

NIELIT Gorakhpur

Course Name: O Level (2nd Sem)

Subject: ICT

Topic: RAID

Date: 13-04-2020

RAID:- RAID is an acronym for **Redundant Array of Independent Disks**. With RAID enabled on a storage system you can connect two or more drives in the system so they act as one large volume fast drive or set them up as one system drive used to automatically and instantaneously duplicate (or mirror) your data for real-time backup. The term was coined by David Patterson, Garth A. Gibson, and Randy Katz at the University of California, Berkeley in 1987.

Data redundancy, although taking up extra space, adds to disk reliability. This means, in case of disk failure, if the same data is also backed up onto another disk, we can retrieve the data and go on with the operation. On the other hand, if the data is spread across just multiple disks without the RAID technique, the loss of a single disk can affect the entire data.

Important evaluation points for a RAID System:

- **Reliability:** How many disk faults can the system tolerate?
- **Availability:** What fraction of the total session time is a system in uptime mode, i.e. how available is the system for actual use?
- **Performance:** How good is the response time? How high is the throughput (rate of processing work)? Note that performance contains a lot of parameters and not just the two.
- **Capacity:** Given a set of N disks each with B blocks, how much useful capacity is available to the user?

Key concepts used in RAID

Mirroring: Mirroring is duplicating data to more than one disk. It can speed read times because the system can read data from more than one disk. But mirroring may slow write times if the system must confirm that data is correctly written to each disk.

Striping :Striping is writing data across a number of disks in parallel, which speeds read/write performance.

Parity: Parity error checking is where redundancy information is calculated for each piece of data stored. If a drive fails, the missing data can be reconstructed from the remaining data and the parity data. Error checking tends to slow the system because data from several locations must be read and compared.

RAID levels are defined by the combination of the techniques used; they also provide varying degrees of reliability (ability to withstand drive failure) and availability (speed of I/O). There are six basic RAID levels:

RAID LEVELS:

RAID 0

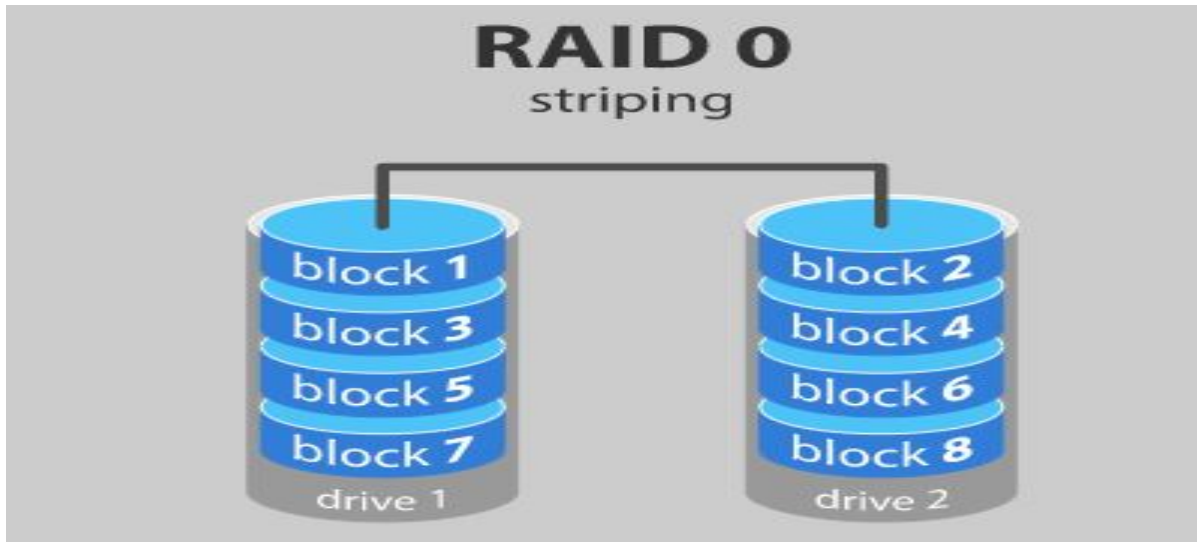
In a RAID 0 system data are split up into blocks that get written across all the drives in the array. By using multiple disks (at least 2) at the same time, this offers superior I/O performance. This performance can be enhanced further by using multiple controllers, ideally one controller per disk.

Advantages:

- RAID 0 offers great performance, both in read and write operations. There is no overhead caused by parity controls.
- All storage capacity is used, there is no overhead.
- The technology is easy to implement.

Disadvantages

- RAID 0 is not fault-tolerant. If one drive fails, all data in the RAID 0 array are lost. It should not be used for mission-critical systems.



