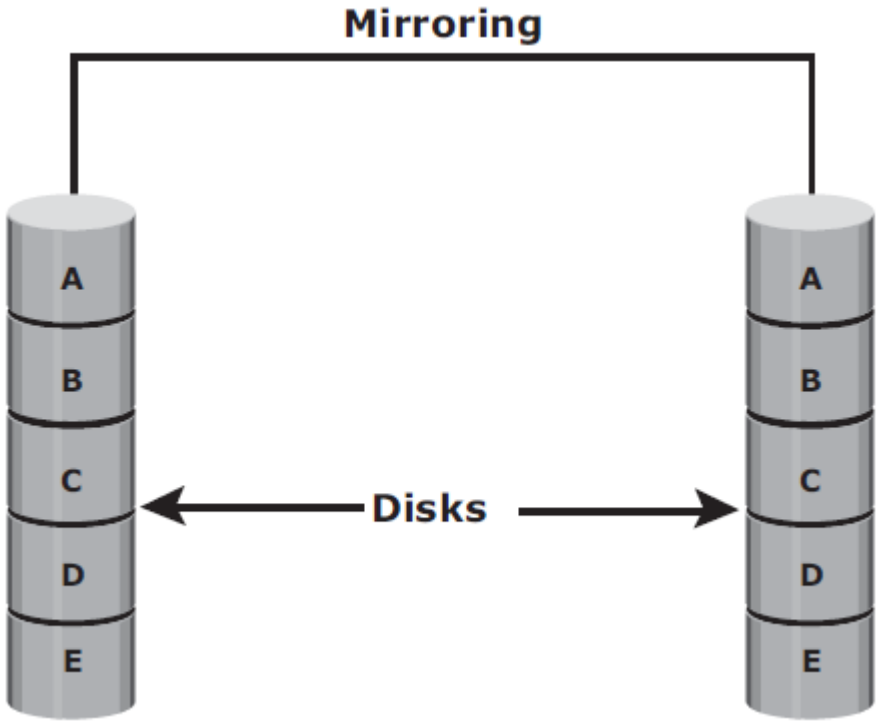
= 320 KB



RAID Techniques - Mirroring

Mirroring is a technique whereby the same data is stored on two different disk drives, yielding two copies of the data.

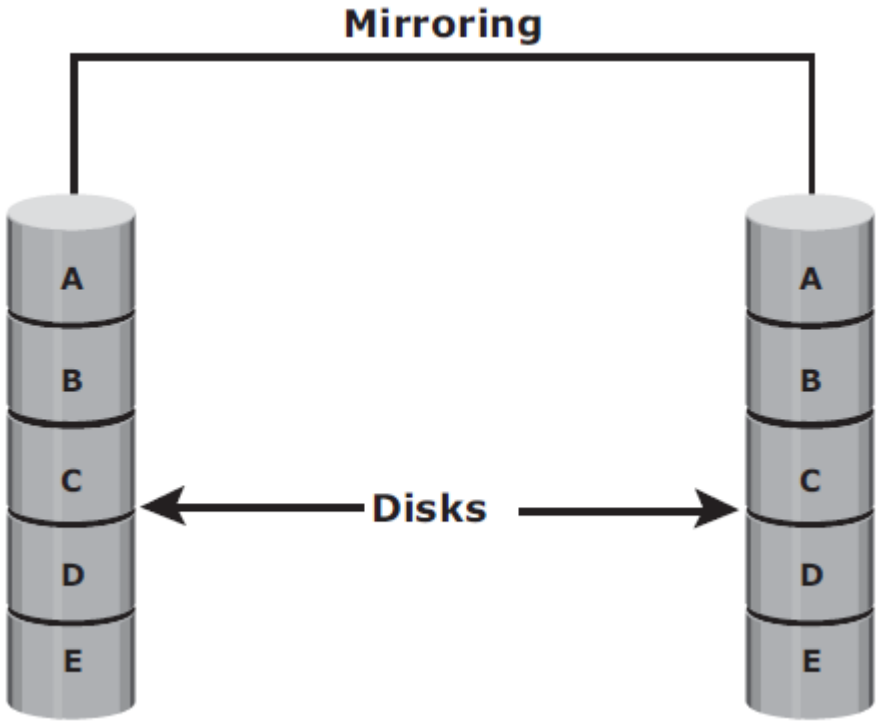
If one disk drive failure occurs, the data is intact on the surviving disk drive and the controller continues to service the host's data requests from the surviving disk of a mirrored pair.



RAID Techniques - Mirroring

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.



RAID Techniques - Mirroring

Advantages

- Provide complete data redundancy
- It enables fast recovery from disk failure.
- Mirroring improves read performance because read requests can be serviced by both disks.

RAMAIAH Institute of Technology

RAID Techniques - Mirroring

Disadvantages

- It only provides data protection and is not a substitute for data backup.
- Mirroring involves duplication of data — the amount of storage capacity needed is twice the amount of data being stored.
- Is expensive and is preferred for mission-critical applications that cannot afford the risk of any data loss.
- Write performance is slightly lower than that in a single disk because each write request manifests as two writes on the disk drives.
- Mirroring does not deliver the same levels of write performance as a striped RAID.

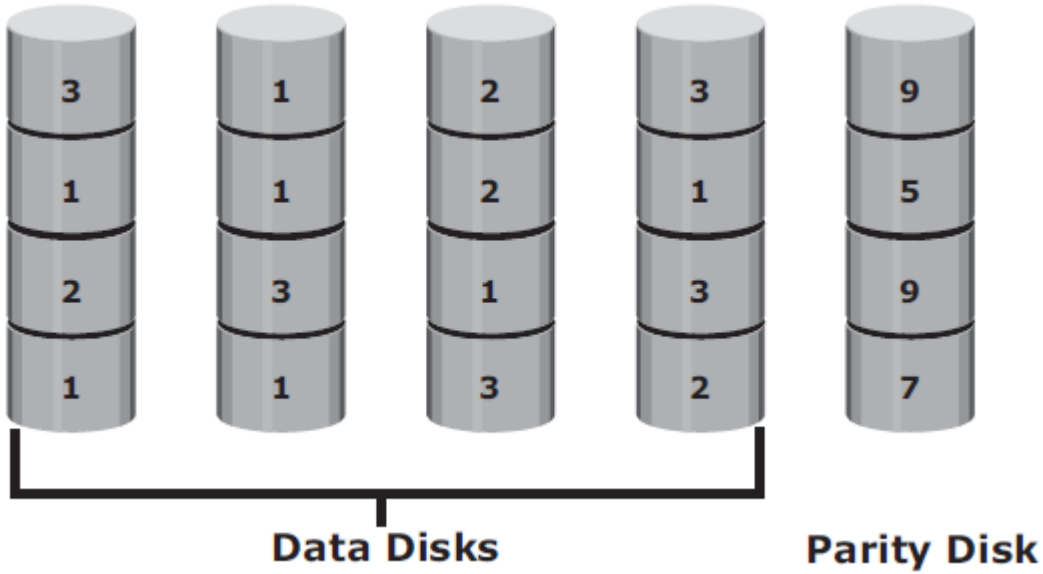
RAID Techniques - Parity

- Mirroring involves high cost, so to protect the data new technique is used with striping called parity.
- Parity is a redundancy technique that ensures protection of data without maintaining a full set of duplicate data.
- Parity is information that is used to rebuild the data in the event of a disk failure
- An additional disk drive is added to hold parity

RAID Techniques - Parity

Parity information can be stored on separate, dedicated disk drives or distributed across all the drives in a RAID set.

Figure shows a parity RAID set.

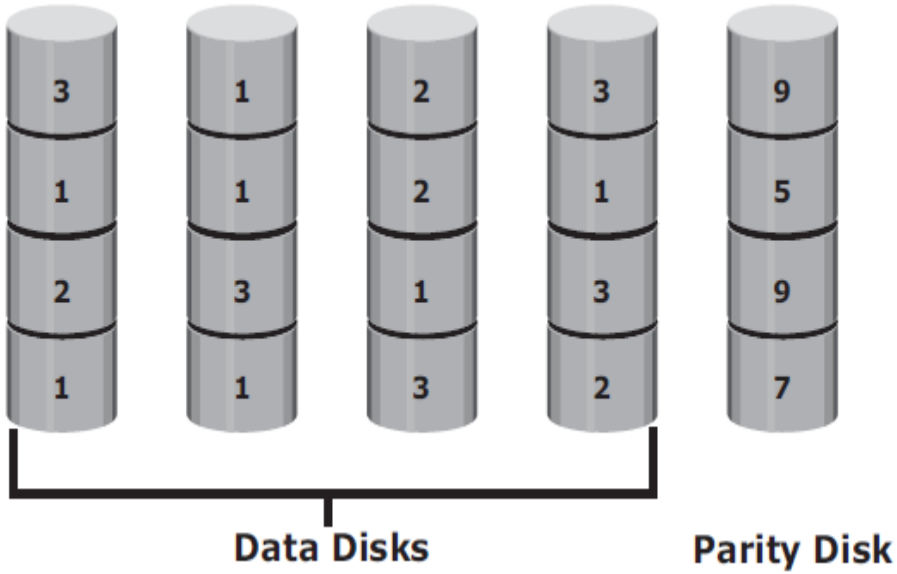


RAID Techniques - Parity

The first four disks, labeled "Data Disks," contain the data.

The fifth disk, labeled "Parity Disk," stores the parity information, which, in this case, is the sum of the elements in each row.

If one of the data disks fails, the missing value can be calculated by subtracting the sum of the rest of the elements from the parity value.



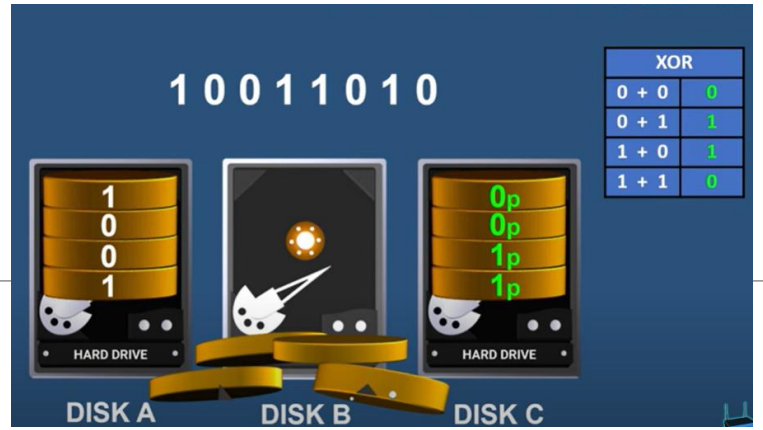
RAID Techniques - Parity

Parity calculation is a bitwise XOR operation.

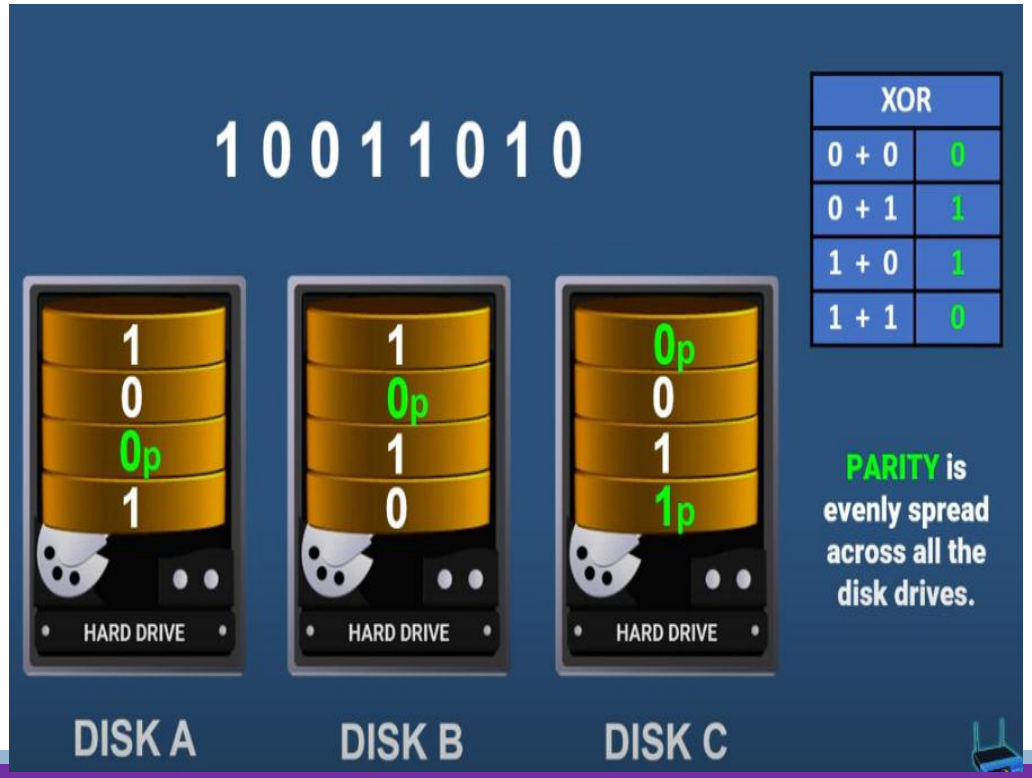
- A and B denote the inputs and C, the output after performing the XOR operation.
- If any of the data from A, B, or C is lost, it can be reproduced by performing an XOR operation on the remaining available data.
- For example, if a disk containing all the data from A fails, the data can be regenerated by performing an XOR between B and C.

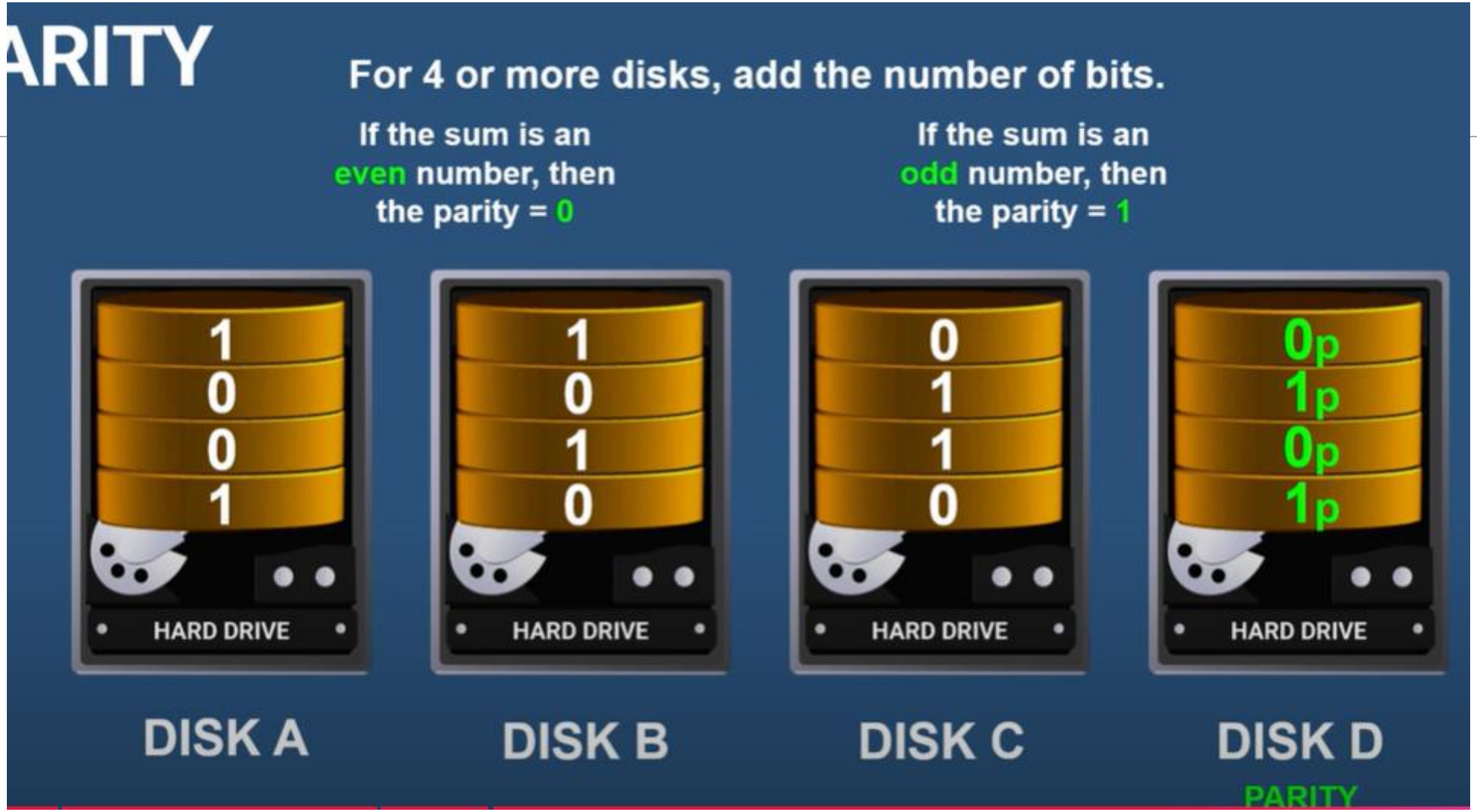
A	B	C
0	0	0
0	1	1
1	0	1
1	1	0


1 0 0 1 1 0 1 0
XOR
0 + 0 0
0 + 1 1
1 + 0 1
1 + 1 0
1
0
0
1
HARD DRIVE
DISK A
1
0
1
0
HARD DRIVE
DISK B
0p
0p
1p
1p
HARD DRIVE
DISK C


1 0 0 1 1 0 1 0
XOR
0 + 0 0
0 + 1 1
1 + 0 1
1 + 1 0
1
0
0
1
0p
0p
1p
1p
HARD DRIVE
HARD DRIVE
DISK A
DISK B
DISK C


1 0 0 1 1 0 1 0
XOR
0 + 0 0
0 + 1 1
1 + 0 1
1 + 1 0
1
0
0
1
1
0
1
0p
0p
1p
1p
HARD DRIVE
HARD DRIVE
HARD DRIVE
DISK A
DISK B
DISK C


1 0 0 1 1 0 1 0
XOR
0 + 0 0
0 + 1 1
1 + 0 1
1 + 1 0
1
0
0p
1
1
0p
1
0
0p
0
1
1p
HARD DRIVE
HARD DRIVE
HARD DRIVE
PARITY is evenly spread across all the disk drives.
DISK A
DISK B
DISK C


ARITY
For 4 or more disks, add the number of bits.
If the sum is an even number, then the parity = 0
If the sum is an odd number, then the parity = 1
1
0
0
1
HARD DRIVE
DISK A
1
0
1
0
HARD DRIVE
DISK B
0
1
1
0
HARD DRIVE
DISK C
0p
1p
0p
1p
HARD DRIVE
DISK D
PARITY

RAID Techniques - Parity

Advantage

Parity implementation considerably reduces the cost associated with data protection.

Disadvantages of using parity.

Parity information is generated from data on the data disk. Therefore, parity is recalculated every time there is a change in data. This recalculation is time-consuming and affects the performance of the RAID array.

RAID Techniques - Parity

For parity RAID, the stripe size calculation does not include the parity strip.

For example in a five (4 + 1) disk parity RAID set with a strip size of 64 KB, the stripe size will be = 64 KB * 4

= 256 KB

UNIT 2- Storage System

Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- RAID Array Components,
- RAID Techniques,
- RAID Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares

RAMAIAH Institute of Technology

RAID Levels

Compare the common RAID types — RAID 0, RAID 1, RAID 3, RAID 4, RAID 5, R and RAID 1+0/0+1 — based on the following parameters:

- Minimum number of disks required
- Storage efficiency
- Read and write performance
- Write penalty
- Data protection

Table shows the commonly used RAID levels.

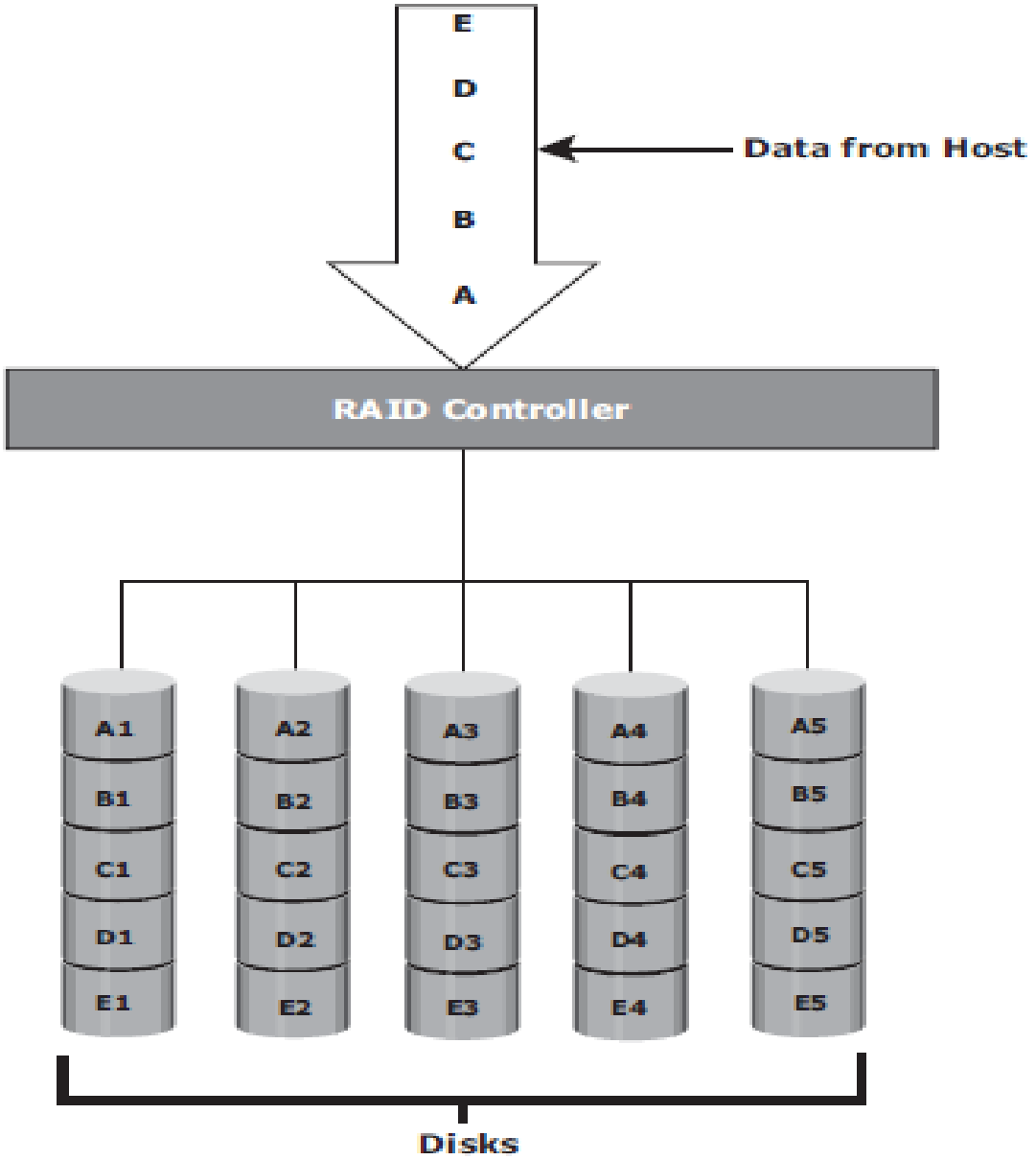
- RAID levels are defined on the basis of striping, mirroring, and parity techniques.
- Some RAID levels use a single technique, whereas others use a combination of techniques.

