Disk Structure

Unix

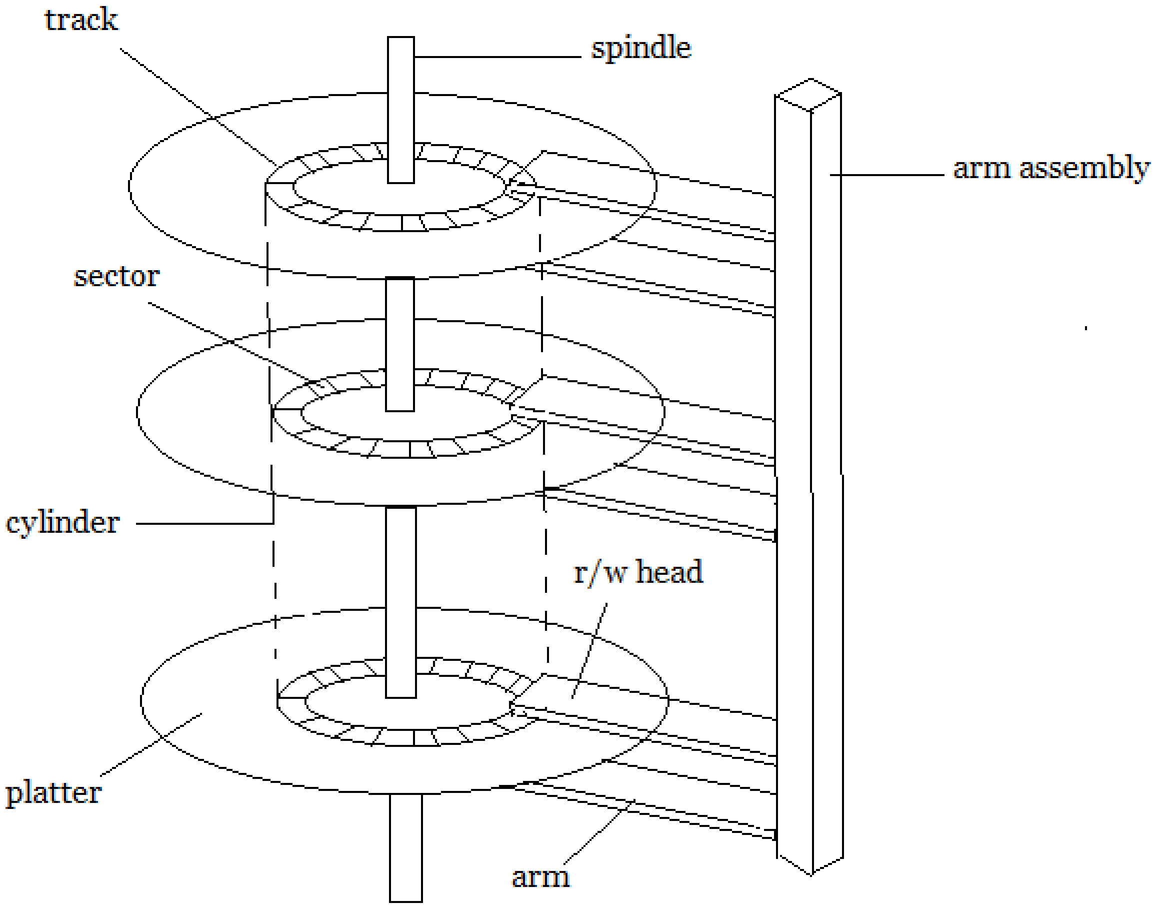
Secondary Storage

SOFTENGER
Strive to Excel

- Secondary storage devices are those devices whose memory is non volatile, meaning, the stored data will be intact even if the system is turned off. Here are a few things worth noting about secondary storage.
  - Secondary storage is less expensive when compared to primary memory like RAMs.
  - The speed of the secondary storage is also lesser than that of primary storage.
  - Hence, the data which is less frequently accessed is kept in the secondary storage.
  - A few examples are magnetic disks, magnetic tapes, etc.