RAID level 1 – Mirroring

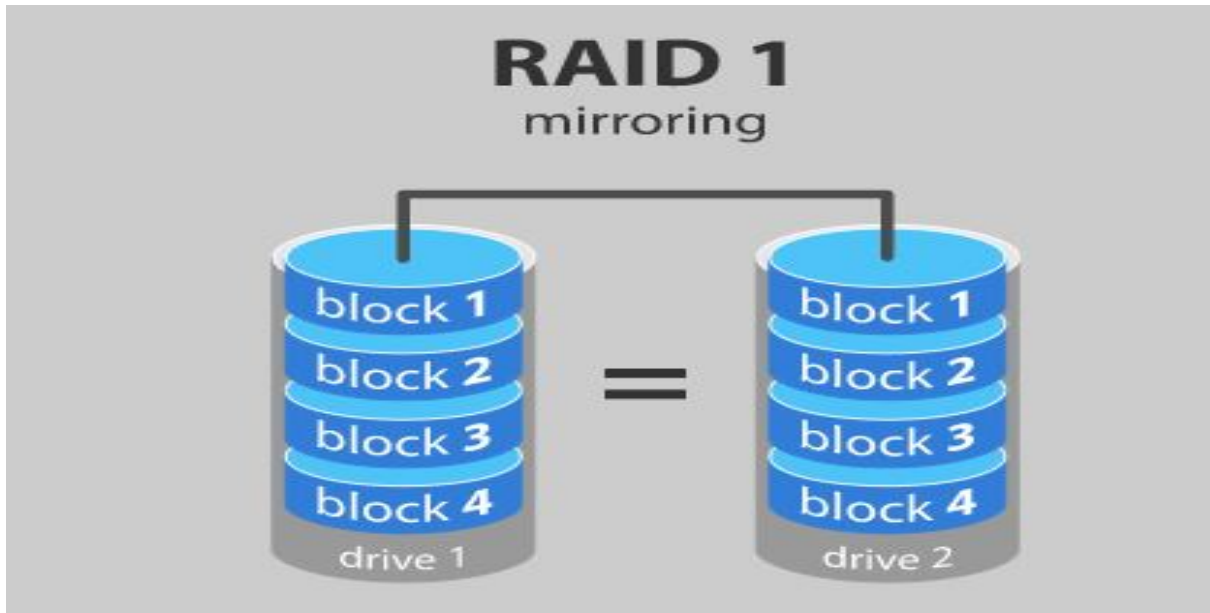
Data are stored twice by writing them to both the data drive (or set of data drives) and a mirror drive (or set of drives). If a drive fails, the controller uses either the data drive or the mirror drive for data recovery and continues operation. You need at least 2 drives for a RAID 1 array.

Advantages

- RAID 1 offers excellent read speed and a write-speed that is comparable to that of a single drive.
- In case a drive fails, data do not have to be rebuilt, they just have to be copied to the replacement drive.
- RAID 1 is a very simple technology.

Disadvantages

- The main disadvantage is that the effective storage capacity is only half of the total drive capacity because all data get written twice.