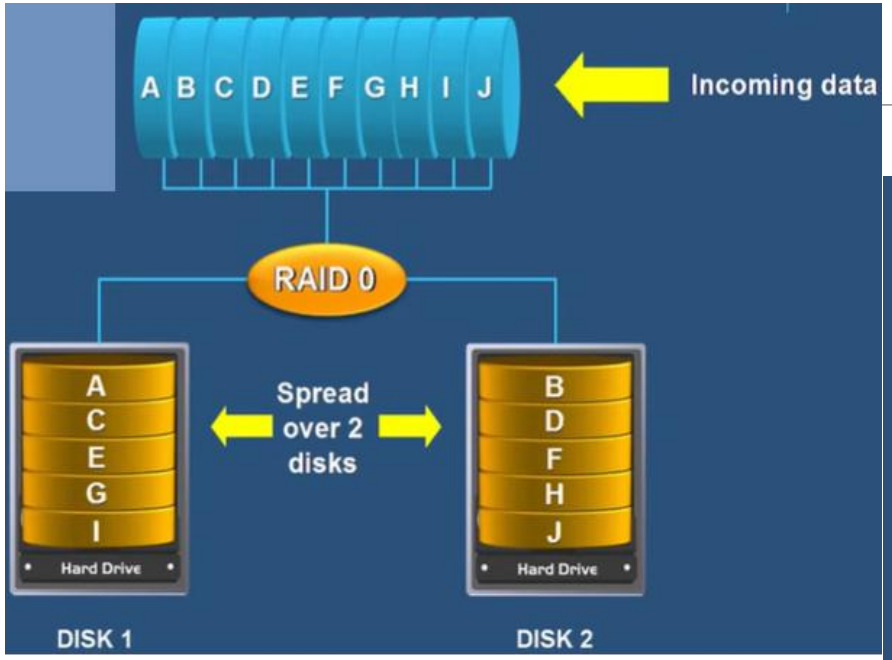
LEVELS	BRIEF DESCRIPTION
RAID 0	Striped set with no fault tolerance
RAID 1	Disk mirroring
Nested	Combinations of RAID levels. Example: RAID 1 + RAID 0
RAID 3	Striped set with parallel access and a dedicated parity disk
RAID 4	Striped set with independent disk access and a dedicated parity disk
RAID 5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

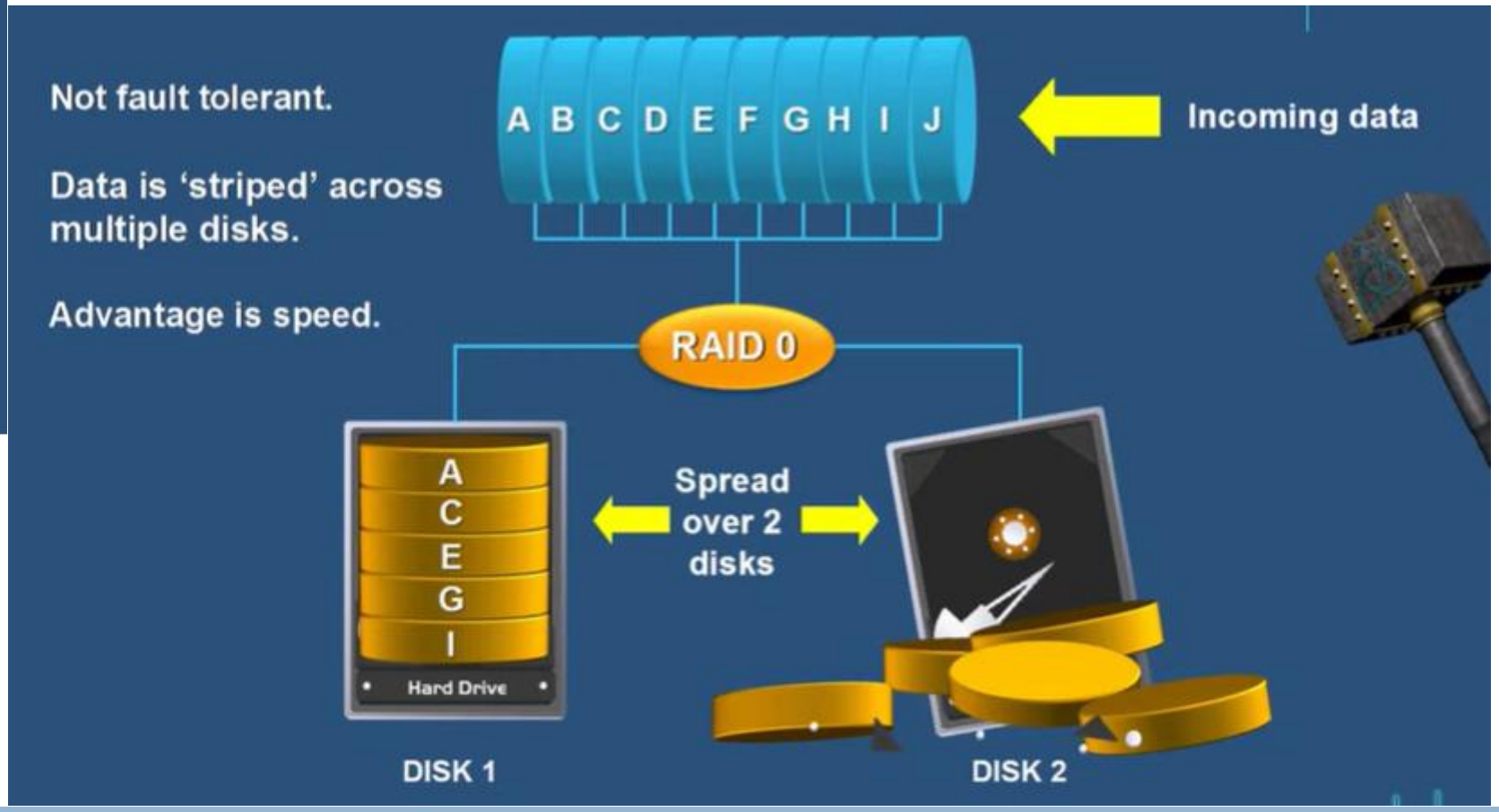
RAMAIAH
Institute of Technology

RAID Levels - 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.




A B C D E F G H I J
Incoming data
RAID 0
A
C
E
G
I
Hard Drive
Spread
over 2
disks
B
D
F
H
J
Hard Drive
DISK 1
DISK 2


Not fault tolerant.
Data is 'striped' across multiple disks.
Advantage is speed.
A B C D E F G H I J
Incoming data
RAID 0
A
C
E
G
I
Hard Drive
Spread
over 2
disks
DISK 1
DISK 2

RAID 0

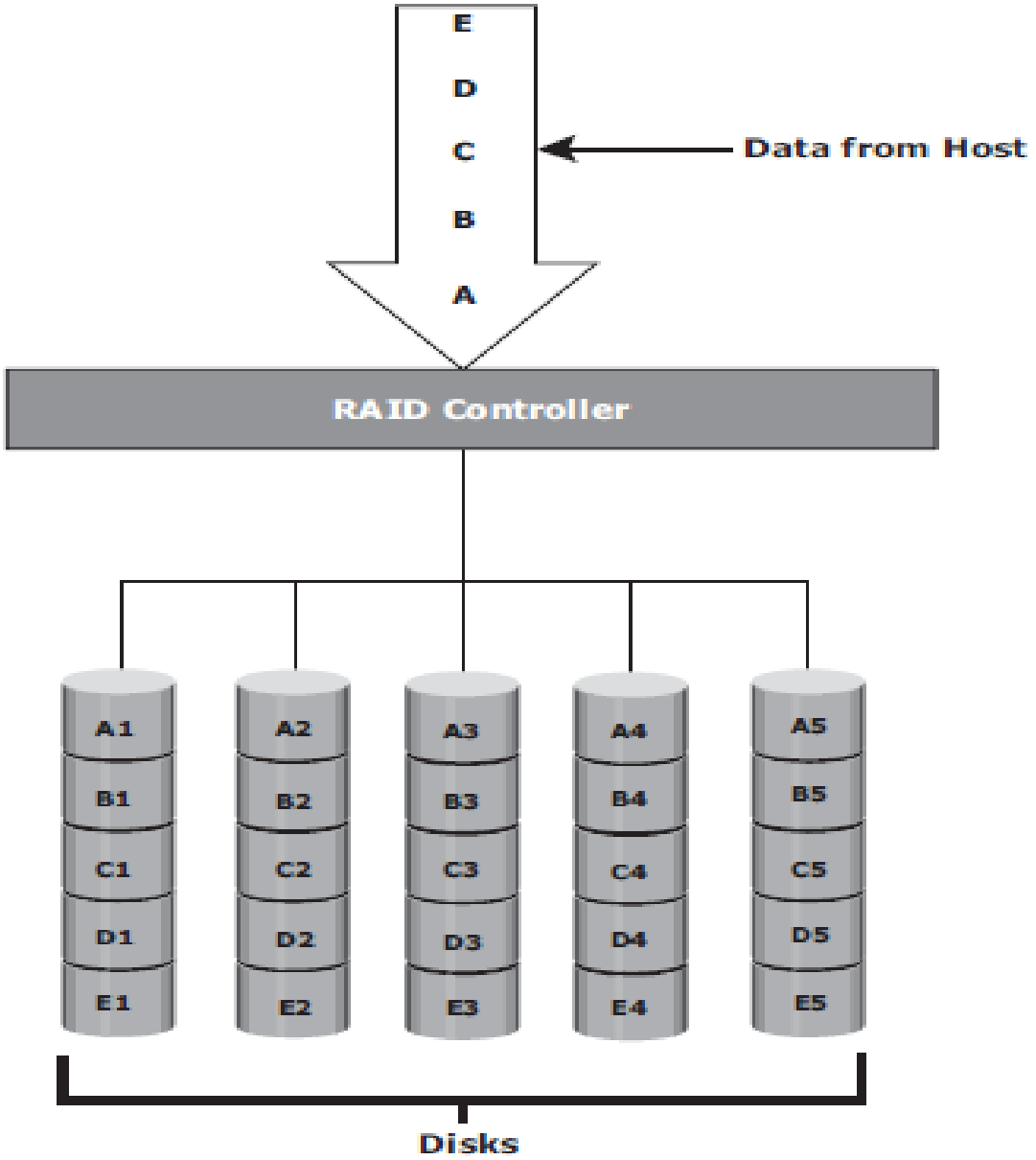
- Advantages
- Disk Utilization is 100%
- It will increase speeds for reading and writing from multiple disks at a time.
- Drawbacks
- Data will be lost, if there is any failure in disk.
- Because data is NOT redundant.

👉 Live streaming
👉 IPTV
👉 VOD Edge Server

RAMAIAH
Institute of Technology

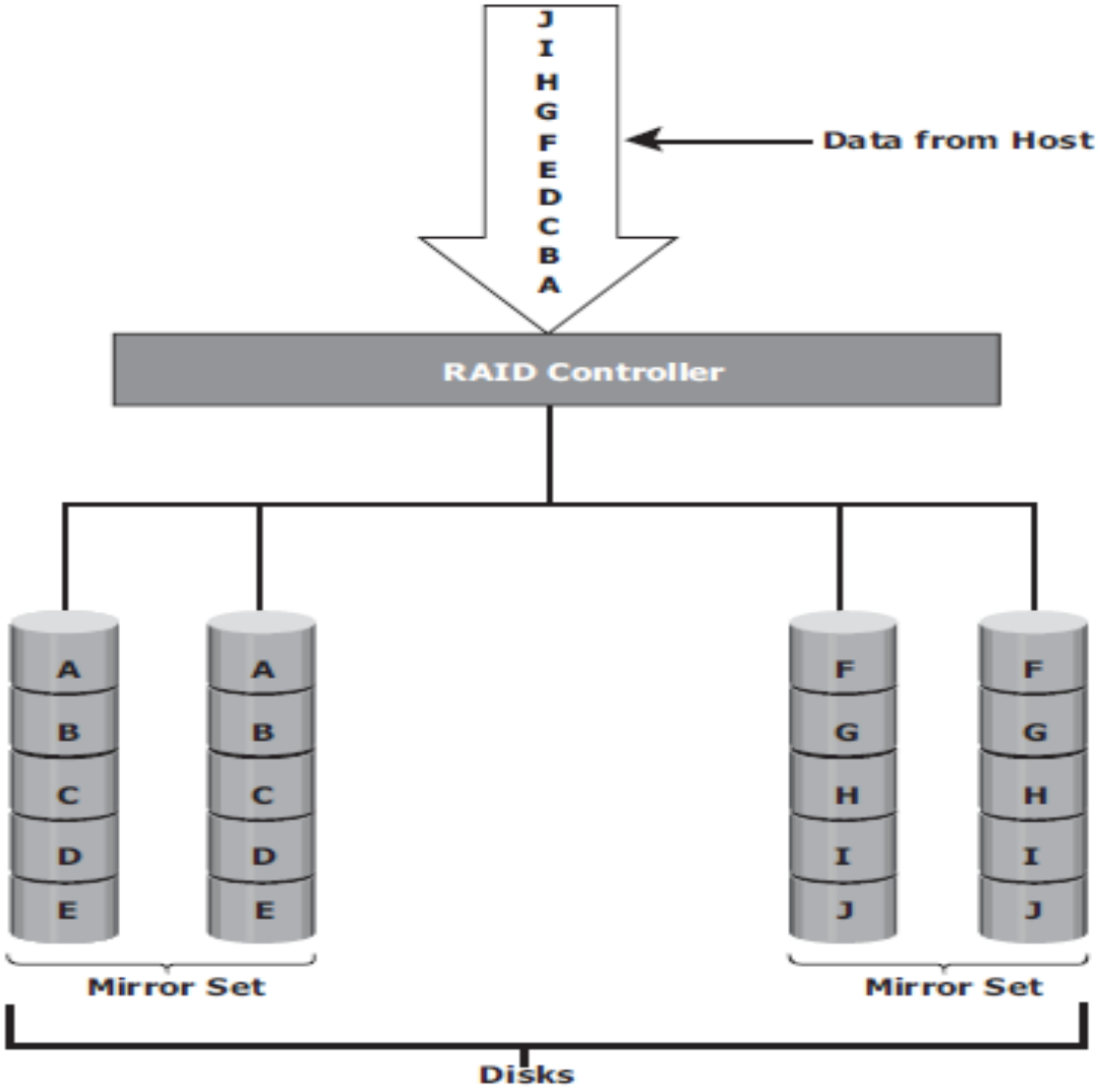
RAID Levels - RAID 0

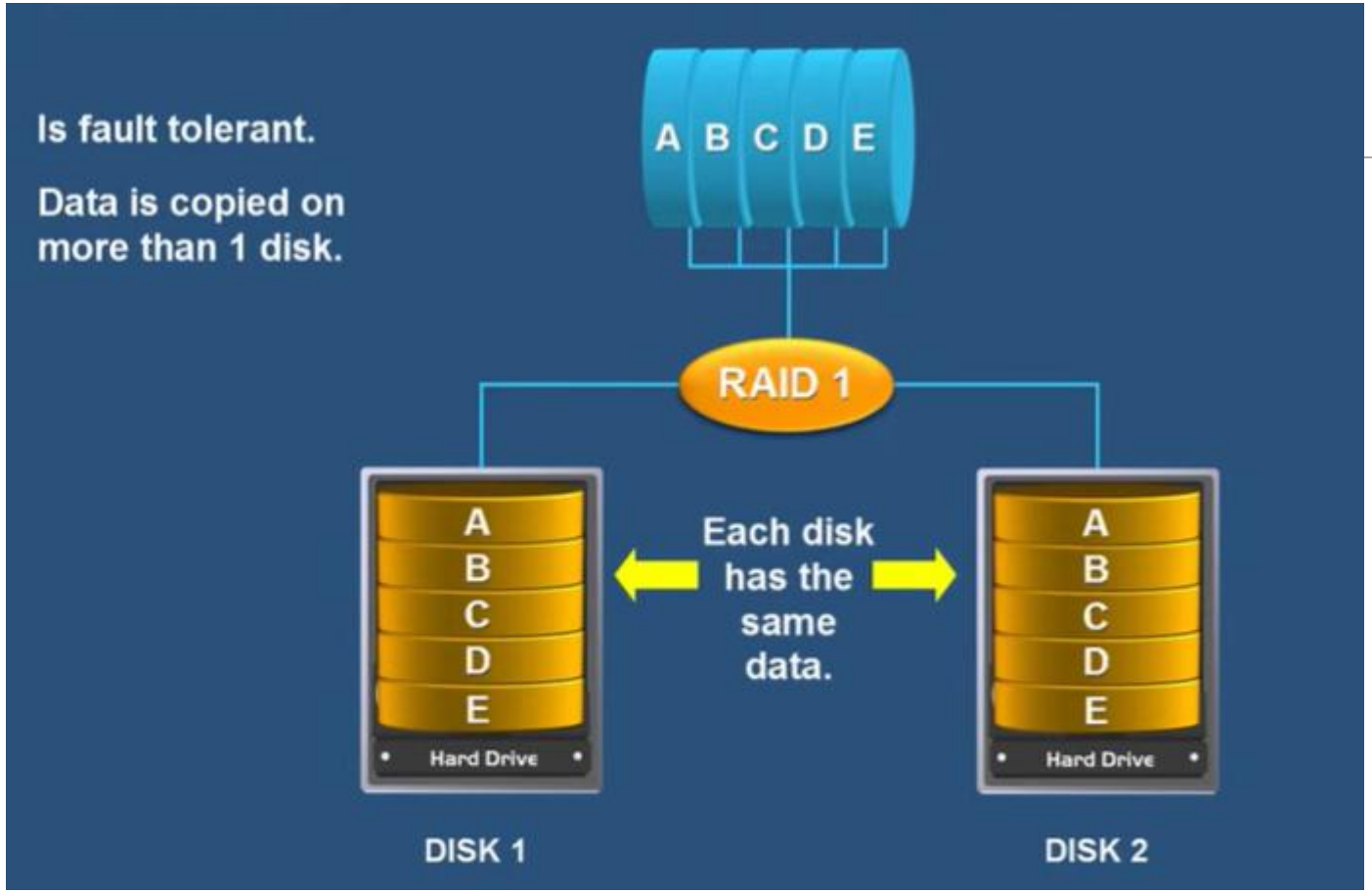
- When the number of drives in the RAID set increases, performance improves because more data can be read or written simultaneously.
- RAID 0 is a good option for applications that need high I/O throughput.
- However, if these applications require high availability during drive failures, RAID 0 does not provide data protection and availability.

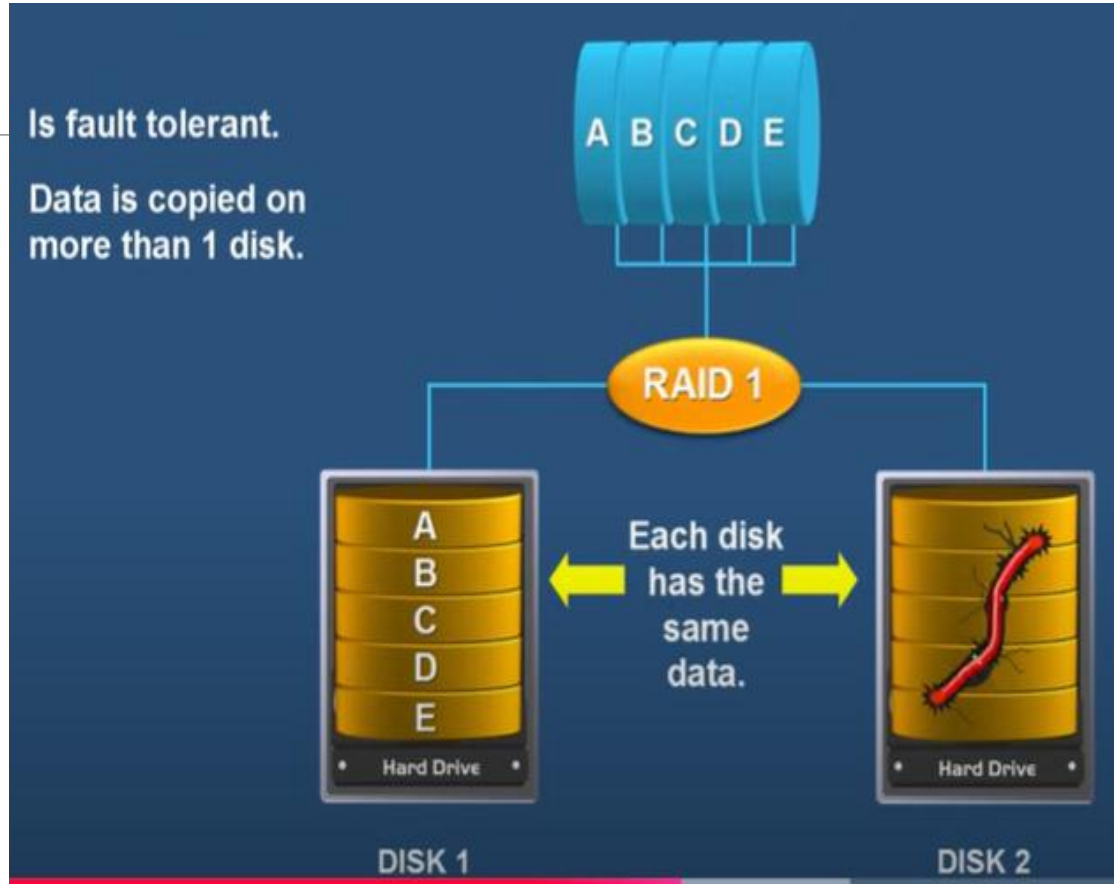


RAID Levels - RAID 1

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

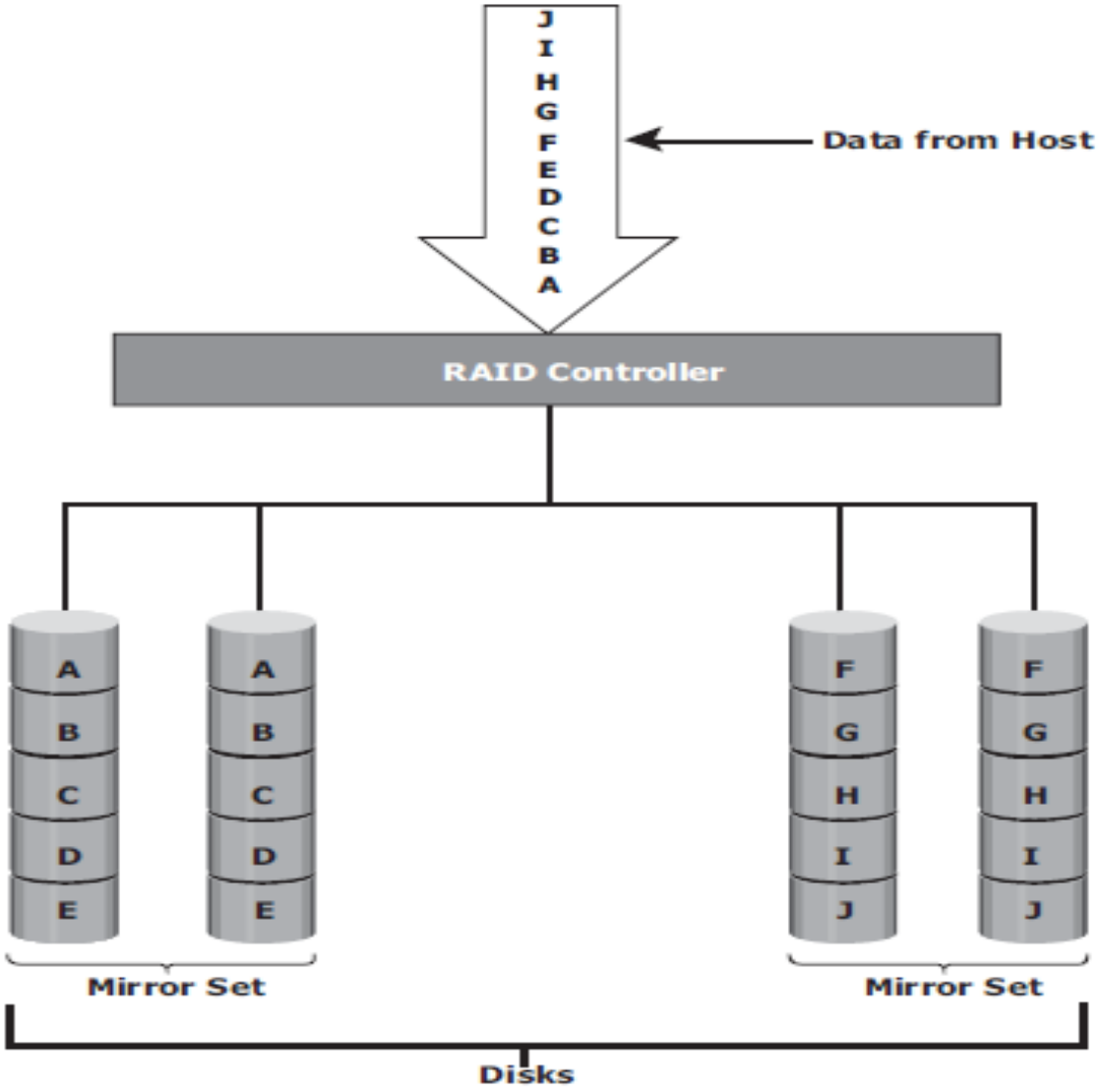



Is fault tolerant.
Data is copied on more than 1 disk.
A B C D E
RAID 1
A
B
C
D
E
Hard Drive
Each disk has the same data.
A
B
C
D
E
Hard Drive
DISK 1
DISK 2


Is fault tolerant.
Data is copied on more than 1 disk.
A B C D E
RAID 1
A
B
C
D
E
Hard Drive
Each disk has the same data.
Hard Drive
DISK 1
DISK 2

RAID Levels - RAID 1

- 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.