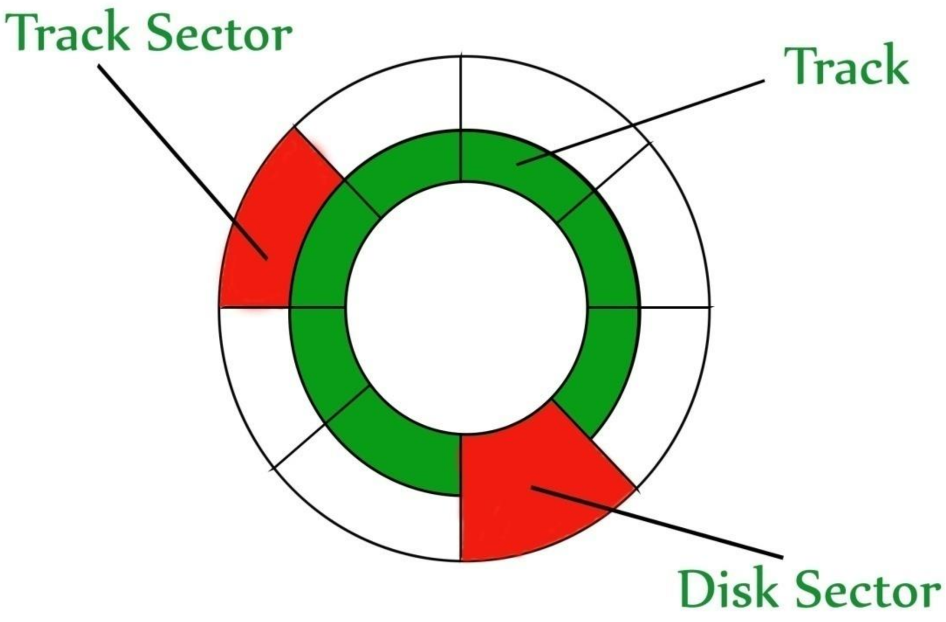
Magnetic Disk Structure

- Most of the secondary storage is in the form of magnetic disks
- A magnetic disk contains several **platters**. Each platter is divided into circular shaped **tracks**. The length of the tracks near the centre is less than the length of the tracks farther from the centre. Each track is further divided into **sectors**, as shown in the figure.
- Tracks of the same distance from centre form a cylinder. A read-write head is used to read data from a sector of the magnetic disk.



Disk Structure

- A hard disk looks like this:



- The disk is divided into tracks.
- Each track is further divided into sectors.
- Outer tracks are bigger in size than the inner tracks but they contain the same number of sectors and have equal storage capacity. This is because the storage density is high in sectors of the inner tracks where as the bits are sparsely arranged in sectors of the outer tracks.
- Some space of every sector is used for formatting. So, the actual capacity of a sector is less than the given capacity.
- Read-Write(R-W) head moves over the rotating hard disk. It is this Read-Write head that performs all the read and write operations on the disk and hence, position of the R-W head is a major concern.
- To perform a read or write operation on a memory location, we need to place the R-W head over that position.