RAID level 5

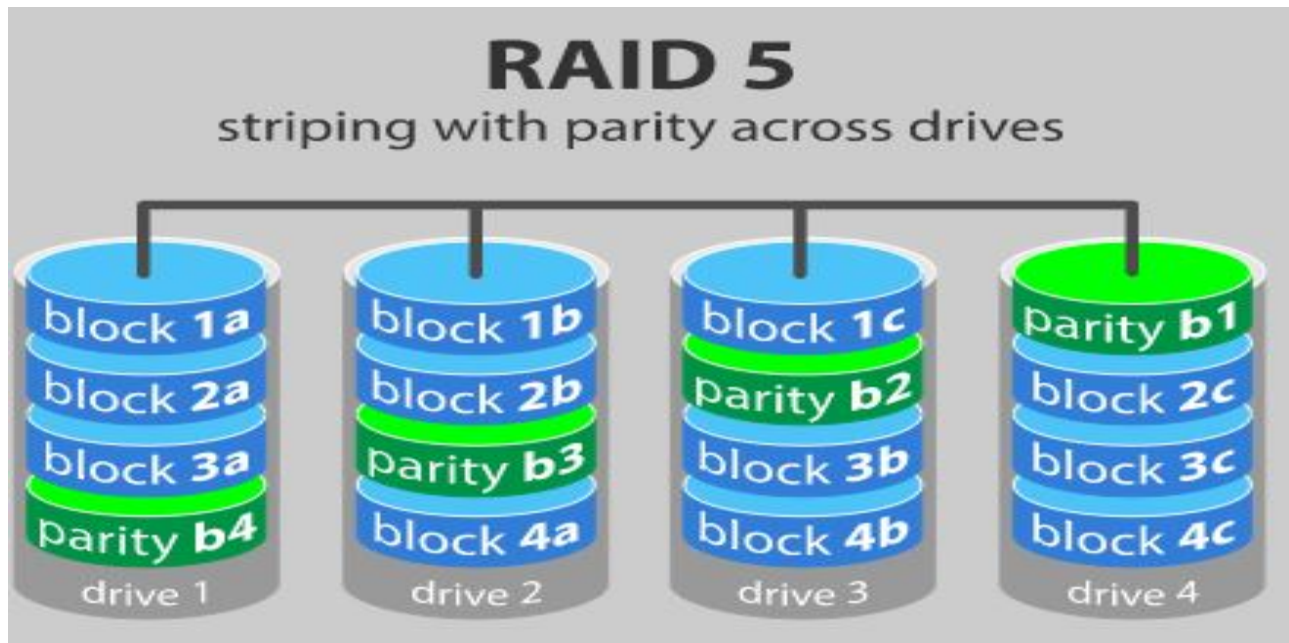
RAID 5 is the most common secure RAID level. It requires at least 3 drives but can work with up to 16. Data blocks are striped across the drives and on one drive a parity checksum of all the block data is written. The parity data are not written to a fixed drive, they are spread across all drives, as the drawing below shows. Using the parity data, the computer can recalculate the data of one of the other data blocks, should those data no longer be available. That means a RAID 5 array can withstand a single drive failure without losing data or access to data. Although RAID 5 can be achieved in software, a hardware controller is recommended. Often extra cache memory is used on these controllers to improve the write performance.

Advantages:

- Read data transactions are very fast while write data transactions are somewhat slower (due to the parity that has to be calculated).
- If a drive fails, you still have access to all data, even while the failed drive is being replaced and the storage controller rebuilds the data on the new drive.

Disadvantages:

- Drive failures have an effect on throughput, although this is still acceptable.
- This is complex technology. If one of the disks in an array using 4TB disks fails and is replaced, restoring the data (the rebuild time) may take a day or longer, depending on the load on the array and the speed of the controller. If another disk goes bad during that time, data are lost forever.



RAID level 6 – Striping with double parity:

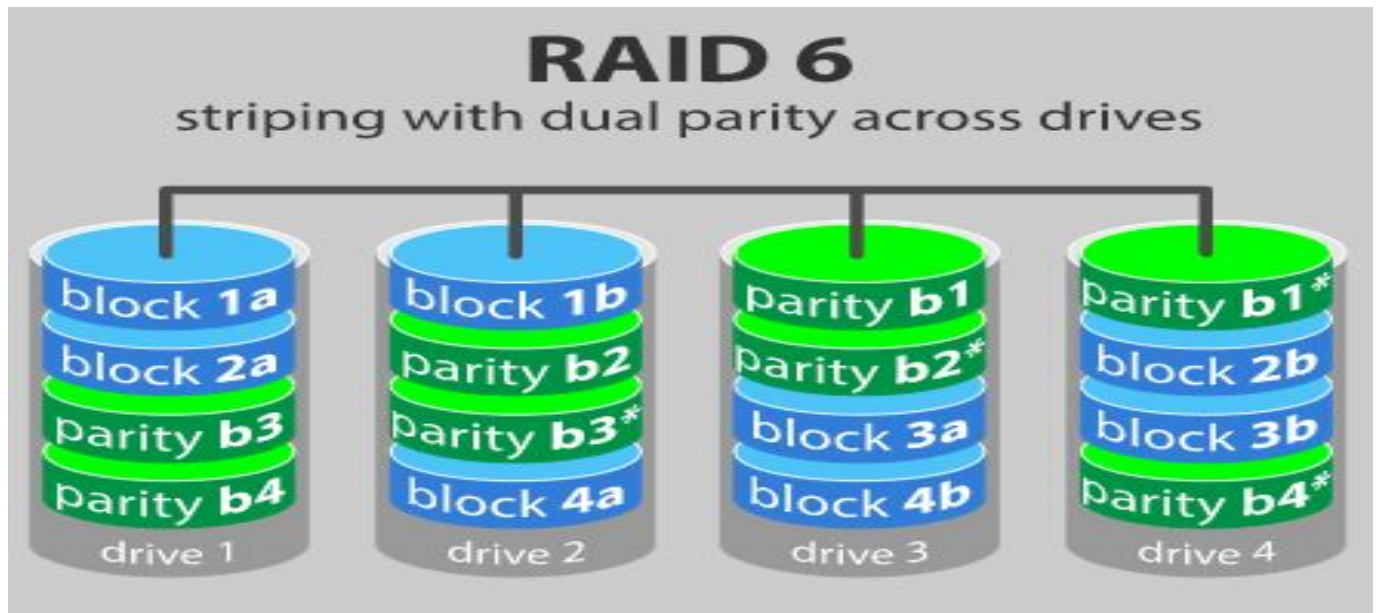
RAID 6 is like RAID 5, but the parity data are written to two drives. That means it requires at least 4 drives and can withstand 2 drives dying simultaneously. The chances that two drives break down at exactly the same moment are of course very small. However, if a drive in a RAID 5 systems dies and is replaced by a new drive, it takes hours or even more than a day to rebuild the swapped drive.