RAID 1

- Advantages
- Read performance is improved, since either disk can be read at the same time.
- If one disk fails, the data will retrieve from another disk.
- Write performance is the same as for single disk storage.
- Disadvantages
- Disk utilization got reduced to 50%

Ideal for mission critical storage
e.g accounting systems

It is also suitable for small servers in which only two data drives will be used

RAID Levels - Nested RAID

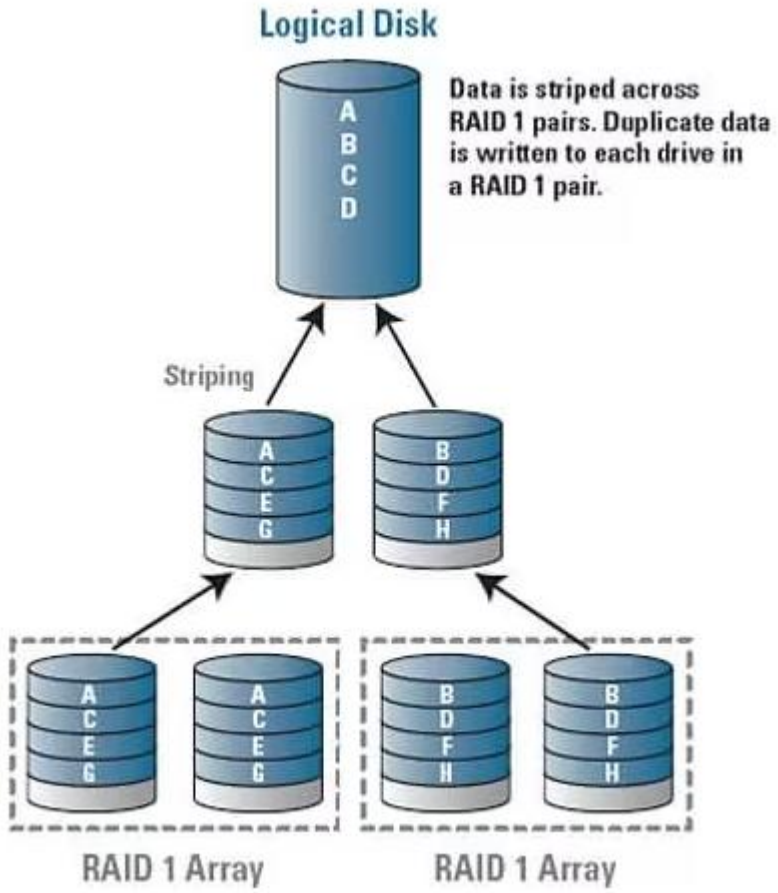
- RAID 1+0 and RAID 0+1 combine the performance benefits of RAID 0 with the redundancy benefits of RAID 1.
- They use striping and mirroring techniques and combine their benefits.
- These types of RAID require an even number of disks, the minimum being four

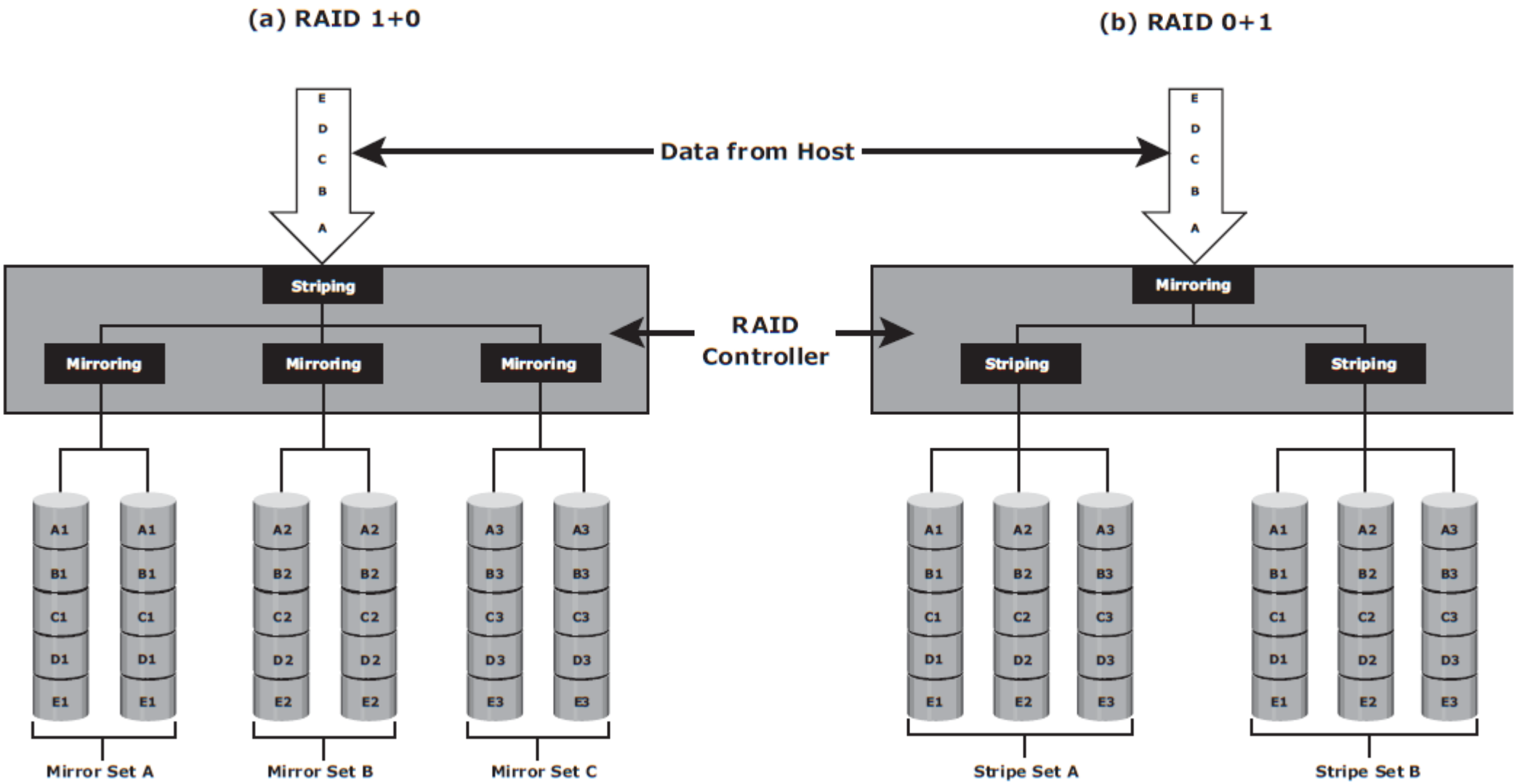
RAID 1+0 is also known as RAID 10 (Ten) or RAID 1/0.

RAID 0+1 is also known as RAID 01 or RAID 0/1.

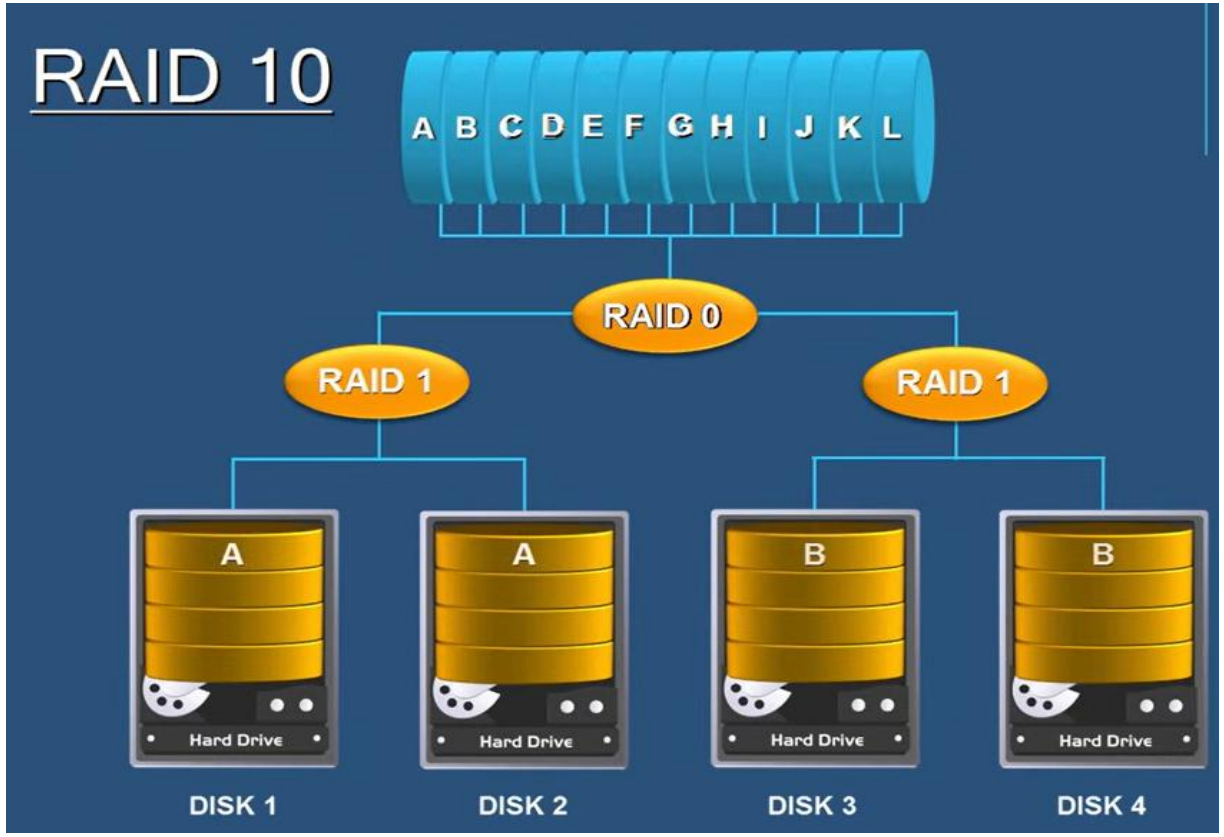
RAID 1/0

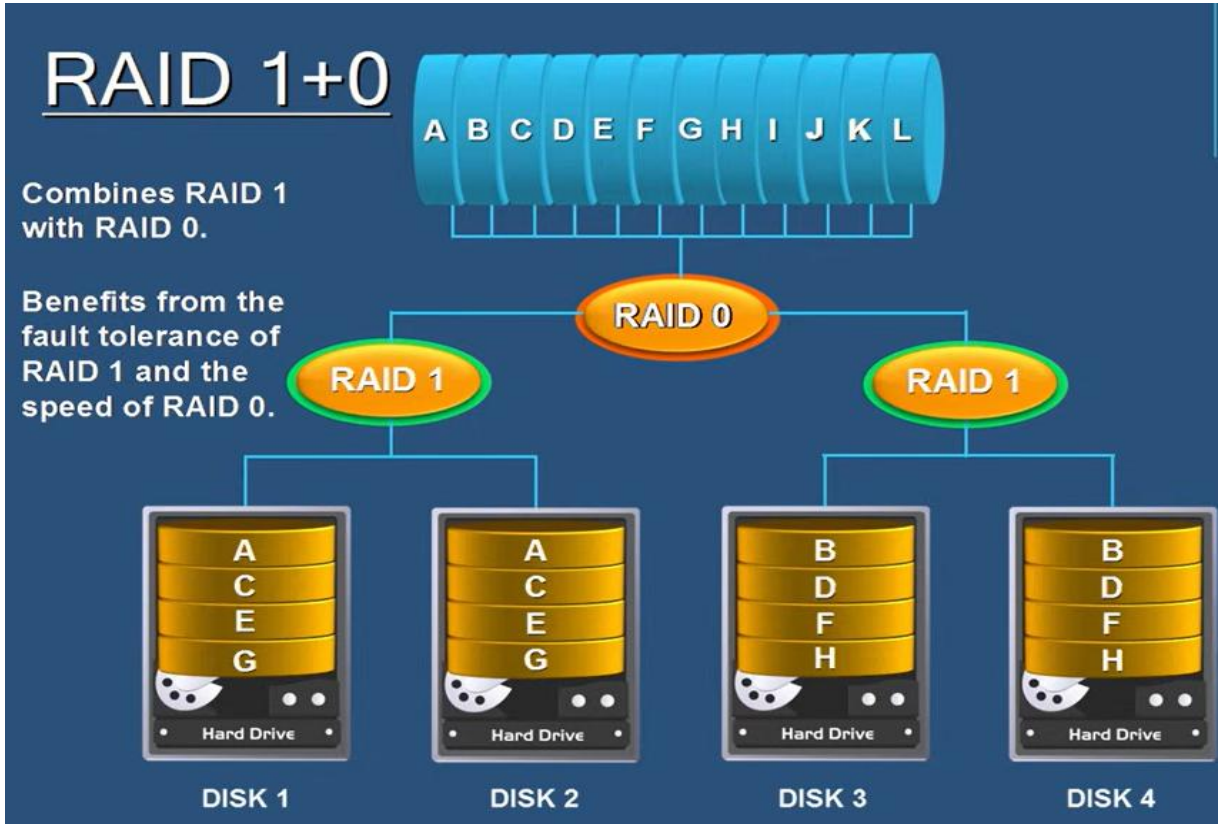
- RAID 10 is a combine of RAID 0 and RAID 1 to form a RAID 10.
- To setup Raid 10, we need at least 4 number of disks.
- The array first stripes the data into blocks (RAID 0) and then creates a mirror image for it in separate drives (RAID 1).
- It is the a good option for I/O-intensive applications email, web servers, databases and operations that require high disk performance.




(a) RAID 1+0
(b) RAID 0+1
Data from Host
E
D
C
B
A
Striping
Mirroring
Mirroring
Mirroring
RAID
Controller
Mirroring
Striping
Striping
A1
B1
C1
D1
E1
A2
B2
C2
D2
E2
A3
B3
C3
D3
E3
Mirror Set A
Mirror Set B
Mirror Set C
Stripe Set A
Stripe Set B

RAID 1/0


RAID 10
A B C D E F G H I J K L
RAID 0
RAID 1
RAID 1
A
A
B
B
Hard Drive
Hard Drive
Hard Drive
Hard Drive
DISK 1
DISK 2
DISK 3
DISK 4


RAID 1+0
A B C D E F G H I J K L
Combines RAID 1 with RAID 0.
Benefits from the fault tolerance of RAID 1 and the speed of RAID 0.
RAID 0
RAID 1
RAID 1
A C E G
A C E G
B D F H
B D F H
Hard Drive
Hard Drive
Hard Drive
Hard Drive
DISK 1
DISK 2
DISK 3
DISK 4

RAMAIAH
Institute of Technology

RAID Levels - Nested RAID

Advantages of **RAID 10**:

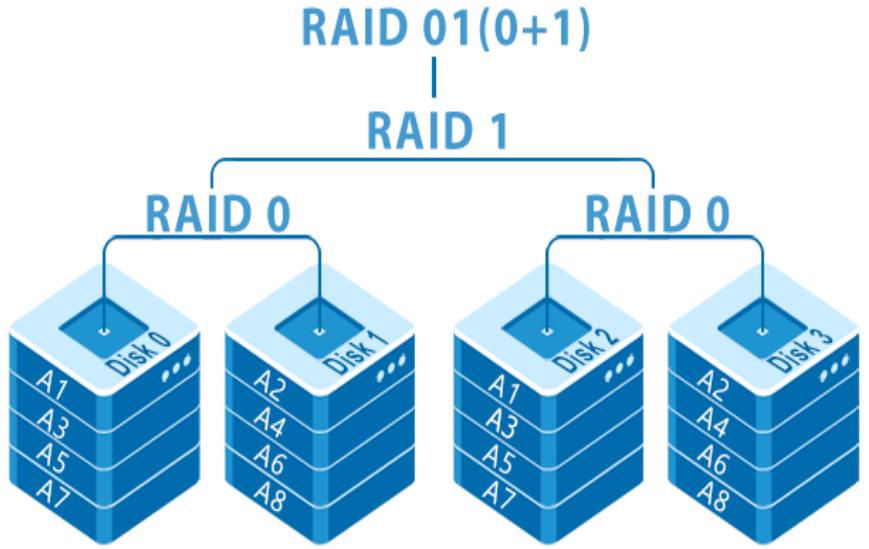
higher reliability level;

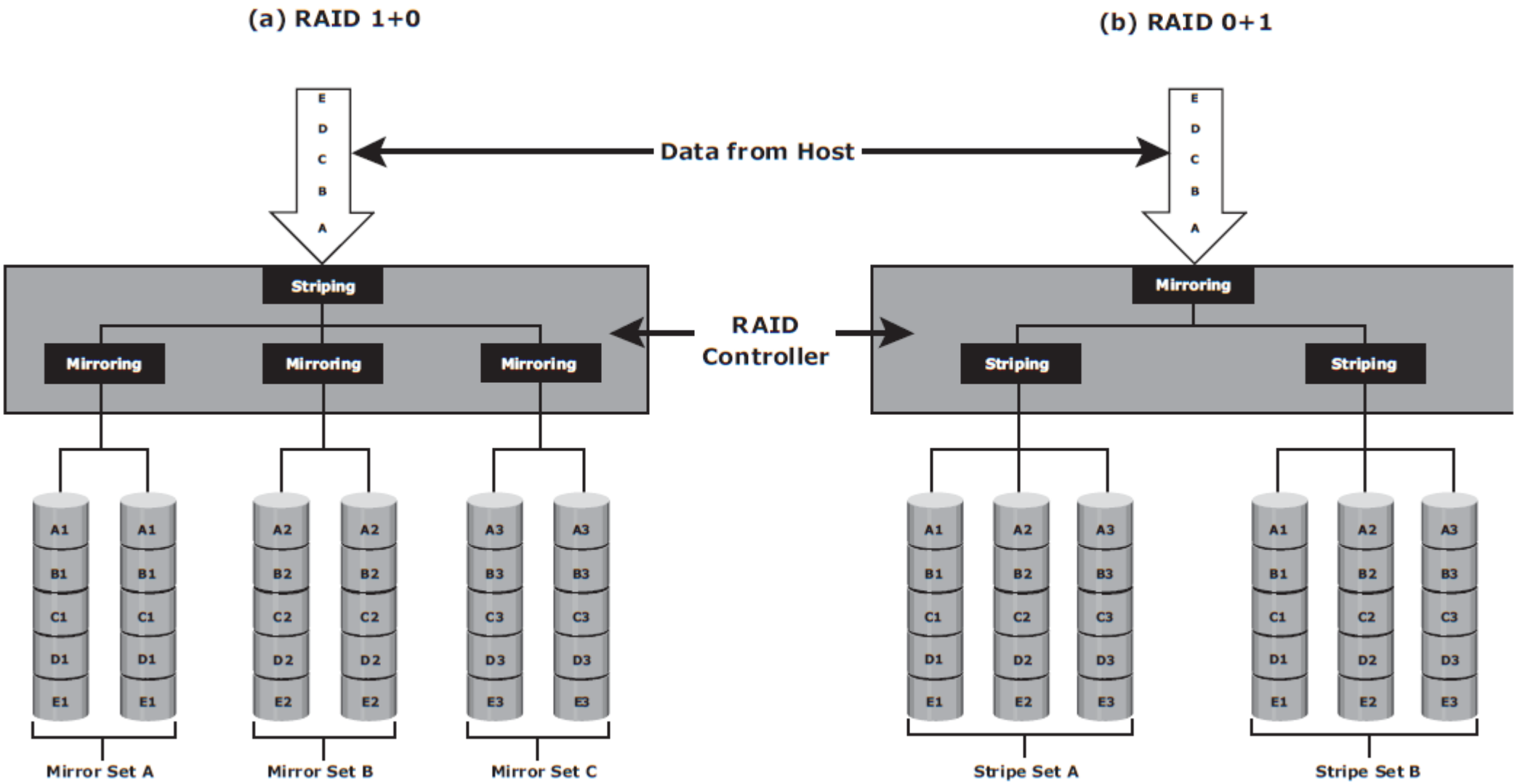
Improves performance

better suited for software controllers;

RAID 0/1

- **RAID 01** (RAID 0+1) is one type of **combined RAID array**.
- It allows you to implement the speed of **RAID 0** and the reliability of **RAID 1** in a single array.
- **RAID 01** is a RAID 1 array with two RAID 0 arrays inside.
- The data stream is first copied and then each copy is striped and written to two (or more) disks. Hence, the minimum number of disks to implement RAID 01 is four.