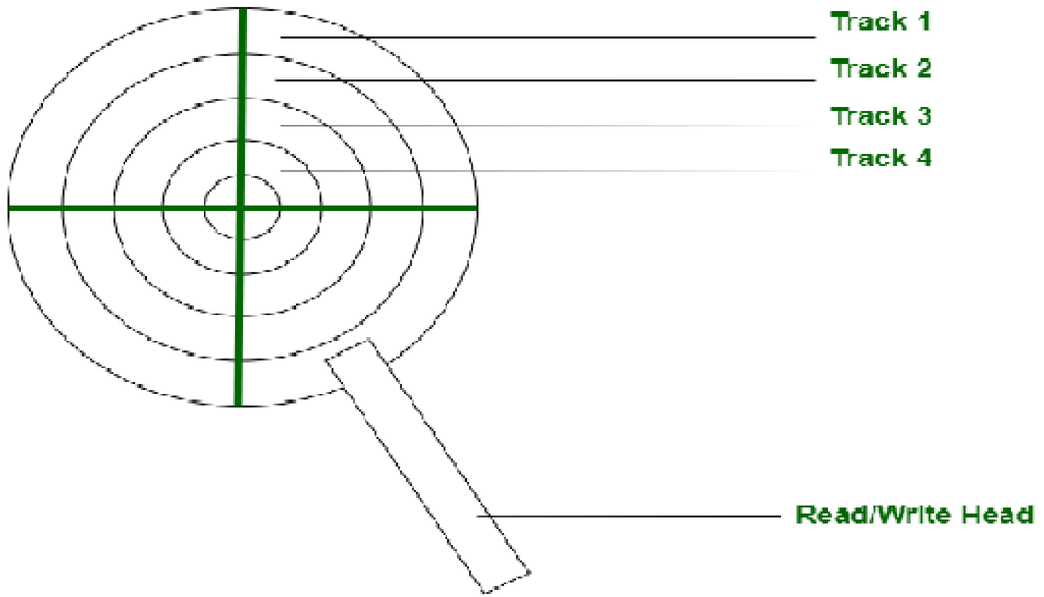
Disk Speed

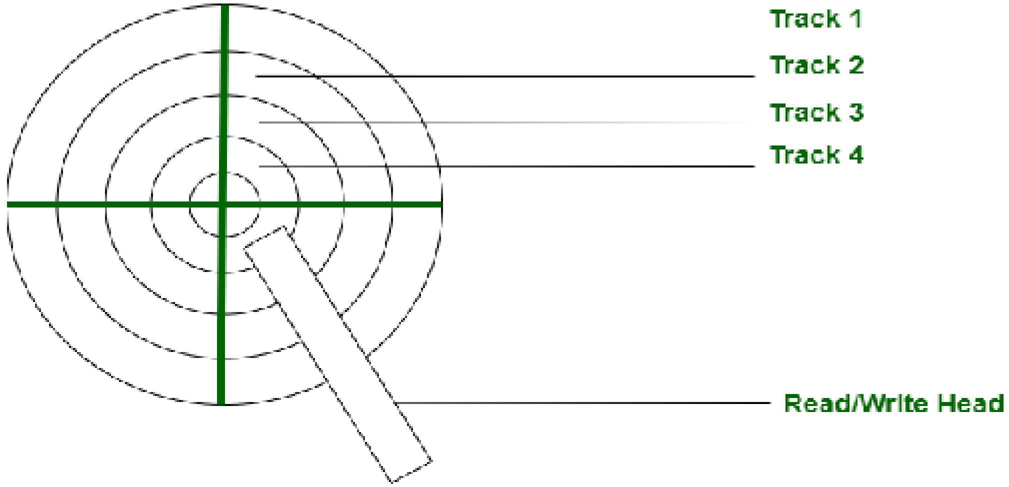
- The speed of the disk is measured as two parts:
  - **Transfer rate:** This is the rate at which the data moves from disk to the computer.
  - **Random access time:** It is the sum of the seek time and rotational latency.
    - **Seek time** is the time taken by the arm to move to the required track. **Rotational latency** is defined as the time taken by the arm to reach the required sector in the track.
    - Even though the disk is arranged as sectors and tracks physically, the data is logically arranged and addressed as an array of blocks of fixed size. The size of a block can be **512** or **1024** bytes. Each logical block is mapped with a sector on the disk, sequentially. In this way, each sector in the disk will have a logical address.

Seek Time and Rotational Latency

SOFTENGER
Strive to Excel

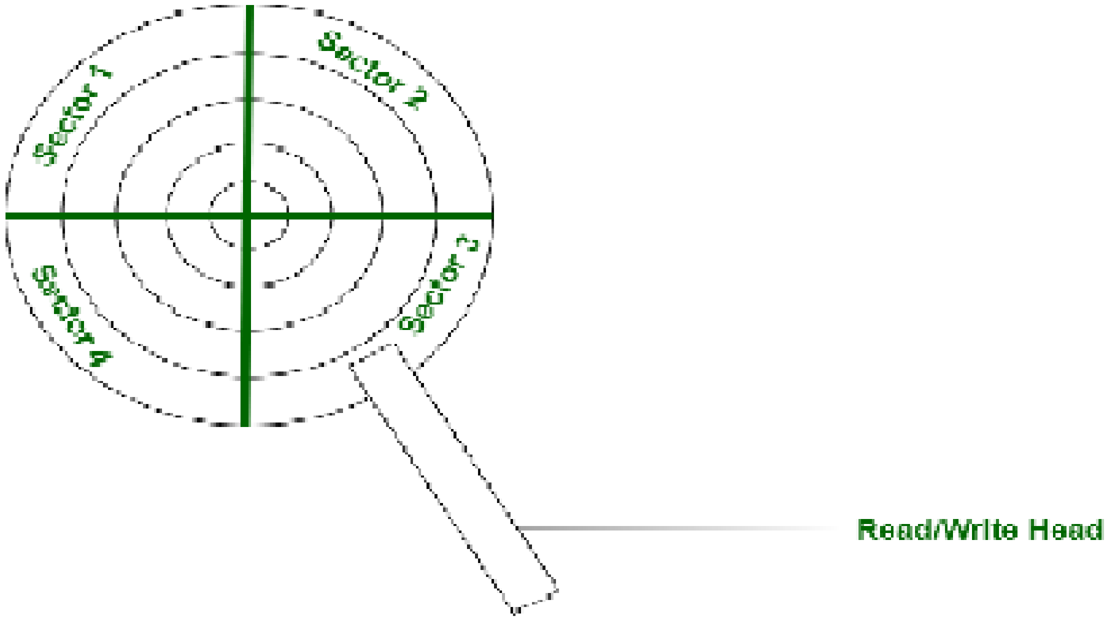
- **Seek Time:**
- The time taken by the R-W head to reach the desired track from it's current position.
- A disk is divided into many circular tracks. Seek Time is defined as the time required by the read/write head to move from one track to another.
- Example,
- Consider the following diagram 1 the read/write head is currently on track 1.
- Now, on the next diagram 2 read/write request, we may want to read data from Track 4, in this case, our read/write head will move to track 4. The time it will take to reach track 4 is the seek time.