Advantages:

1. Like with RAID 5, read data transactions are very fast.
2. If two drives fail, you still have access to all data, even while the failed drives are being replaced. So RAID 6 is more secure than RAID 5.

Disadvantages:

- Write data transactions are slower than RAID 5 due to the additional parity data that have to be calculated. In one report I read the write performance was 20% lower.
- Drive failures have an effect on throughput, although this is still acceptable.



RAID level 10 – combining RAID 1 & RAID 0

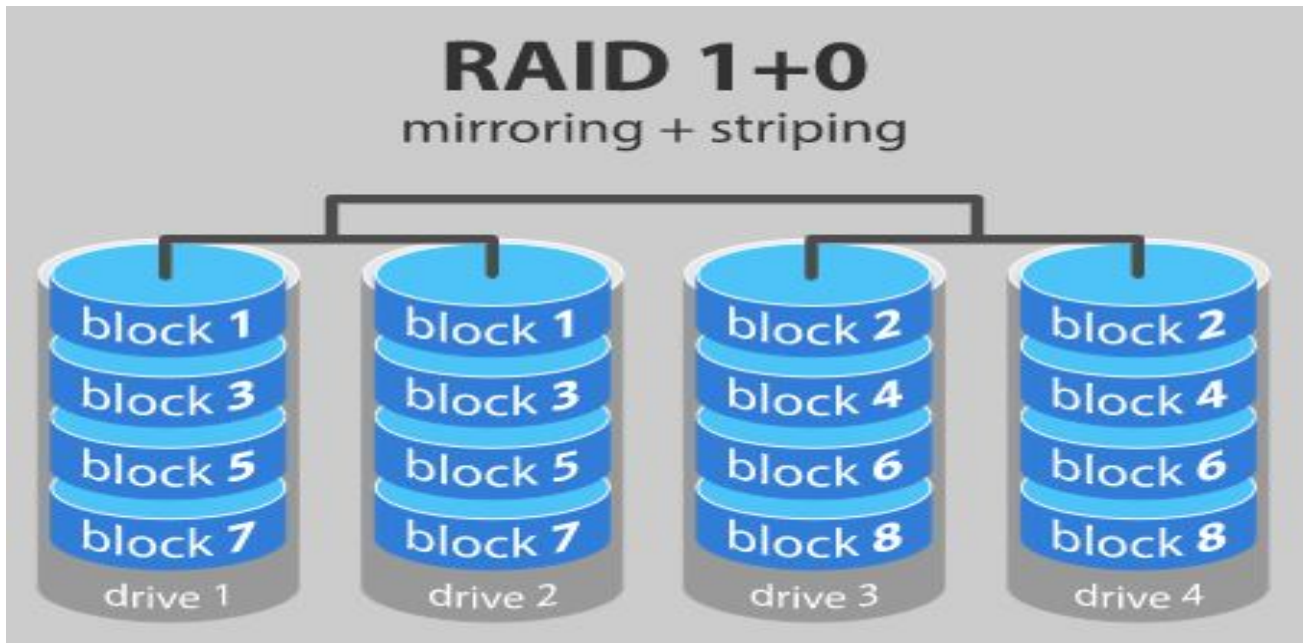
It is possible to combine the advantages (and disadvantages) of RAID 0 and RAID 1 in one single system. This is a nested or hybrid RAID configuration. It provides security by mirroring all data on secondary drives while using striping across each set of drives to speed up data transfers.

Advantages:

- If something goes wrong with one of the disks in a RAID 10 configuration, the rebuild time is very fast since all that is needed is copying all the data from the surviving mirror to a new drive. This can take as little as 30 minutes for drives of 1 TB.

Disadvantages:

- Half of the storage capacity goes to mirroring, so compared to large RAID 5 or RAID 6 arrays, this is an expensive way to have redundancy.



Exercise:

1. What is RAID? Explain.
2. Write the Difference between mirroring and striping