(a) RAID 1+0
(b) RAID 0+1
Data from Host
E
D
C
B
A
Striping
Mirroring
Mirroring
Mirroring
RAID
Controller
Mirroring
Striping
Striping
A1
B1
C1
D1
E1
A2
B2
C2
D2
E2
A3
B3
C3
D3
E3
Mirror Set A
Mirror Set B
Mirror Set C
Stripe Set A
Stripe Set B

RAID Levels - Nested RAID

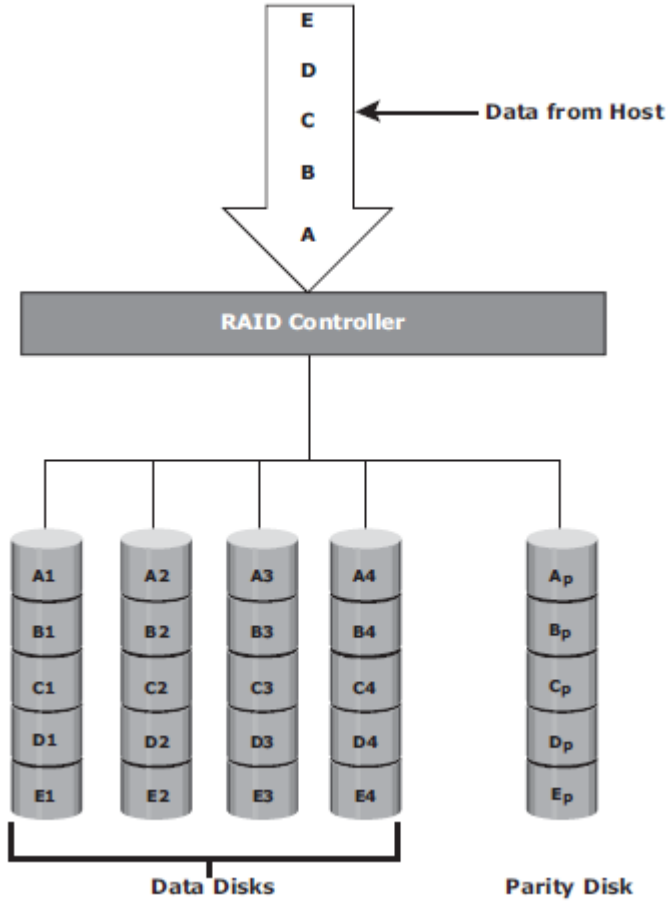
The advantage of **RAID 01**:

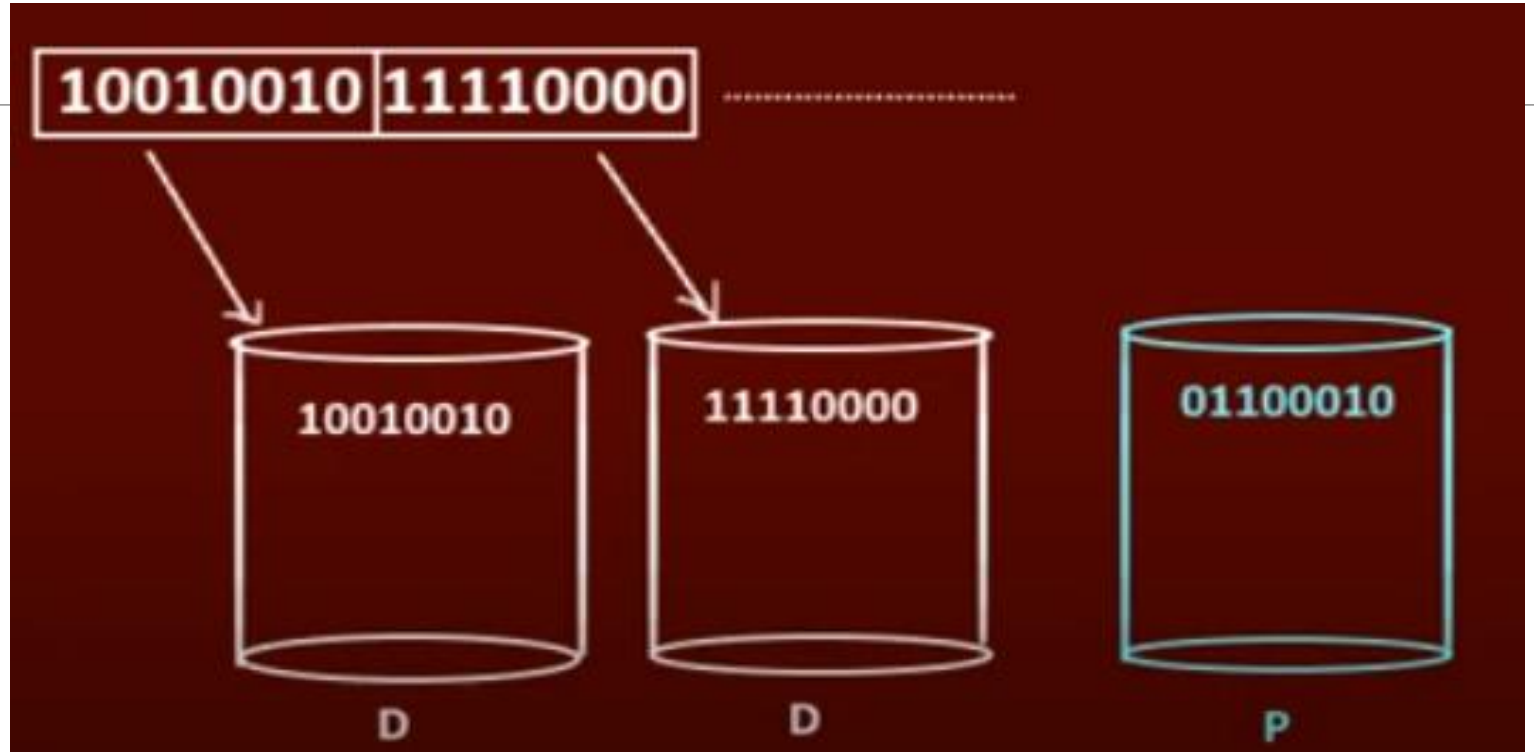
faster performance;

data remains available as long as at least one group of disks is in working order;

RAID Levels – RAID 3

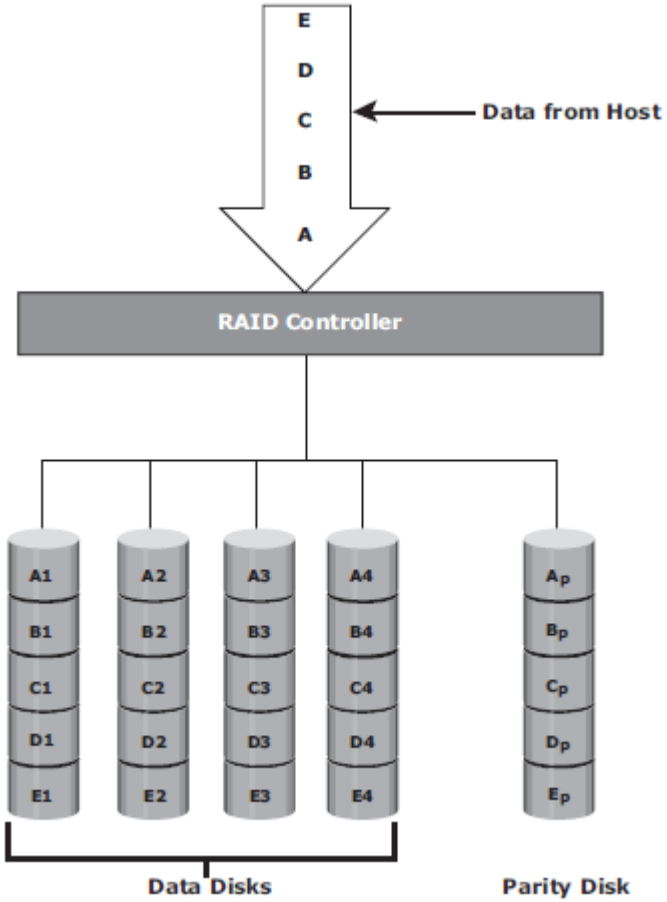
- RAID 3 stores parity information for data redundancy on a separate parity disk and uses byte-level striping.
- Byte-Level Striping: Data is divided into bytes and distributed across multiple disks.
- In this setup, parity bits—which are needed to recover data in the event of a disk failure—are kept on a different drive and are striped across many disks at the byte level.




10010010 11110000
10010010
11110000
01100010
D
D
P

RAID Levels – RAID 3

- 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.



RAID Levels – RAID 3

Advantages of RAID 3

High Data Transfer Rates: Suitable for applications with large file sizes due to its ability to transfer data in bulk.

Data can be accessed from multiple disks in parallel, which speeds up read operations.

Disadvantages of RAID 3

Parity Disk Bottleneck: The dedicated parity disk can become a performance bottleneck, especially during write operations.

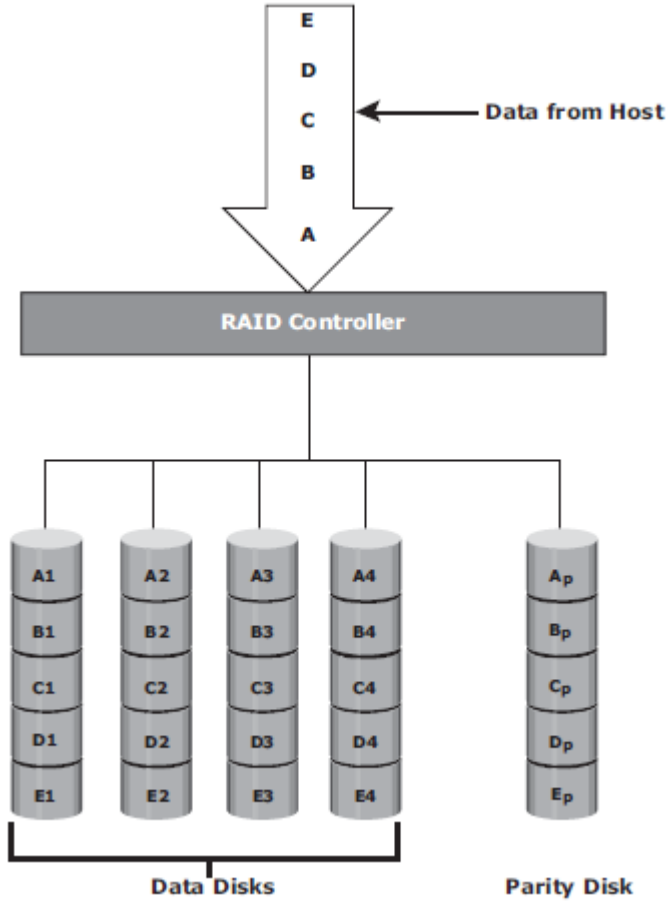
Low Performance with Small Files: The overhead of byte-level striping causes slower performance when working with small files.

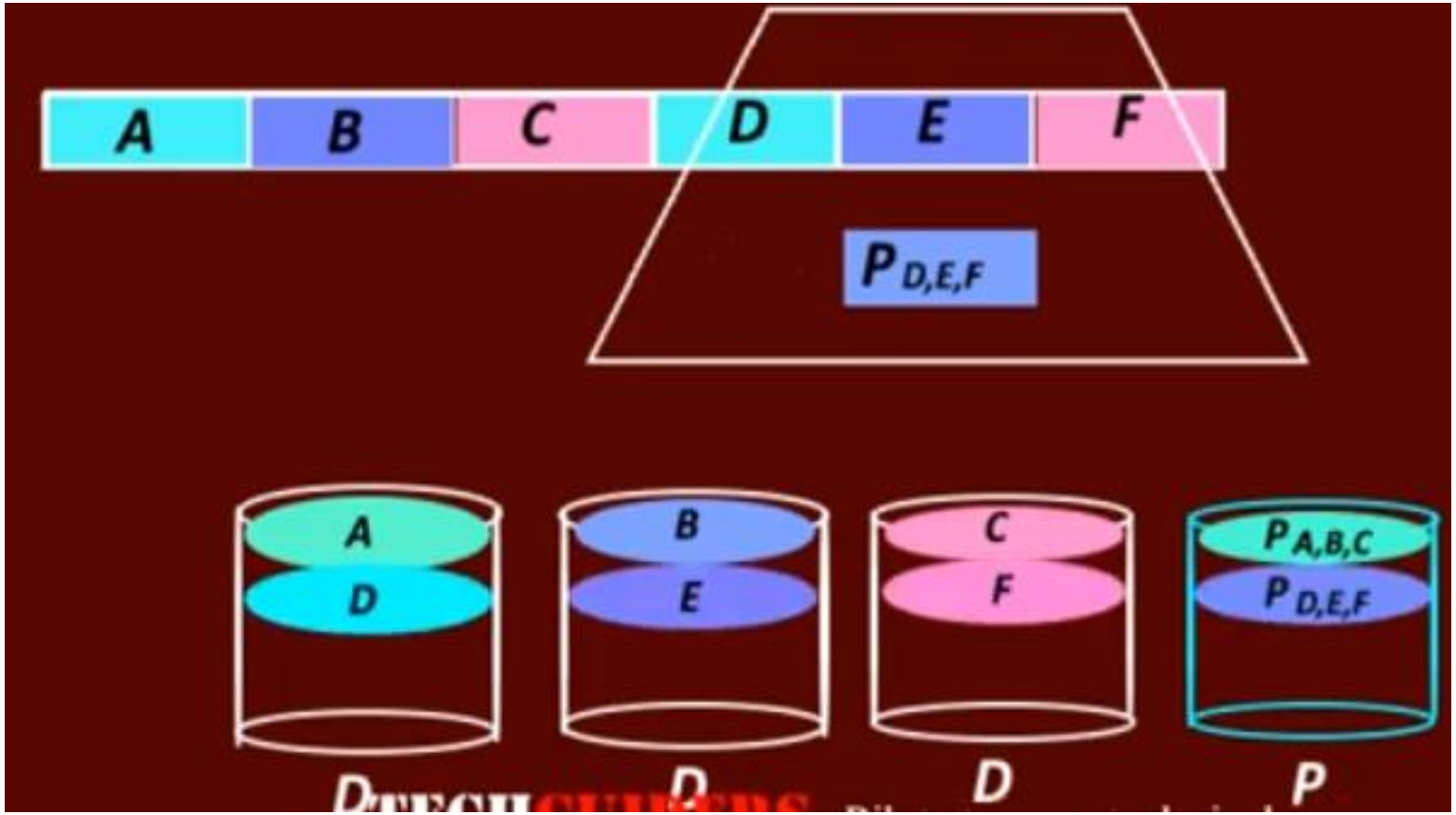
RAID Levels – RAID 4

Block-level striping and a separate parity disk are used in RAID 4.

Block-Level Striping: Data is divided into blocks and distributed across multiple disks.

This configuration stores the parity information for each data block on a different disk, and divides the data into blocks that are striped across multiple disks.

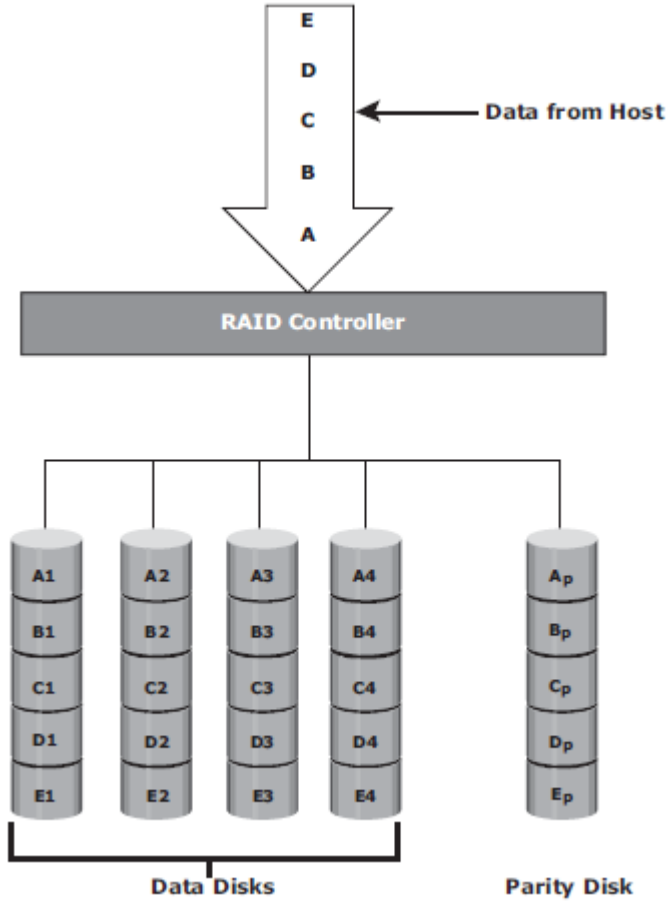



A
B
C
D
E
F
P D,E,F
A
D
B
E
C
F
P A,B,C
P D,E,F
D
D
P

RAID Levels – RAID 4

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 Levels – RAID 4

Advantages of RAID 4

Effective Block-Level Access: Block-level striping enables simultaneous I/O requests.

Low Storage Overhead: In comparison to RAID 5, parity storage needs are reduced.

Disadvantages of RAID 4

sluggish Random Writes: Because of distinct block parity, sluggish random write operations may have an effect on performance.

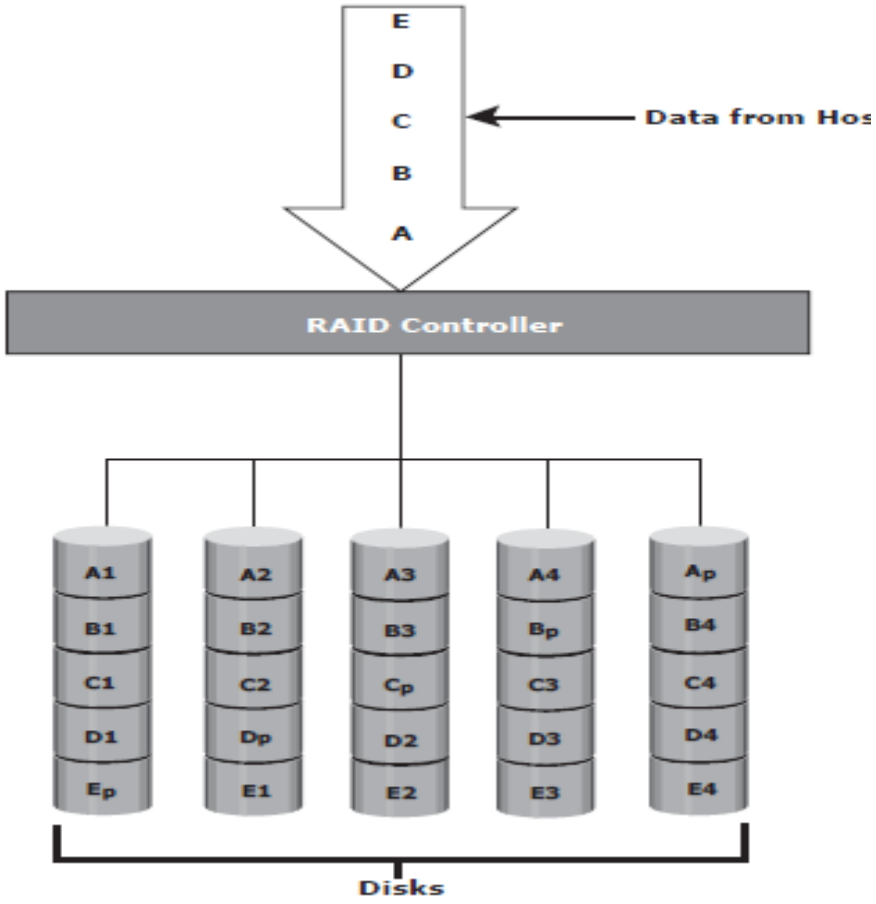
RAID 3	RAID 4
RAID 3 stands for Redundant Array of Independent Disk level 3.	RAID 4 stands for Redundant Array of Independent Disk level 4.
In RAID 3 technology, Byte-level Striping is used.	In RAID 4 technology, Block-level Striping is used.
In this level, parity bits are generated for each disk section and stored on a different disk.	In this level, parity bits are generated for the entire block of data and stored on a different disk
Random read will have worst performance.	Good Random reads, as the data blocks are striped.
Performance is good in case of large sized files.	Performance is low because only one block is accessed at a time

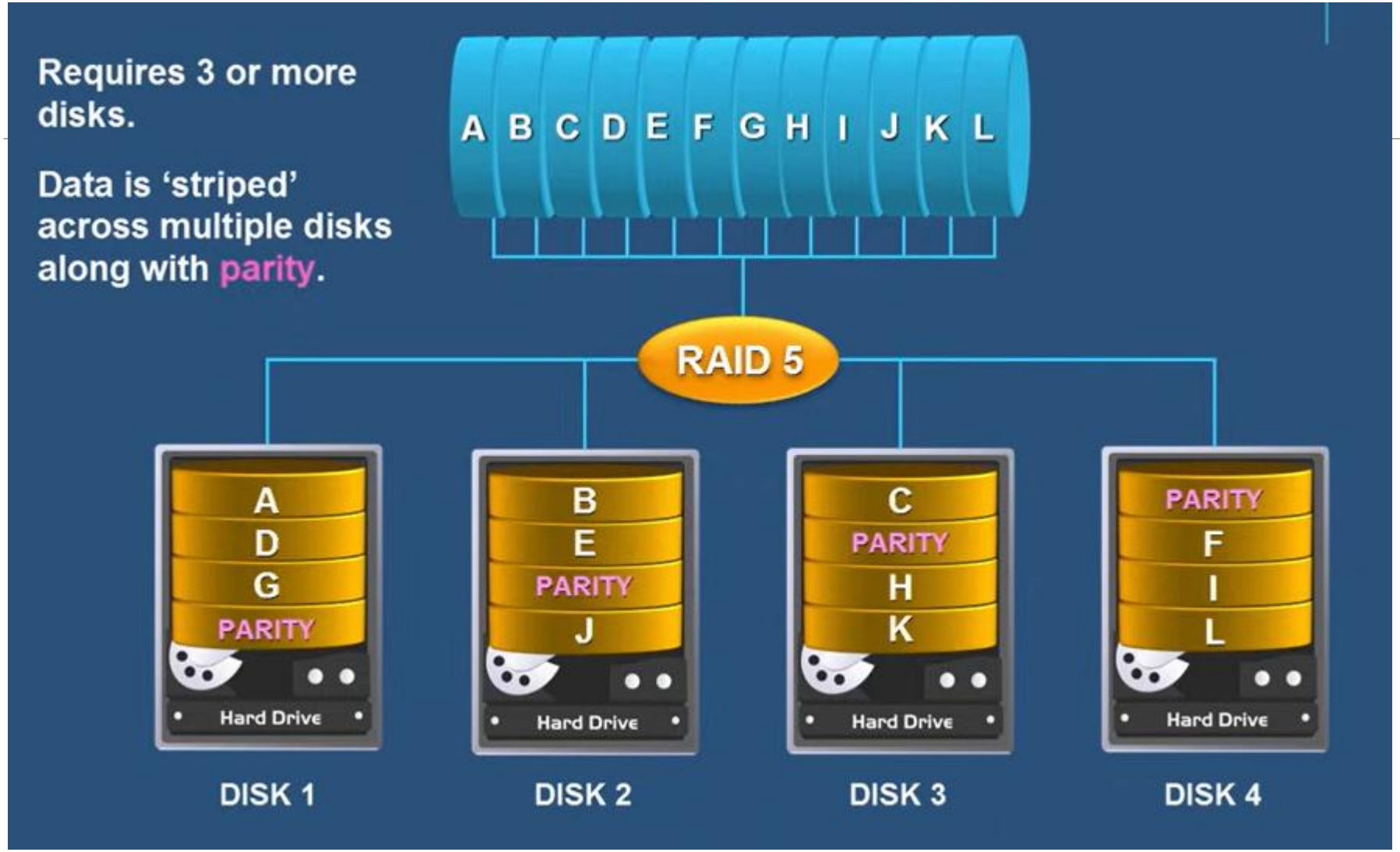
RAID Levels – RAID 5

RAID 5 uses striping. The drives (strips) are also independently accessible.

The difference between RAID 4 and RAID 5 is the parity location.

In RAID 5, parity is distributed across all disks to overcome the write bottleneck of a dedicated parity disk.

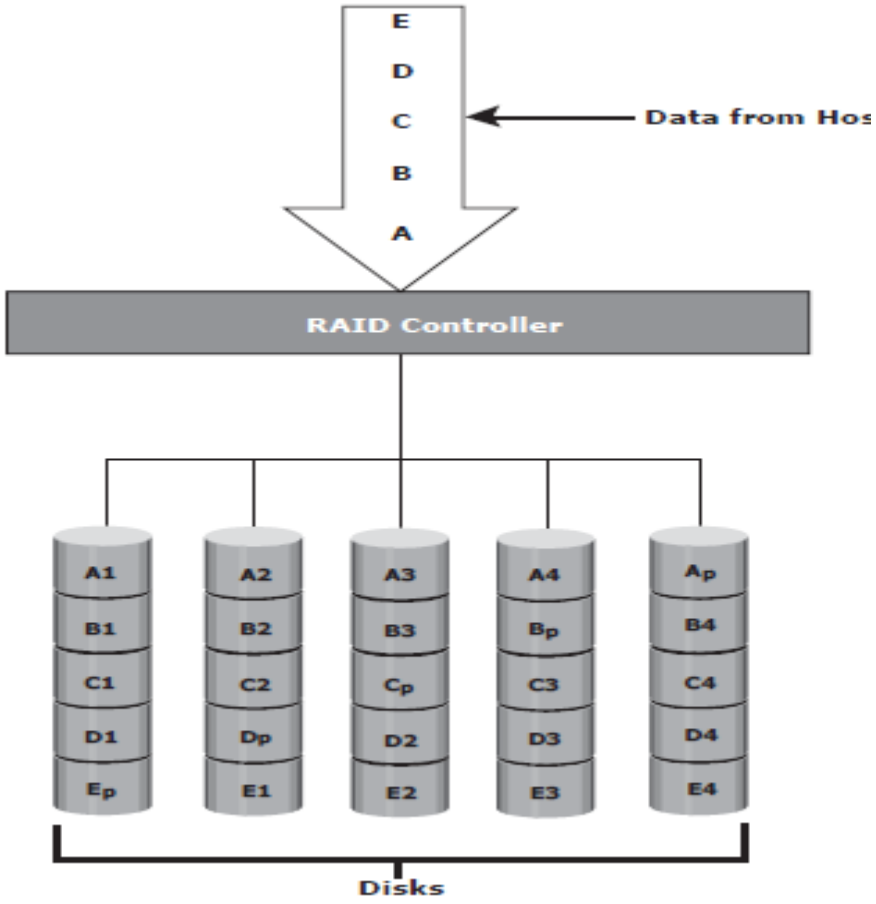



Requires 3 or more disks.
Data is 'striped' across multiple disks along with parity.
A B C D E F G H I J K L
RAID 5
A
D
G
PARITY
Hard Drive
DISK 1
B
E
PARITY
J
Hard Drive
DISK 2
C
PARITY
H
K
Hard Drive
DISK 3
PARITY
F
I
L
Hard Drive
DISK 4

RAID Levels – RAID 5

RAID 5 is good for

- Random, read-intensive I/O applications preferred for messaging,
- Data mining,
- Medium-performance media serving,
- Relational database management system (RDBMS)



RAID Levels – RAID 6

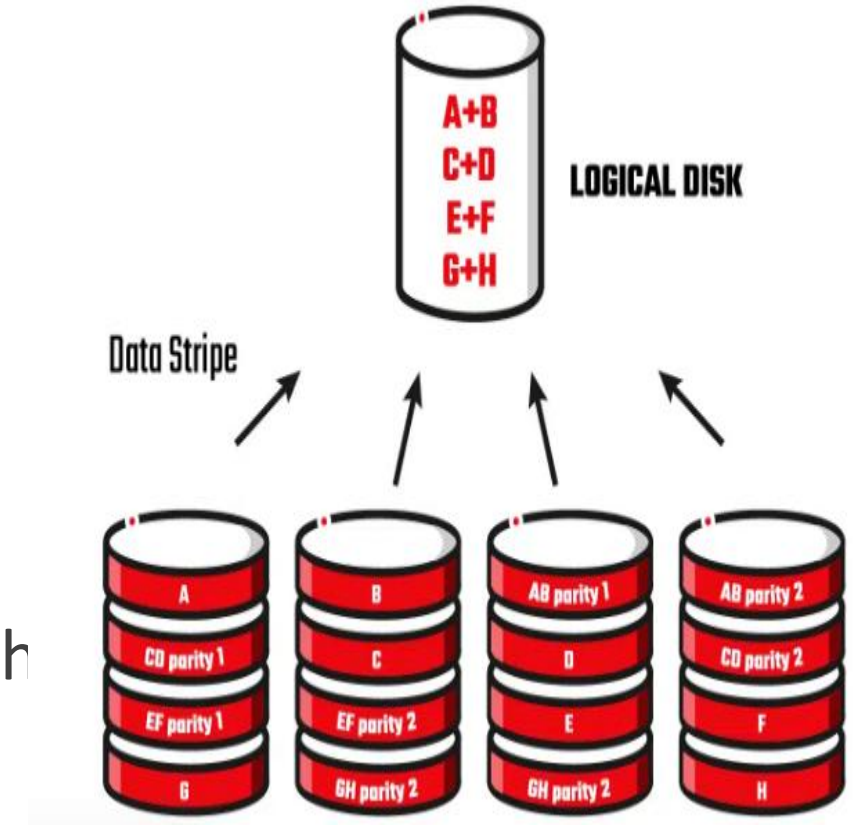
RAID 6 is an array similar to RAID 5 with an addition of its double parity feature.

It is also referred to as the double-parity RAID.

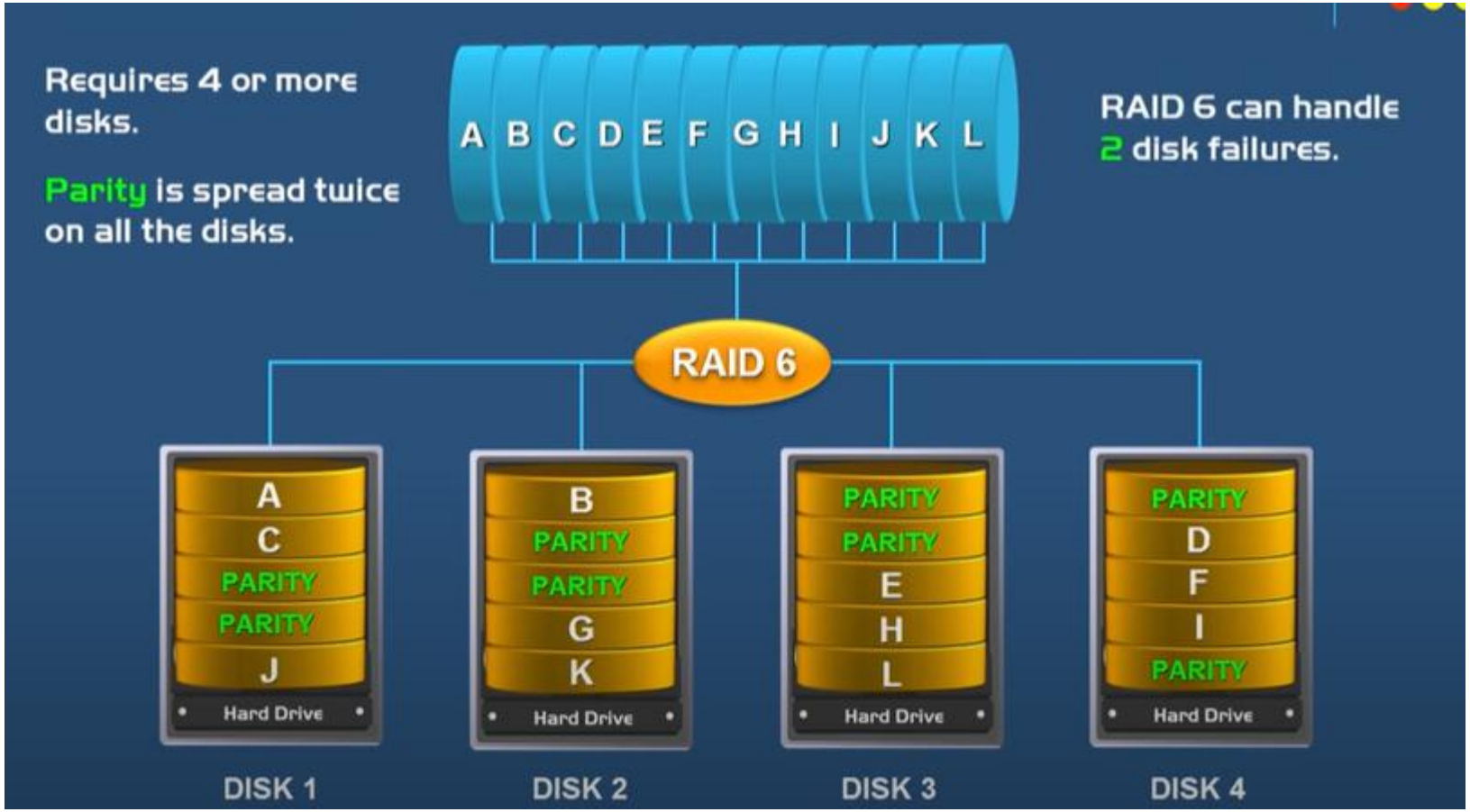
This setup requires a minimum of four drives.

The setup resembles RAID 5 but includes two additional parity blocks distributed across the disk.

It uses block-level striping to distribute the data across the array and stores two parity blocks for each data block.



RAID Levels – RAID 6



UNIT 2- Storage System

Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- RAID Array Components,
- RAID Techniques,
- Levels,
- Impact on Disk Performance,
- Comparison
- Hot Spares

RAID Impact on Disk Performance

When choosing a RAID type, it is imperative to consider its impact on

- Disk performance
- Application IOPS

RAID Impact on Disk Performance

Write penalty

- Write penalty in a Storage Area Network refers to the extra I/O overhead that occurs when writing data to disk due to the parity check and mirroring implemented by the RAID algorithm

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

RAID Impact on Disk Performance

- RAID 1 implementation, every **write operation** must be performed on two disks configured as a mirrored pair,
- RAID 5 implementation, a write operation may manifest as four I/O operations.
- When performing I/Os to a disk configured with RAID 5, the controller has to read, recalculate, and write a parity segment for every data write operation.

RAMAIAH
Institute of Technology

RAID Impact on Disk Performance

Figure illustrates a single write operation on RAID 5 that contains a group of five disks.

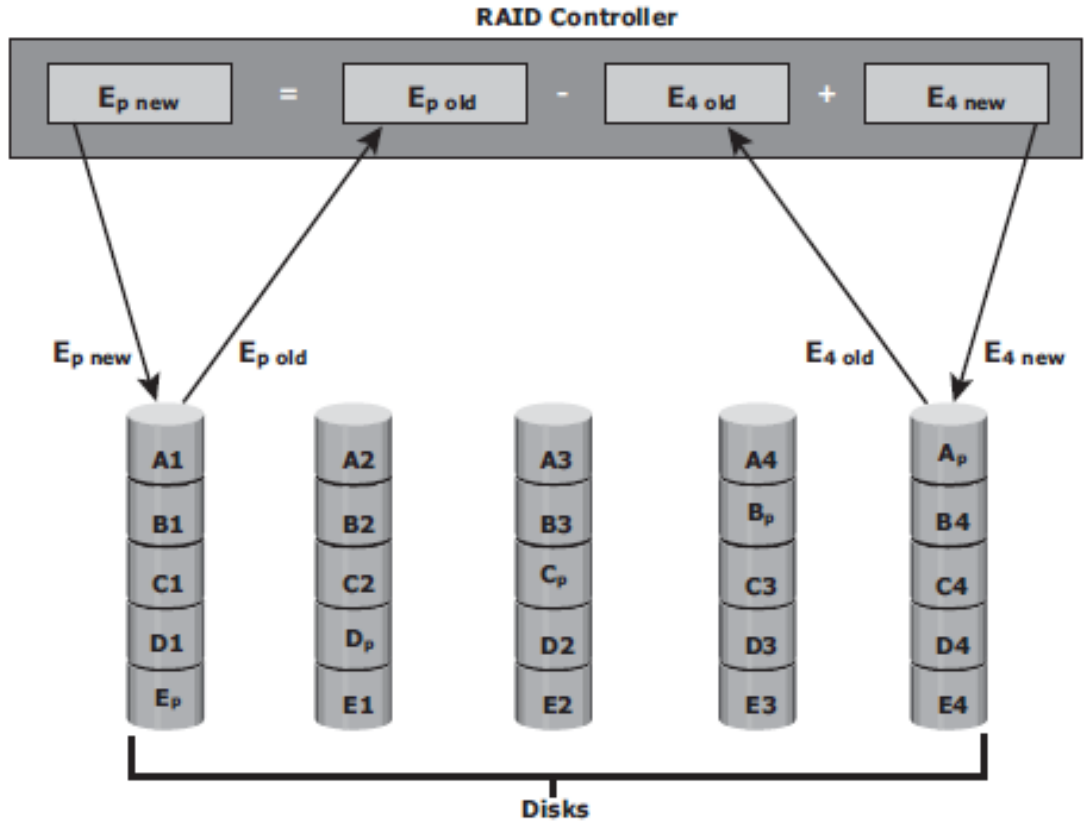


Figure 3-11: Write penalty in RAID 5

RAID Impact on Disk Performance

The parity (P) at the controller is calculated as follows:

$$E_p = E_1 + E_2 + E_3 + E_4 \text{ (XOR operations)}$$

Whenever the controller performs a write I/O, parity must be computed by reading the old parity (E_p old) and the old data (E_4 old) from the disk(Two read I/Os)

Then, the new parity (E_p new) is computed as follows:

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

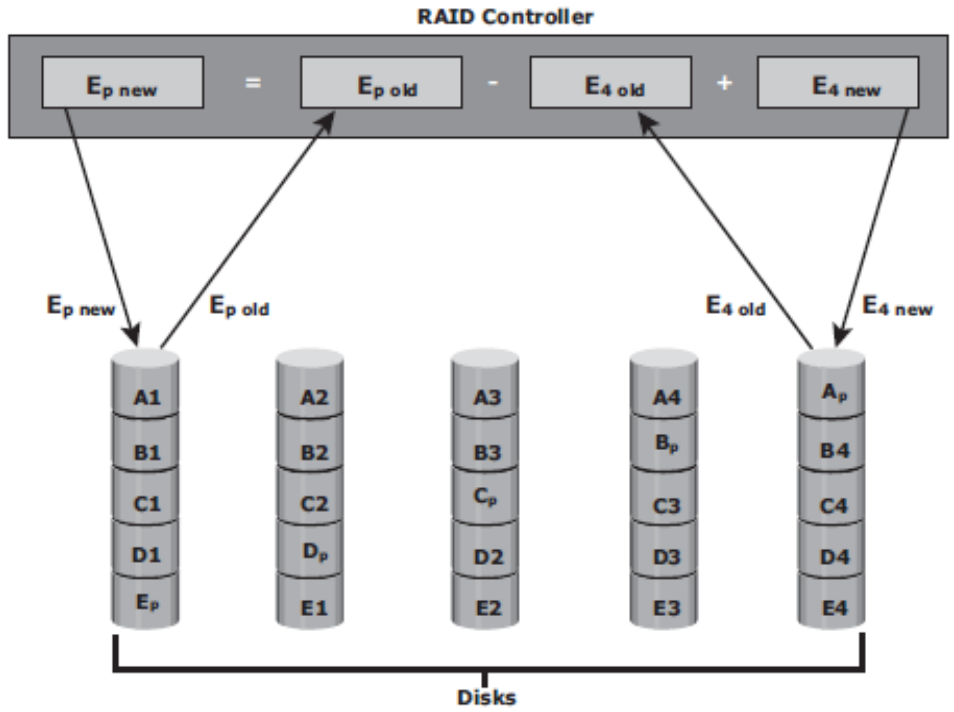


Figure 3-11: Write penalty in RAID 5

RAID Impact on Disk Performance

The parity (P) at the controller is calculated as follows:

$$E_p = E_1 + E_2 + E_3 + E_4 \text{ (XOR operations)}$$

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

AP →A1 +A2+ A3+ A4
EP → E1 + E2+ E3+ E4
Assume Disk2 fails – A2
Consider AP → old parity value
A2 → old disk value

APnew → APold-A2old + A2 new

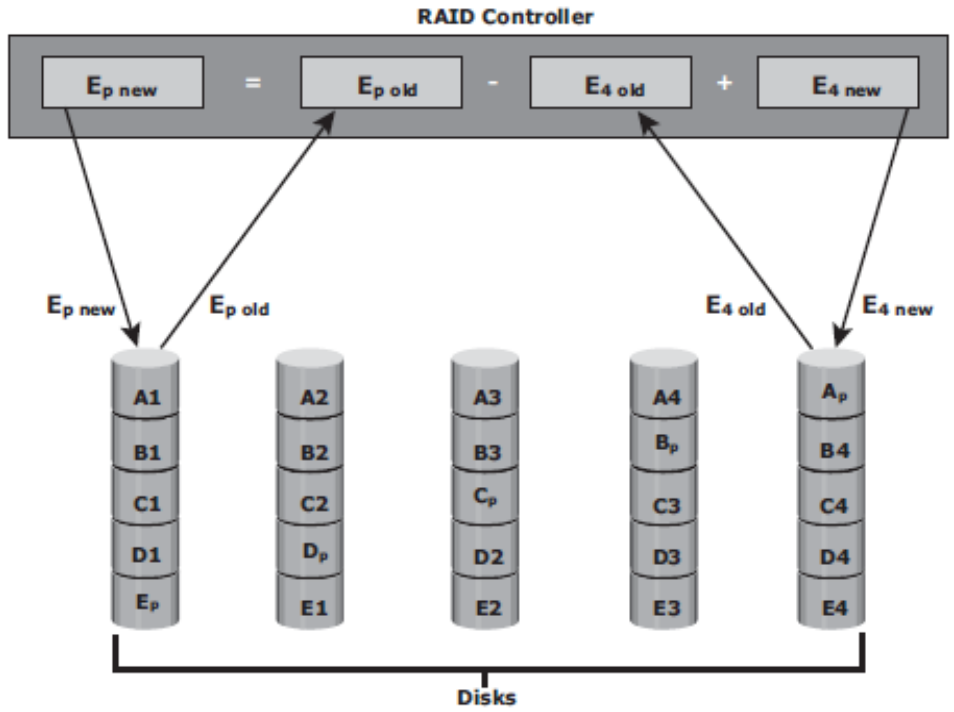


Figure 3-11: Write penalty in RAID 5

RAID Impact on Disk Performance

Then, the new parity (E_p new) is computed as follows:

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

After computing the new parity, the controller completes the write I/O by writing the new data and the new parity onto the disks, amounting to two write I/Os.

Therefore, the controller performs two disk reads and two disk writes for every write operation, and the write penalty is 4(RAID 5).

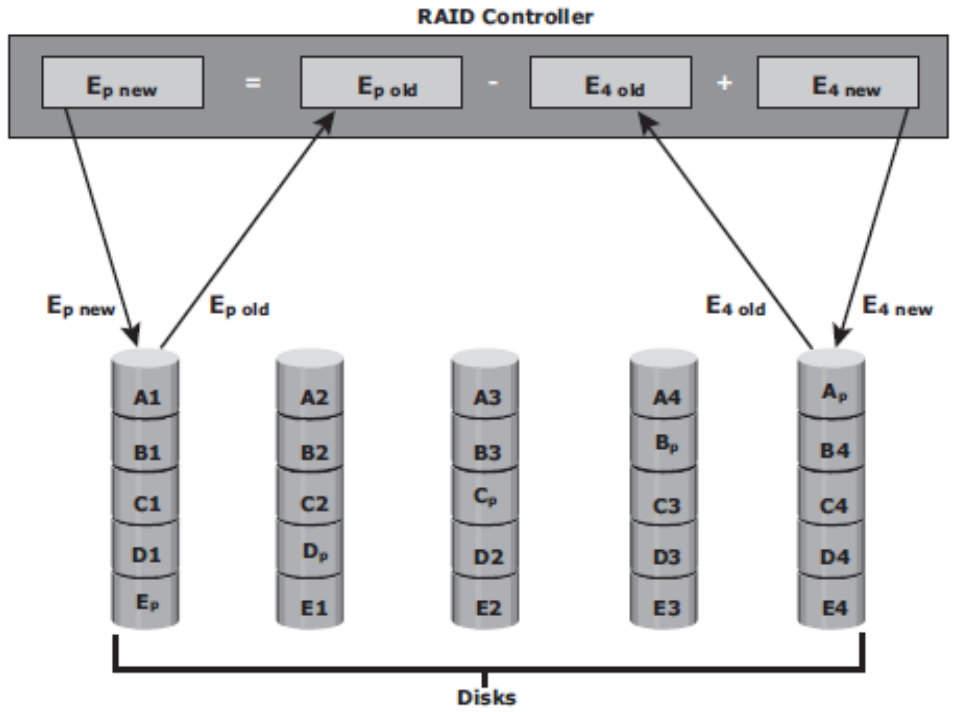


Figure 3-11: Write penalty in RAID 5

RAID Impact on Disk Performance

In RAID 6, which maintains dual parity, a disk write requires three read operations: two parity and one data.

After calculating both new parities, the controller performs three write operations: two parity and an I/O.

Therefore, in a RAID 6 implementation, the controller performs six I/O operations for each write I/O, and the write penalty is 6.

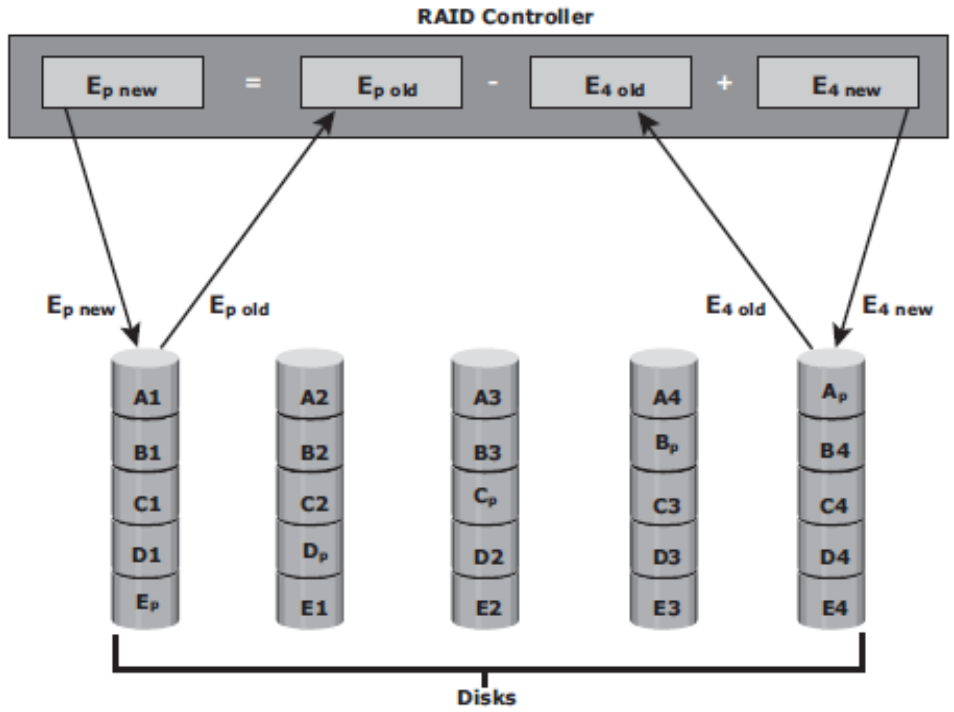


Figure 3-11: Write penalty in RAID 5

RAID Impact on Disk Performance

Application IOPS and RAID Configurations

When deciding the number of disks required for an application, it is important to consider the impact of RAID based on IOPS generated by the application.

Consider an application that generates 5,200 IOPS, with 60 percent of them being r
i.
Calculate the disk load in RAID 5, RAID 1 and RAID 6. ii. Calculate the number of dis
required for the application. HDD (hard Disk Drive) with a specification of a maximu
IOPS is used.

RAID Impact on Disk Performance

Application IOPS and RAID Configurations

Example illustrates the method to compute the disk load in different types of RAID.

Consider an application that generates 5,200 IOPS, with 60 percent of them being reads.

The disk load in RAID 5 is calculated as follows:

RAID Impact on Disk Performance

Consider an application that generates 5,200 IOPS, with 60 percent of them being reads.

The disk load in RAID 5 is calculated as follows:

Write Penalty for RAID 5

RAID 5 disk load (reads + writes) = 0.6 * 5,200 + **4** * (0.4 * 5,200)

= 3,120 + 4 * 2,080

= 3,120 + 8,320

= 11,440 IOPS

RAID Impact on Disk Performance

Consider an application that generates 5,200 IOPS, with 60 percent of them being reads.

The disk load in RAID 1 is calculated as follows:

Write Penalty for RAID 1

RAID 1 disk load = 0.6 * 5,200 + **2** * (0.4 * 5,200)

= 3,120 + 2 * 2,080

= 3,120 + 4,160

= 7,280 IOPS

RAID Impact on Disk Performance

Consider an application that generates 5,200 IOPS, with 60 percent of them being reads.

If a disk drive with a specification of a maximum 180 IOPS needs to be used, the number of disks required to meet the workload for the RAID configuration would be as follows:

RAID 5 = 11,440/180

= 64 disks

RAID 1 = 7,280/180

= 42 disks (approximated to the nearest even number)

RAID Impact on Disk Performance

Consider an application that generates 7,200 IOPS, with 60 percent of them being reads. Calculate the disk load in RAID1, RAID5 and RAID6. If a HDD with a maximum of 180 IOPS for the application need. Calculate the number of disks required to meet the workload for RAID1, RAID5 and RAID6

RAMAIAH
Institute of Technology

RAID Impact on Disk Performance

Consider an application that generates 7,200 IOPS, with 60 percent of them being reads. Calculate the disk load in RAID1, RAID5 and RAID6. If a HDD with a maximum of 180 IOPS for the application need. Calculate the number of disks required to meet the workload for RAID1, RAID5 and RAID6

Disk load for RAID 1 = 0.6 * 7,200 + **2** * (0.4 * 7,200) =10080 IOPS

Disk load for RAID 5 = 0.6 * 7,200 + **4** * (0.4 * 7,200) = 15480 IOPS

Disk load for RAID 6 = 0.6 * 7,200 + **6*** (0.4 * 7,200)=21600 IOPS

RAID Impact on Disk Performance

Disk load for RAID 1 = 0.6 * 7,200 + **2** * (0.4 * 7,200) =10080 IOPS

Disk load for RAID 5 = 0.6 * 7,200 + **4** * (0.4 * 7,200) = 15480 IOPS

Disk load for RAID 6 = 0.6 * 7,200 + **6*** (0.4 * 7,200)=21600 IOPS

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

For RAID 5 = 15480/180=88

For RAID 6 = 21600/180=120

RAID Impact on Disk Performance

An application has 1000 heavy users at a peak of 2 IOPS each and 2000 typical users at a peak of 1 IOPS each, with a read/write of 2:1. It is estimated that that application also experiences an overhead of 20% for other workloads. Calculate the IOPS required for RAID1, RAID3,RAID5 and RAID 6. Also compute the number of drives required to support the application in different RAID environments if 10K rpm drives with a rating of 130 IOPS per drive were used.

RAID Impact on Disk Performance

1000 heavy users at a peak of 2 IOPS each = 2000 IOPS

2000 heavy users at a peak of 1 IOPS each = 2000 IOPS

Assume maximum concurrency of 90%

=(2000+2000)*0.90

= 3600 host based IOPS for 3000 users during peak activity period

read-write ratio 2:1

For RAID1 = 3600*2/3 +(2*1/3*3600) = 4800 IOPS
For RAID3 = 3600*2/3 +(4*1/3*3600) = 7200 IOPS
For RAID5 = 3600*2/3 +(4*1/3*3600) = 7200 IOPS
For RAID6 = 3600*2/3 +(6*1/3*3600) = 9600 IOPS

Number of drives required to support the application in different RAID environments, if 10k rpm drives with a rating of 130 IOPS per drive.

For RAID1 = 4800/130 = 37drives
For RAID3 = 7200/130 = 55drives
For RAID5 = 7200/130 = 55drives
For RAID6 = 9600/130 = 74drives

RAMAIAH Institute of Technology

Example -2

Total IOPS at peak workload is 1200. Read/write ratio is 2:1. Calculate the disk load at peak activity for RAID 5 and RAID 1/0

Hint :

2+1=3 so read operations is 2/3 and write operations is 1/3

Total IOPS at peak workload is 1200. Read/write ratio is 2:1. Calculate the disk load at peak activity for RAID 5 and RAID 1/0

2+1=3 so read operations is 2/3 and write operations is 1/3

RAID 1/0 : (2/3*1200) + 2 (1/3*1200)

= 1600 IOPS

RAID 5: (2/3*1200) + 4(1/3*1200)

= 2400 IOPS

Caclulate the disk requirements for an application that requires 2500 IOPS and 2 TB of strorage given that we have to implement Raid 5 and we have 40% writes of the total IOPS required and the disk available is having disk storage of 140 GB and 120 IOPS.

Caclulate the disk requirements for an application that requires 2500 IOPS and 2 TB of strorage given that we have to implement Raid 5 and we have 40% writes of the total IOPS required and the disk available is having disk storage of 140 GB and 120 IOPS.

Disk IOPS= (2500*0.6+4*0.4*2500)/120 =

Disk Capacity =2000/140=

Calculate the Disk requirement given that an application requires 2500 IOPS and requires 2TB of storage. There are disk with 180 IOPS and 140 GB available and disk utilization is 60%

RAID Comparison - Types of RAID levels.

Comparison of Common RAID types

RAID	Min Disks	Storage Efficiency %	Cost	Read Performance	Write Performance	Write* Penalty	Protection
0	2	100	Low	Good for both random and sequential reads	Excellent	1	No Protection
1	2	50	High	Better than a single disk	Good - Slower than a single disk because every write must be committed to all disks	2	Mirror Protection
5	3	67 - 97%	Moderate - Low	Good for random and sequential reads	Fair for random and sequential writes	4	Parity protection for single disk failure
6	3	33 - 94%	Extremely High - Low	Good for random and sequential reads	Poor to fair for random writes and fair for sequential writes	6	Parity protection for two disk failures
10	4	50	High	Good	Good	2	Mirror Protection
50	6	67 - 94%	Moderate-High	Better	Moderate for random and sequential writes	4	Parity protection for single disk failure
60	6	33 - 88%	Extremely High - Moderate	Better	Fair for random and sequential writes	6	Parity protection for two disk failures

UNIT 2- Storage System

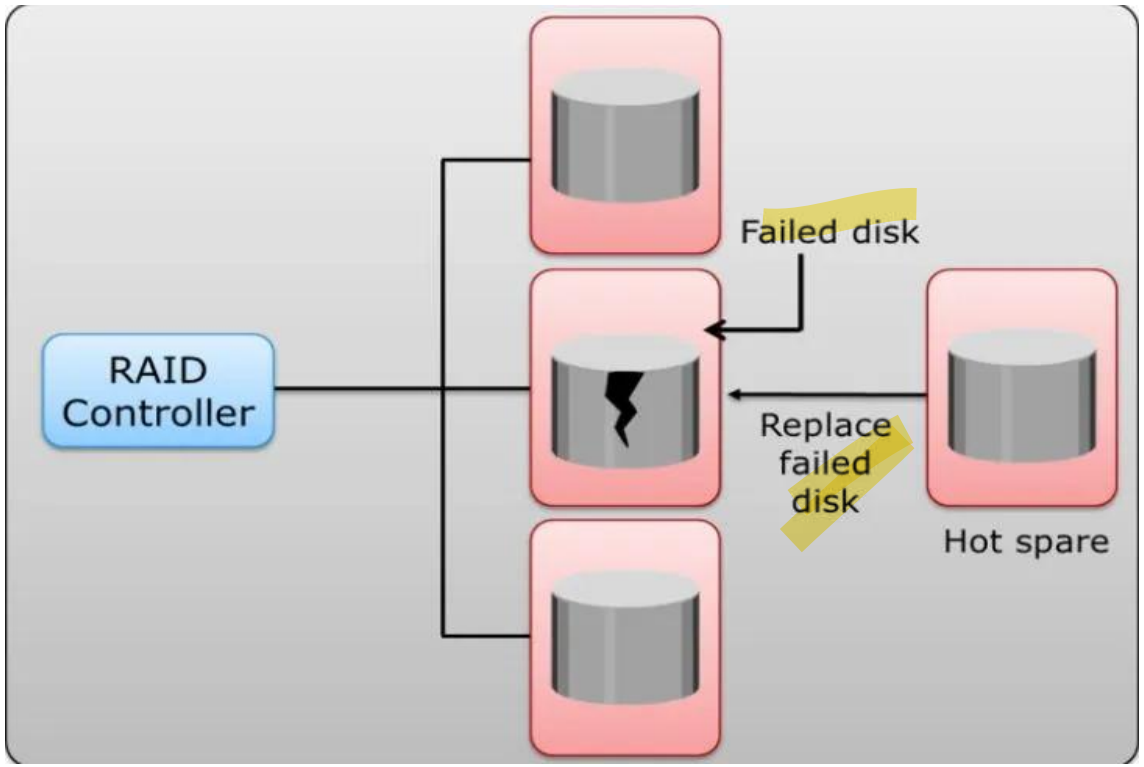
Data Protection: RAID (Chapter 3)

- RAID(Redundant array of independent disks) Implementation Methods,
- RAID Array Components,
- RAID Techniques,
- RAID Levels,
- RAID Impact on Disk Performance,
- RAID Comparison
- Hot Spares

What is a Hot Spare? Explain how a Hot Spare is utilized in the event of a disk failure.Justify

Hot Spares

- A hot spare refers to a spare drive in a RAID array that temporarily replaces a failed disk drive by taking the identity of the failed disk drive.
- A hot spare should be large enough to accommodate data from a failed drive.
- Some systems implement multiple hot spares to improve data availability.



Hot Spares

Different Methods of data recovery performed depending on the RAID implementation:

1. If parity RAID is used, the data is rebuilt onto the hot spare from the parity and the data on the surviving disk drives in the RAID set.
2. If mirroring is used, the data from the surviving mirror is used to copy the data onto the hot spare.

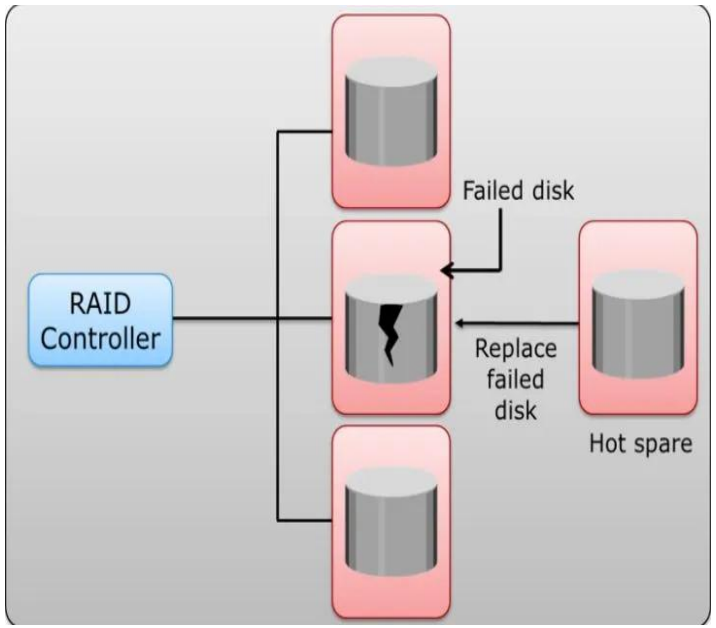
Hot Spares

1.When a new disk drive is added to the system, data from the hot spare is copied to it.

- The hot spare returns to its idle state, ready to replace the next failed drive.

2.Alternatively, the hot spare replaces the failed disk drive permanently.

- This means that it is no longer a hot spare, and a new hot spare must be configured on the array.



Hot Spares

A hot spare can be configured as automatic or user initiated, which specifies how it will be used in the event of disk failure.

Automatic configuration

- When the recoverable error rates for a disk exceed a predetermined threshold, the disk subsystem tries to copy data from the failing disk to the hot spare automatically.
- If this task is completed before the damaged disk fails, the subsystem switches to the hot spare and marks the failing disk as unusable.
- Otherwise, it uses parity or the mirrored disk to recover the data.

Hot Spares

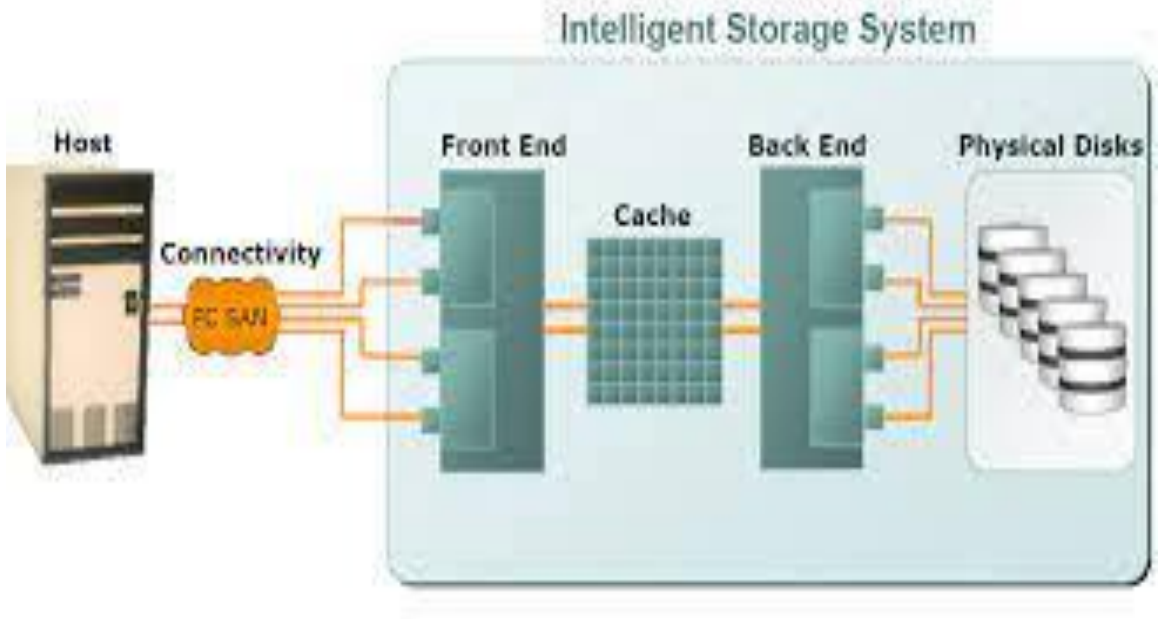
User initiated

- The administrator has control of the rebuild process.
- For example, the rebuild could occur overnight to prevent any degradation of system performance.
- However, the system is at risk of data loss if another disk failure occurs.

UNIT 2- Storage System

Intelligent Storage System(Chapter 4)

- Components of an Intelligent Storage System
- Storage Provisioning
- Types of Intelligent Storage Systems



Introduction

- A disk drive is a core element of storage that governs the performance of any storage system.
- Some of the older disk-array technologies could not overcome performance constraints due to the limitations of disk drives and their mechanical components.
- With advancements in technology, a new breed of storage solutions, known as intelligent storage systems, has evolved.
- These intelligent storage systems are feature-rich RAID arrays that provide highly optimized I/O processing capabilities.

Introduction

- These storage systems are configured with a large amount of memory (called cache) and multiple I/O paths and use sophisticated algorithms to meet the requirements of performance-sensitive applications.
- These arrays have an operating environment that intelligently and optimally handles the management, allocation, and utilization of storage resources.

Explain the components of an intelligent storage system with the help of a neat di

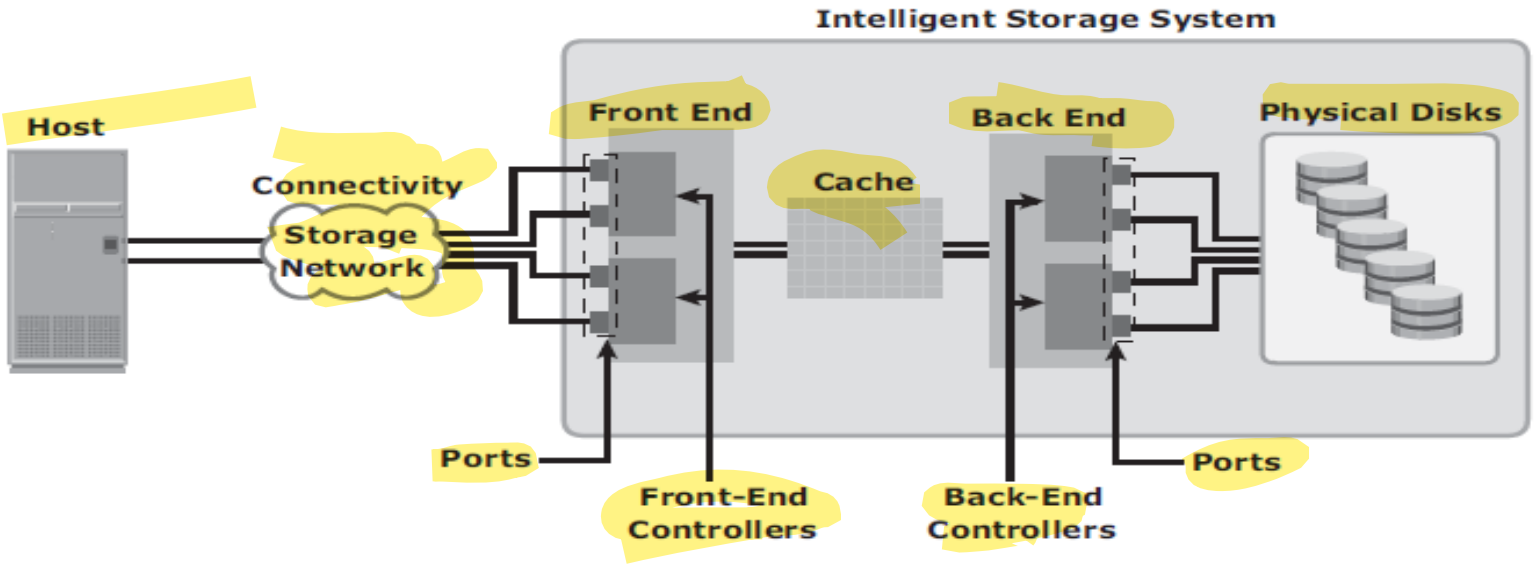
Explain the components of an intelligent storage system with the help of a neat di

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

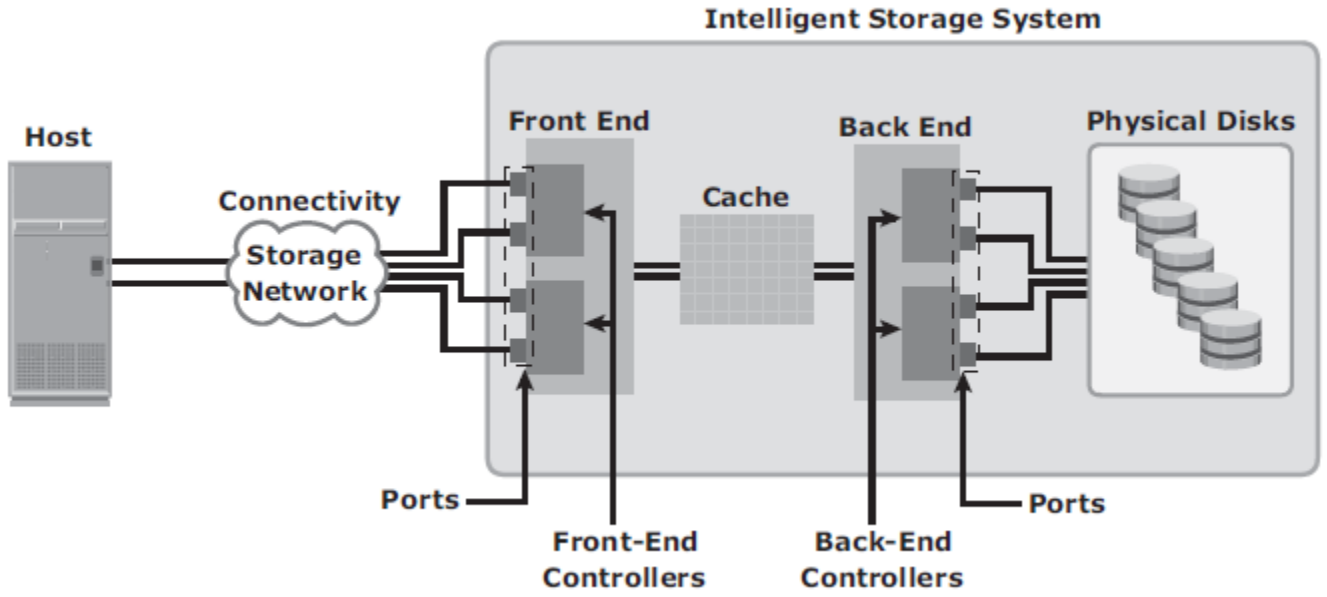
F C B P



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).

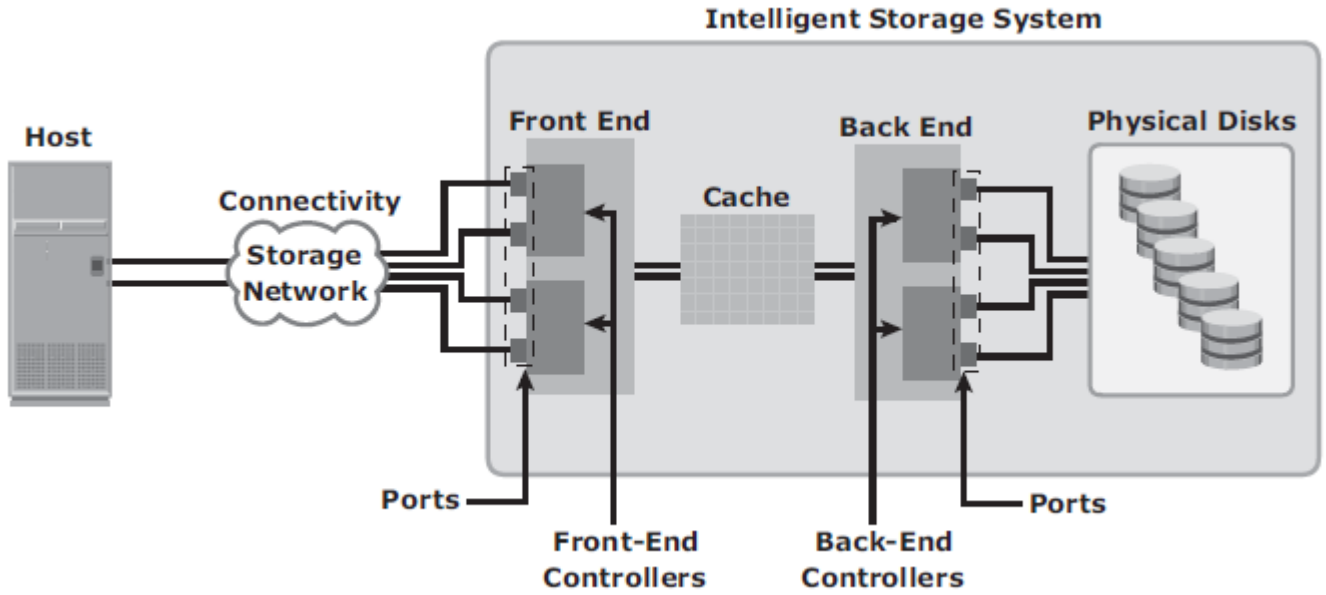
Components of an Intelligent Storage System

- Figure 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.



Components of an Intelligent Storage System

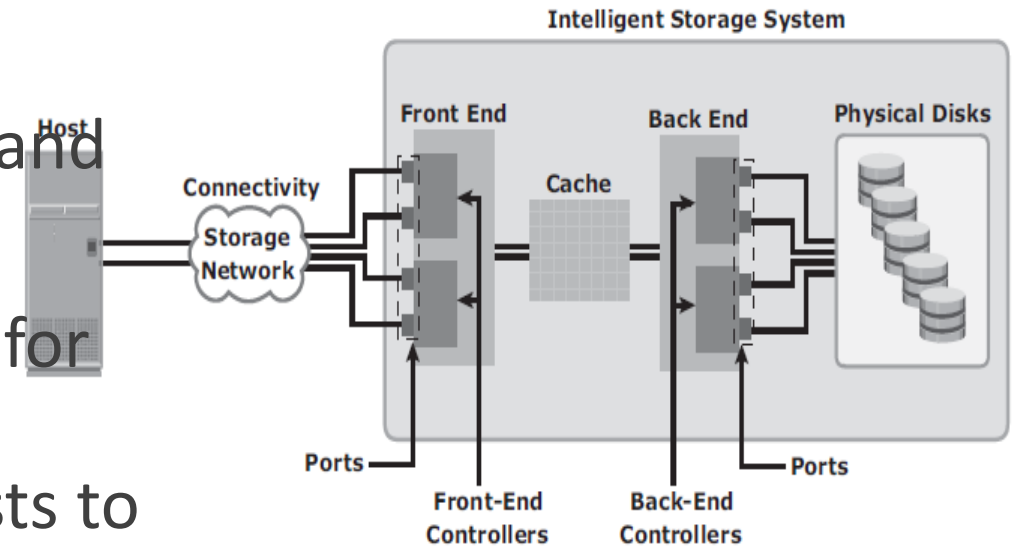
- Read request can be serviced directly from cache if the requested data is found in the cache.



Components of an Intelligent Storage System

Front end

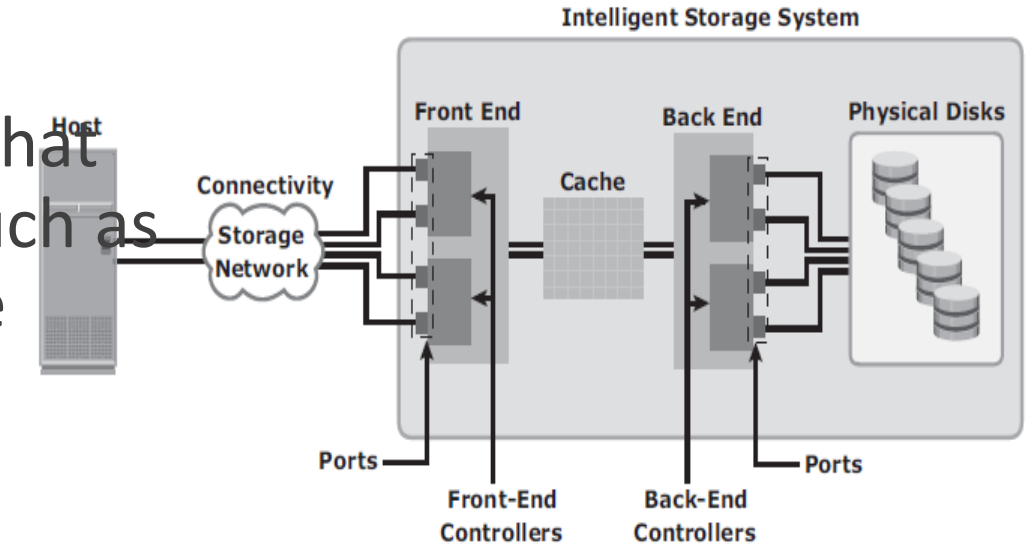
- The front end provides the interface between the storage system and the host.
- It consists of two components: front-end ports and front-end controllers.
- Typically, a front end has redundant controllers for high availability, and each controller contains multiple ports that enable large numbers of hosts to connect to the intelligent storage system.



Components of an Intelligent Storage System

Front end

- Front-end controllers route data to and from cache via the internal data bus.
- Each front-end controller has processing logic that executes the appropriate transport protocol, such as Fibre Channel, iSCSI, FICON, or FCoE for storage connections.



Components of an Intelligent Storage System

Cache

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

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.

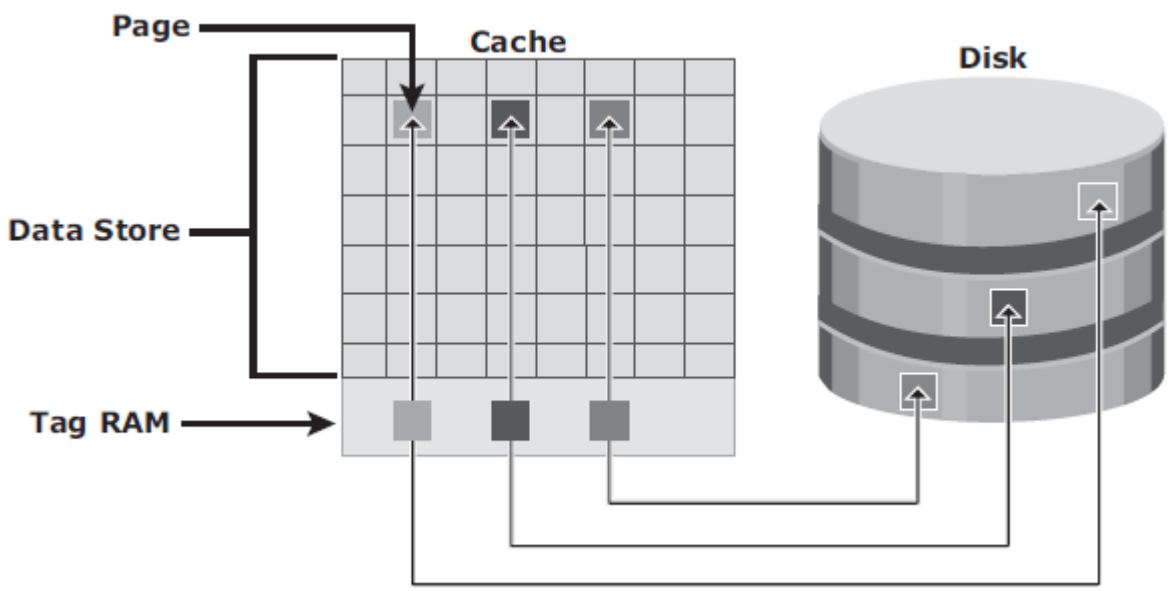
Components of an Intelligent Storage System

Structure of Cache

Cache is organized into pages, which is the smallest unit of cache allocation.

Cache consists of the

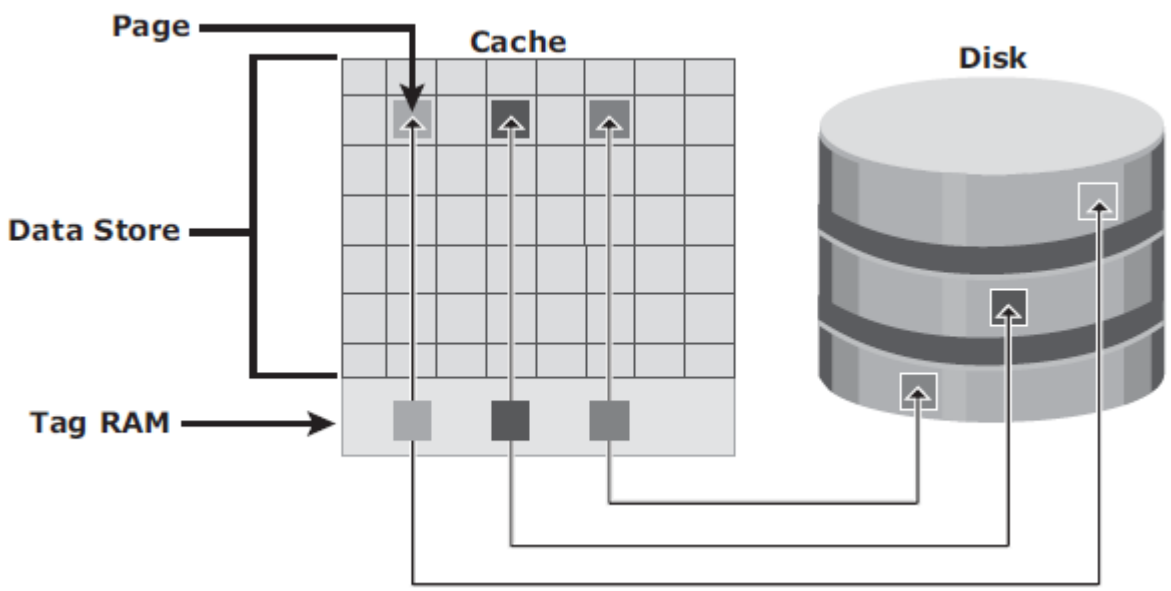
1. Data store - holds the data
2. Tag RAM - tracks the location of the data in the data store and in the disk.



Components of an Intelligent Storage System

Structure of Cache

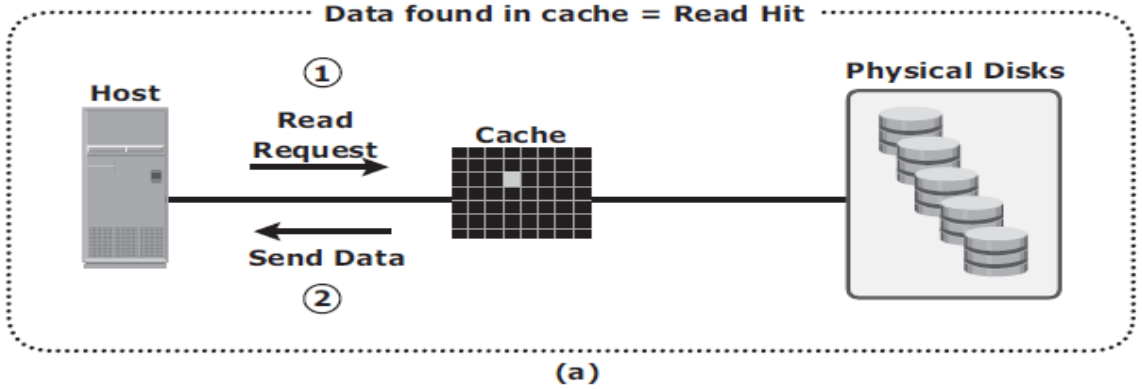
- Entries in tag RAM indicate where data is found in cache and where the data belongs on the disk.
- Tag RAM includes a dirty bit flag, which indicates whether the data in cache has been committed to the disk.
- It also contains time-based information, such as the time of last access, which is used to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



Components of an Intelligent Storage System

Read Operation with Cache

- 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
- Read performance is measured in terms of the read hit ratio(hit rate)
- Hit rate - ratio of the number of read hits with respect to the total number of read requests.

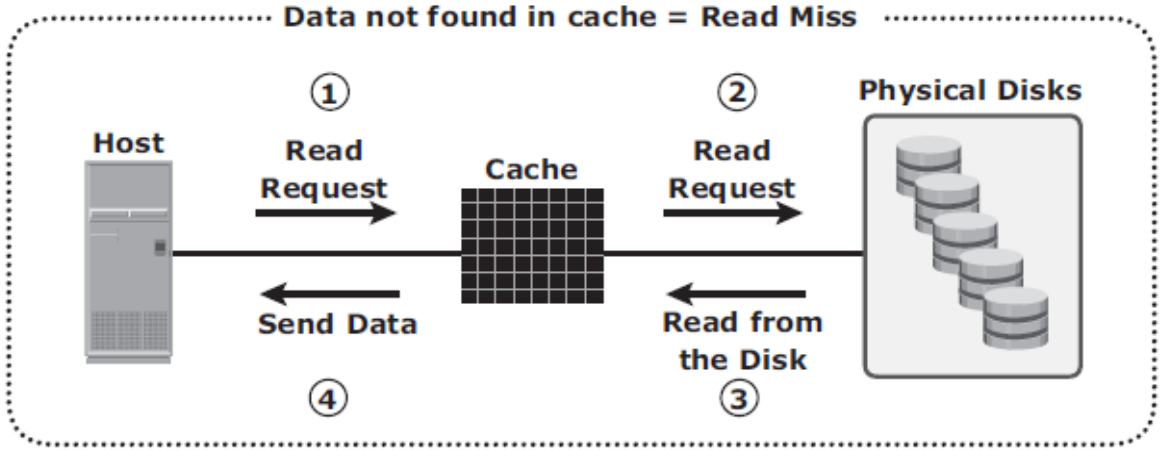


RAMAIAH Institute of Technology

Components of an Intelligent Storage System

Read Operation with Cache

- 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.



(b)

Components of an Intelligent Storage System

Read Operation with Cache

- Prefetch or Read-ahead algorithm
  - Fixed prefetch size
  - Variable prefetch size

Components of an Intelligent Storage System

Read Operation with Cache

- A prefetch or read-ahead algorithm is used when read requests are sequential.
- In a sequential read request, a contiguous set of associated blocks is retrieved.
- Several other blocks that have not yet been requested by the host can be read from the disk and placed into cache in advance.
- When the host subsequently requests these blocks, the read operations will be read hits.
- The intelligent storage system offers fixed and variable prefetch sizes.

Components of an Intelligent Storage System

Read Operation with Cache

A prefetch or read-ahead algorithm

- In fixed prefetch, the intelligent storage system prefetches a fixed amount of data.
- In variable prefetch, the storage system prefetches an amount of data in multiples of the size of the host request.

Components of an Intelligent Storage System

Write Operation with Cache

- Write operations with cache provide performance advantages over writing directly to disks.
- When the cache receives the write data, the controller sends an acknowledgment message back to the host.
- When an I/O is written to cache and acknowledged, it is completed in far less time (from the host's perspective) than it would take to write directly to disk.

Components of an Intelligent Storage System

Write Operation with Cache

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

Write-back cache:

Data is placed in cache and an acknowledgment is sent to the host immediately. Later, data from several writes are committed (de-staged) 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.

Components of an Intelligent Storage System

Write Operation with Cache

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

Write-through cache:

Data is placed in the cache and immediately written to the disk, and an acknowledgment is sent to the host.

Because data is committed to disk as it arrives, the risks of data loss are low, but the write-response time is longer because of the disk operations.

Components of an Intelligent Storage System

- Cache can be bypassed under certain conditions, such as large size write I/O.
- If the size of an I/O request exceeds the predefined size, called write aside size, writes are sent to the disk directly to reduce the impact of large writes consuming a large cache space.

Components of an Intelligent Storage System

Cache Implementation

- Cache can be implemented as either dedicated cache or global cache.
- With dedicated cache, separate sets of memory locations are reserved for reads and writes.
- In global cache, both reads and writes can use any of the available memory addresses

Components of an Intelligent Storage System

Cache Management

Even though modern intelligent storage systems come with a large amount of cache, when all cache pages are filled, some pages have to be freed up to accommodate new data and avoid performance degradation.

Components of an Intelligent Storage System

Cache Management

Various cache management algorithms are implemented in intelligent storage systems

Least Recently Used (LRU): An algorithm that continuously monitors data access in cache and identifies the cache pages that have not been accessed for a long time.

LRU either frees up these pages or marks them for reuse.

This algorithm is based on the assumption that data that has not been accessed for a while will not be requested by the host.

Components of an Intelligent Storage System

Cache Management

Various cache management algorithms are implemented in intelligent storage systems

Most Recently Used (MRU):

The pages that have been accessed most recently are freed up or marked for reuse.

This algorithm is based on the assumption that recently accessed data may not be required for a while.

Describe the different flushing techniques used to clear dirty pages when the cache becomes full to manage space availability.

Components of an Intelligent Storage System

Cache Management

As cache fills, the storage system must take action to flush dirty pages (data written into the cache but not yet written to the disk) to manage space availability.

Flushing is the process that commits data from cache to the disk.

Components of an Intelligent Storage System

Cache Management

- 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) is the cache utilization level at which the storage system starts high-speed flushing of cache data.
- Low watermark (LWM) is the point at which the storage system stops flushing data to the disks.

Components of an Intelligent Storage System

Cache Management

The cache utilization level, as shown in Figure 4-4, drives the mode of flushing to be used:

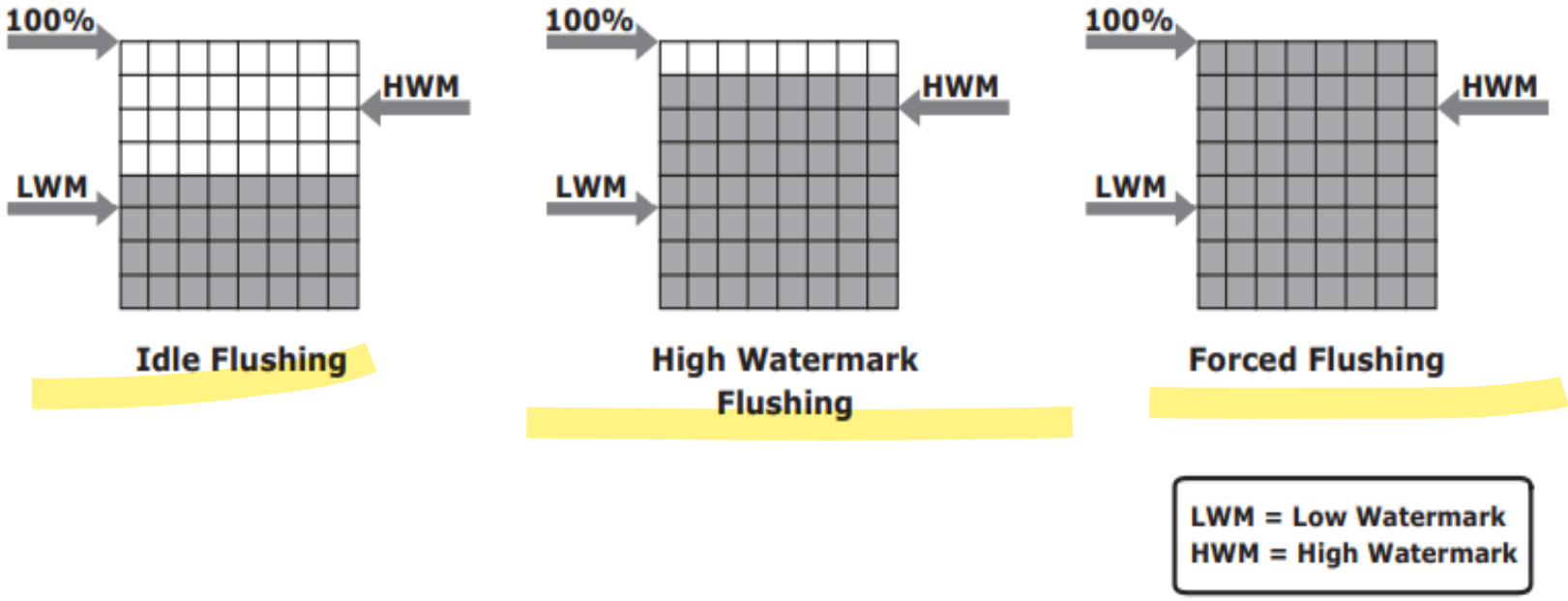
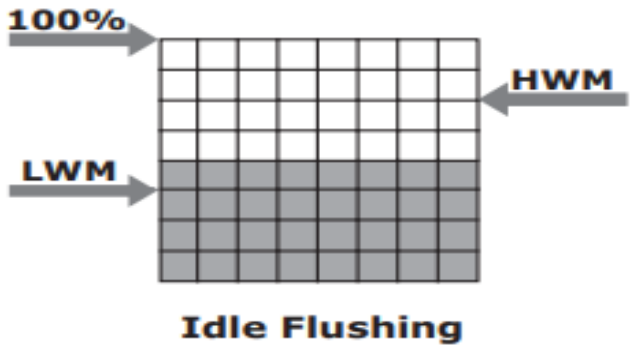


Figure 4-4: Types of flushing

Components of an Intelligent Storage System

Cache Management

Idle flushing: Occurs continuously, at a modest rate, when the cache utilization level is between the high and low watermark.

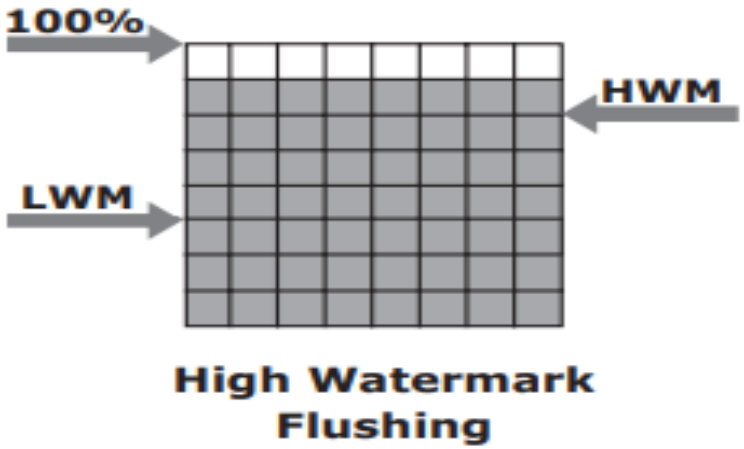


Idle Flushing

Components of an Intelligent Storage System

Cache Management

High watermark flushing: Activated when cache utilization hits the high watermark. The storage system dedicates some additional resources for flushing.



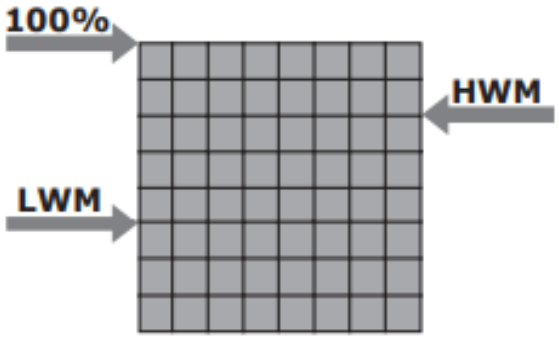
High Watermark Flushing

Components of an Intelligent Storage System

Cache Management

Forced flushing: Occurs in the event of a large I/O burst when cache reaches 100 percent of its capacity, which significantly affects the I/O response time.

In forced flushing, system flushes the cache on priority by allocating more resources.



Forced Flushing

Components of an Intelligent Storage System

Cache Data Protection

- Cache is volatile memory, so a power failure or any kind of cache failure will cause loss of the data that is not yet committed to the disk.
- This risk of losing uncommitted data held in cache can be mitigated using

1. cache mirroring
2. cache vaulting

Components of an Intelligent Storage System

Cache Data Protection

- **Cache mirroring:**
- Each write to cache is held in two different memory locations on two independent memory cards.
- If a cache failure occurs, the write data will still be safe in the mirrored location and can be committed to the disk.
- Reads are staged from the disk to the cache; therefore, if a cache failure occurs, the data can still be accessed from the disk.
- Because only writes are mirrored, this method results in better utilization of the available cache.

Components of an Intelligent Storage System

Cache Data Protection

- **Cache vaulting:**
- The risk of data loss due to power failure can be addressed in various ways: powering the memory with a battery until the AC power is restored or using battery power to write the cache content to the disk.
- If an extended power failure occurs, using batteries is not a viable option.

RAMAIAH
Institute of Technology

Components of an Intelligent Storage System

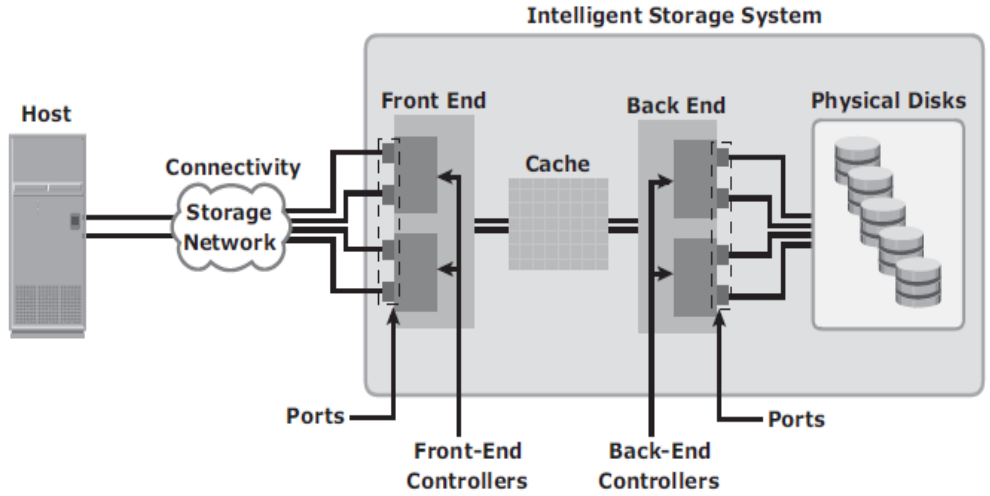
Back end

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.



RAMAIAH
Institute of Technology

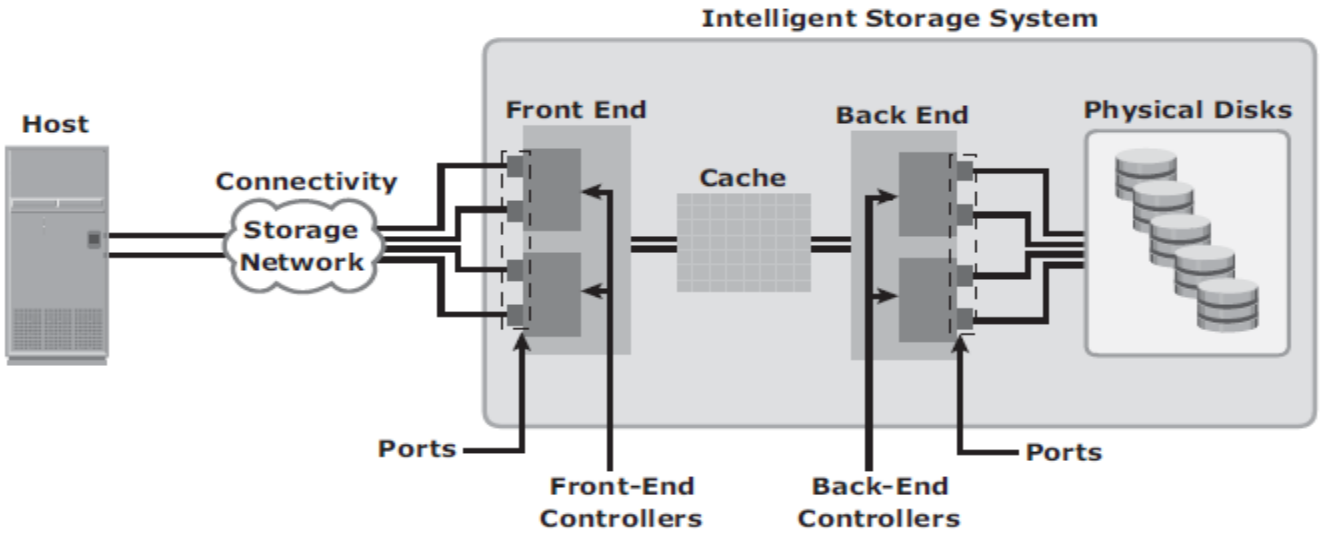
Components of an Intelligent Storage System

Back end

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 functionalit‘



RAMAIAH
Institute of Technology

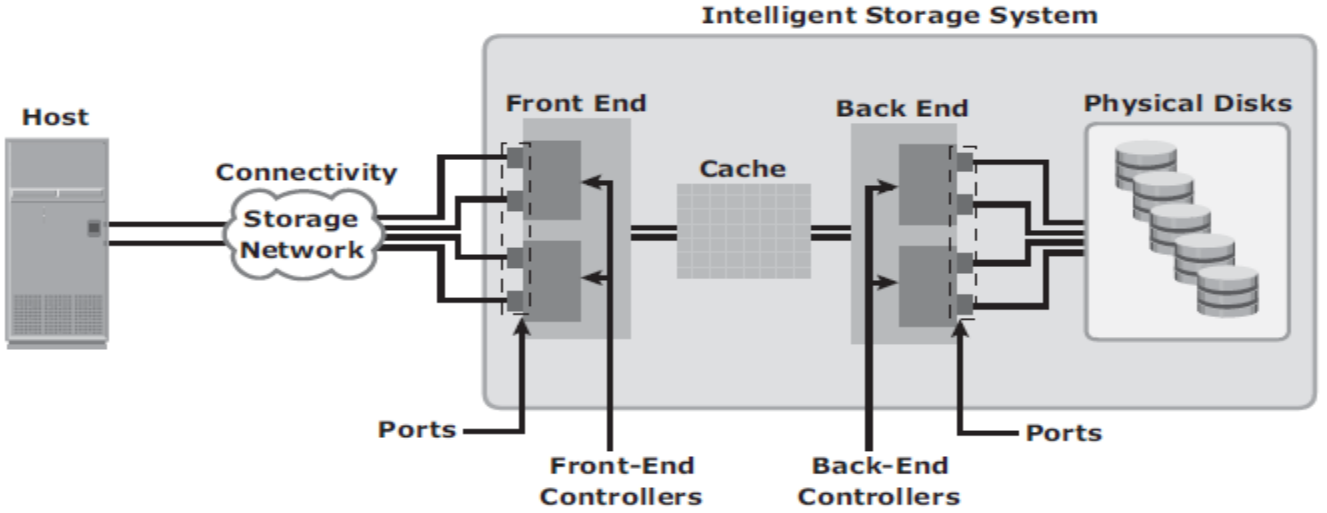
Components of an Intelligent Storage System

Physical disks

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, SATA, SAS, and flash drives.

They also support the use of a mix of flash. FC. or SATA within the same arrav.



RAMAIAH Institute of Technology

UNIT 2- Storage System

Intelligent Storage System(Chapter 4)

- Components of an Intelligent Storage System
- Storage Provisioning
- Types of Intelligent Storage Systems

Storage Provisioning

- Storage provisioning is the process of assigning storage resources to hosts based on capacity, availability, and performance requirements of applications running on the hosts.
- Storage provisioning can be performed in two ways: traditional and virtual.
- Virtual provisioning leverages virtualization technology for provisioning storage for applications.

Storage Provisioning

Traditional Storage Provisioning

Physical disks are logically grouped together and a required RAID level is applied to form a set, called a RAID set.

The number of drives in the RAID set and the RAID level determine the availability, capacity, and performance of the RAID set.

RAID set be created from drives of the same type, speed, and capacity to ensure maximum usable capacity, reliability, and consistency in performance

RAMAIAH Institute of Technology

Storage Provisioning

Traditional Storage Provisioning

Logical units are created from the RAID sets by partitioning (seen as slices of the RAID set) the available capacity into smaller units.

These units are then assigned to the host based on their storage requirements.

Each logical unit created from the RAID set is assigned a unique ID, called a logical unit number (LUN).

LUNs created by traditional storage provisioning methods are also referred to as thick LUNs to distinguish them from the LUNs created by virtual provisioning methods.

Storage Provisioning

Traditional Storage Provisioning

Figure 4-5 shows a RAID set consisting of five disks that have been sliced, or partitioned, into two LUNs:

1. LUN 0
2. LUN 1

These LUNs are then assigned to Host1 and Host 2 for their storage requirements.

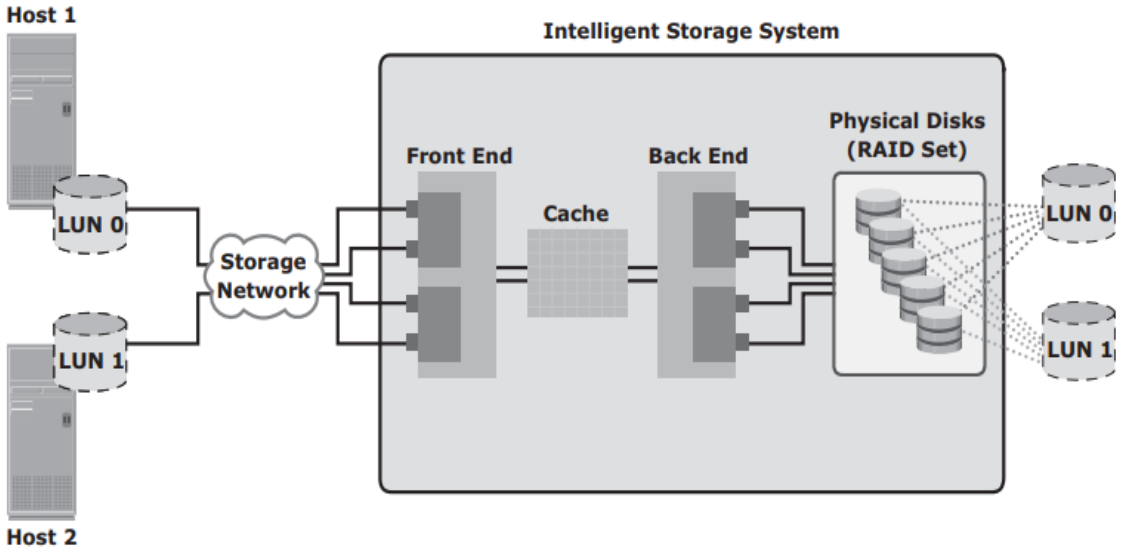


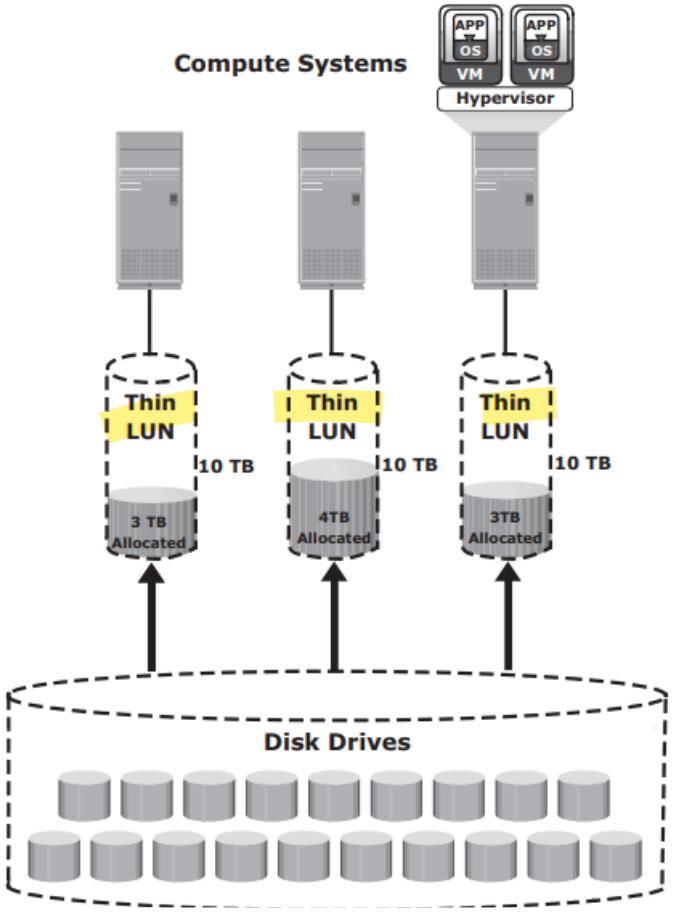
Figure 4-5: RAID set and LUNs

why Thin LUNs are suitable for applications that can tolerate performance variations.

Storage Provisioning

Virtual Storage Provisioning

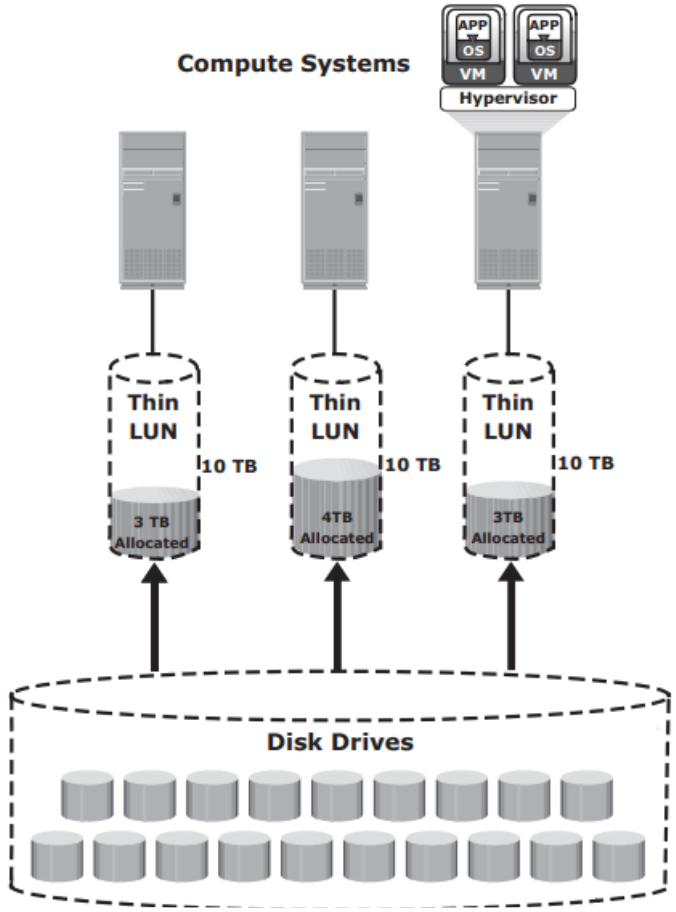
- Virtual provisioning enables creating and presenting a LUN with more capacity than is physically allocated to it on the storage array.
- The LUN created using virtual provisioning is called a thin LUN



Storage Provisioning

Virtual Storage Provisioning

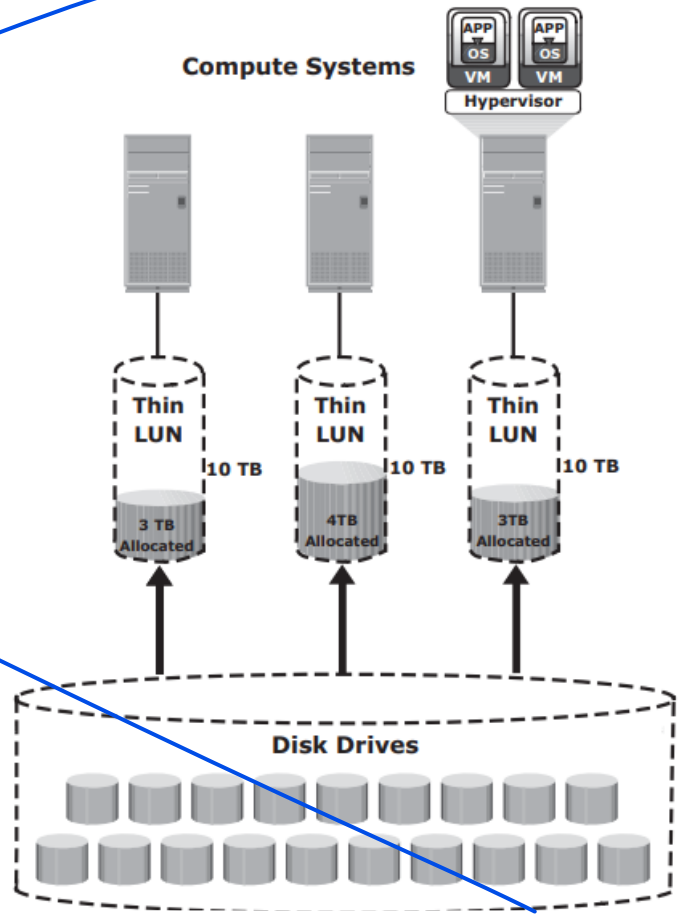
- Thin LUNs do not require physical storage to be completely allocated to them at the time they are created and presented to a host.
- Physical storage is allocated to the host "on-demand" from a shared pool of physical capacity.



Storage Provisioning

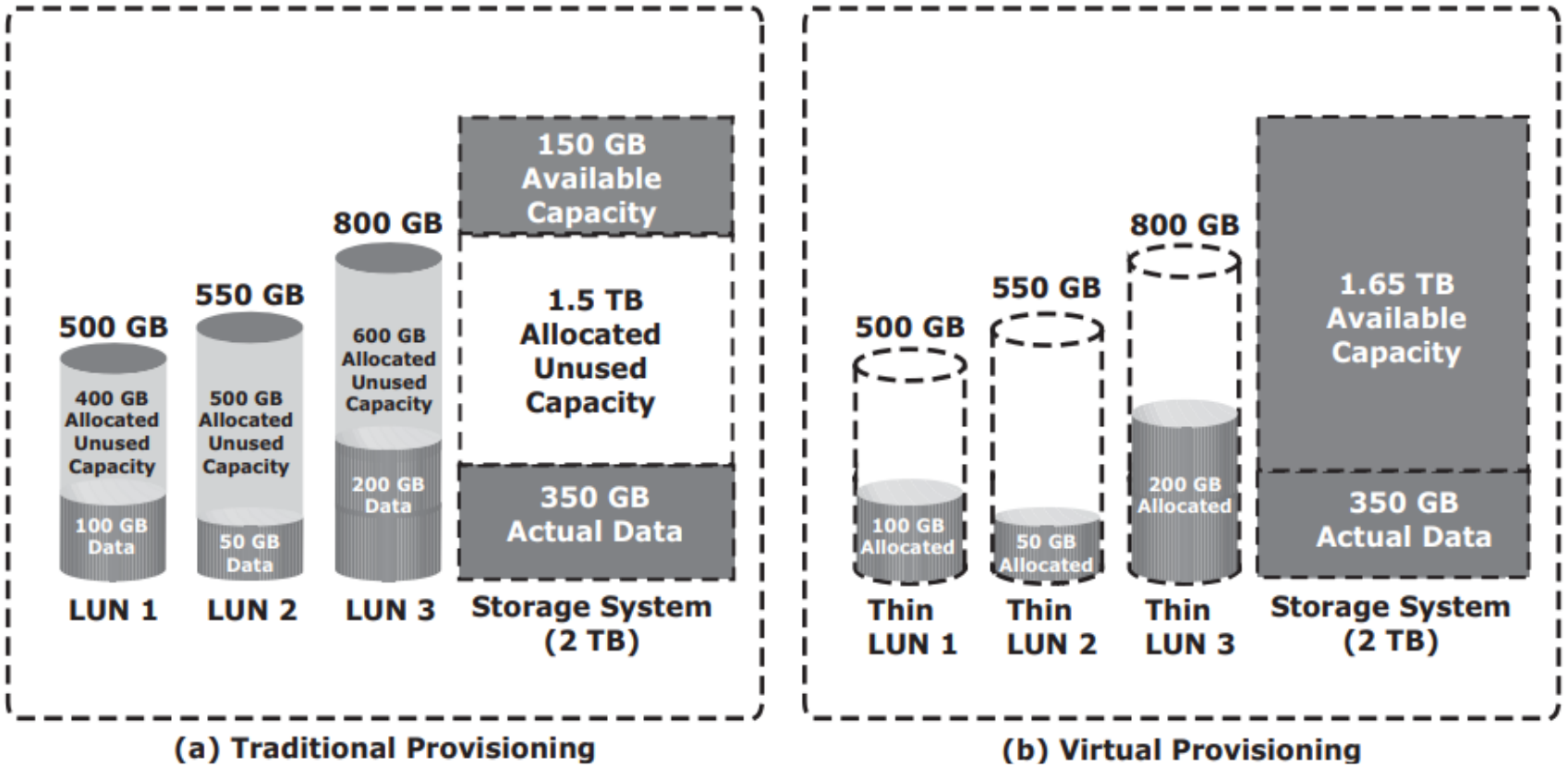
Virtual Storage Provisioning

- Virtual provisioning enables creating and presenting a LUN with more capacity than is physically allocated to it on the storage array.
- The LUN created using virtual provisioning is called a thin LUN



Storage Provisioning

Comparison between Virtual and Traditional Storage Provisioning



RAMAIAH Institute of Technology

UNIT 2- Storage System

Intelligent Storage System(Chapter 4)

- Components of an Intelligent Storage System
- Storage Provisioning
- Types of Intelligent Storage Systems

Types of Intelligent Storage Systems

Intelligent storage systems generally fall into one of the following two categories:

1. High-end storage systems(active-active configuration)
2. Midrange storage systems(active-passive configuration)

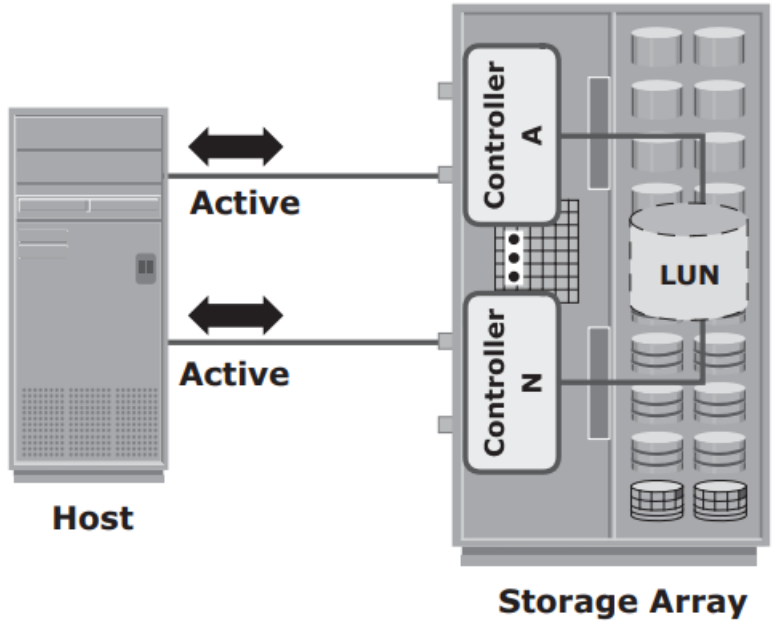
Types of Intelligent Storage Systems

High-end storage systems

High-end storage systems(Active-active arrays) are generally aimed at large enterprise applications.

These systems are designed with a large number of controllers and cache memory.

An active-active array implies that the host can perform I/Os to its LUNs through any of the available controllers

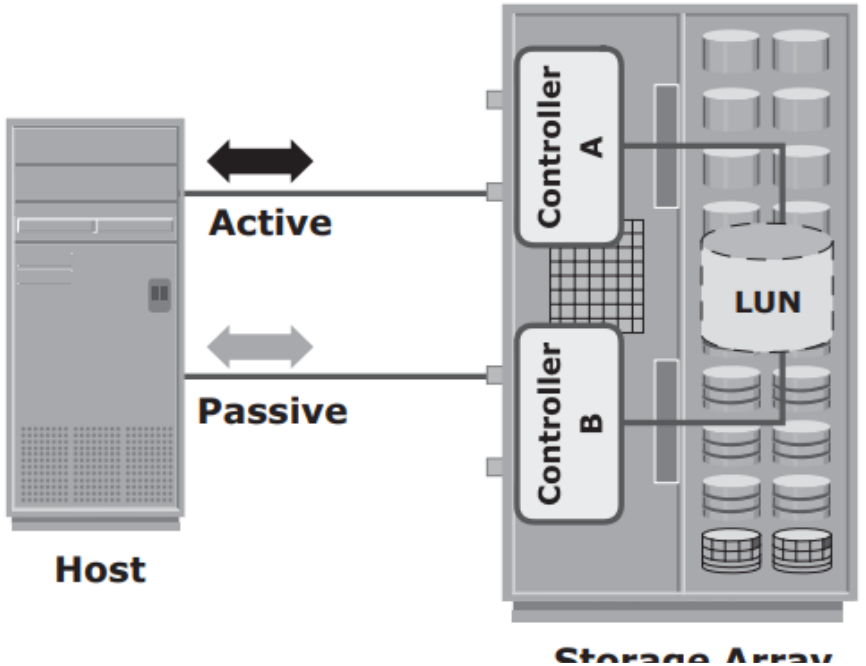


Types of Intelligent Storage Systems

Midrange Storage Systems

Midrange storage systems(Active-passive arrays)

In an active-passive array, a host can perform I/Os to a LUN only through the controller that owns the LUN.

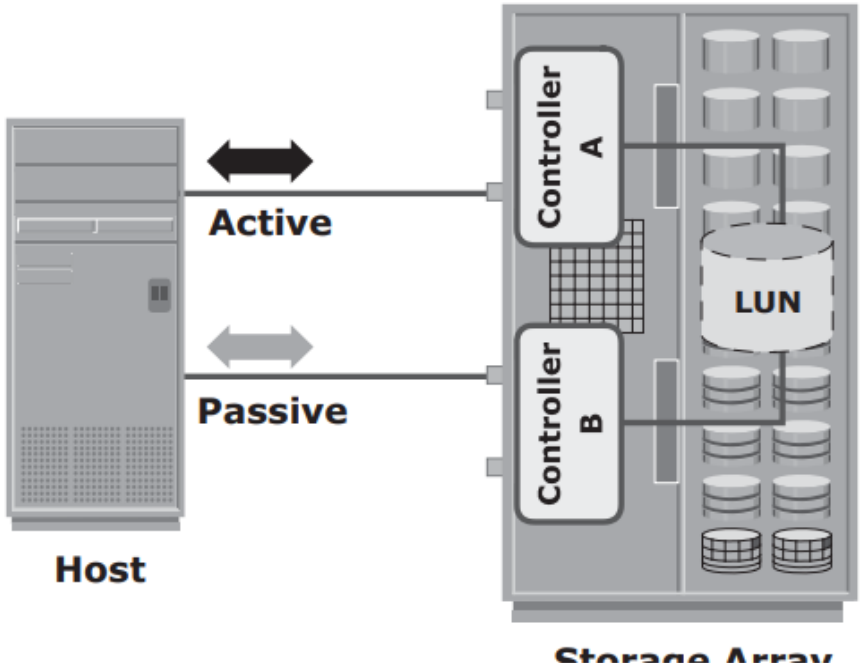


Types of Intelligent Storage Systems

Midrange Storage Systems

The host can perform reads or writes to the LUN only through the path to controller A because controller A is the owner of that LUN.

The path to controller B remains passive and no I/O activity is performed through this path.




Thank
you