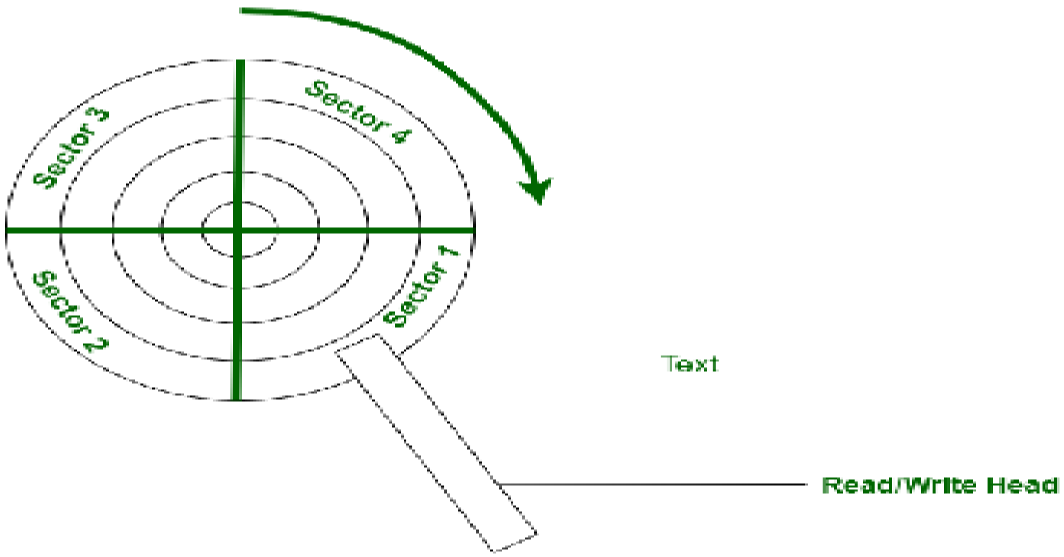




Rotational Latency:

- Time taken by the sector to come under the R-W head.
- The disk is divided into many circular tracks, and these tracks are further divided into blocks knows as sectors. The time required by the read/write head to rotate to the requested sector from the current position is called Rotational Latency.
- Example,
- Consider the diagram 1, We have divided each track into 4 sectors.
- The systems get a request to read a sector from track 1, thus the read/write head will move to track 1 and this time will be seek time.
- The read/write head is currently in sector 3.
- But the data may not be in sector 3. The data block may be present in sector 1. The time required by read/write head to move from sector 3 to sector 1 is the **rotational latency**.
-





Disk Access Time :

- Disk Access Time is the total time taken by the computer to process a data request from the processor and then transfer the data. When a program has to read information from the disk, the system should rotate the circle, the track and the sector on which the data is resides.
- Disk Access Time, = Seek time + Rotation time + Rotational Latency + Transfer time
- Disk Access Time includes –
- **Seek Time –**
  - It is the time taken by the read/write head to reach to the desired output. It is known to be the most important time because it cannot create a gap. Seek time is inversely proportional to the performance. Lesser the seek time, the more good will be the performance. Average seek time, = (1/3) * time taken for one full stroke

Disk Access Time :

SOFTENGER
Strive to Excel

- **Rotation Time –**
- It is the total time taken for one complete rotation (360 degree). For reaching to the desired sector, the disk will rotate by spindle in either clockwise direction or anti-clockwise direction, but can be move in only one direction at a time.
- **Rotational Latency –**
- The amount of time taken by the disk to rotate the track when the read/write head comes to exact sector. Average rotational latency, = (1/2) * one rotation time . **Note:**Average Rotational latency is mostly 1/2*(Rotational latency).
- **Data Transfer Time –**
- It is the time taken to transfer the data. Data transfer time, = (Data to be transfer) / (Transfer rate)

RAID

- RAID is a technology that is used to increase the performance and/or reliability of data storage.
- The abbreviation stands for either *Redundant Array of Inexpensive Disks* or *Redundant Array of Independent Drives*.
- A RAID system consists of two or more drives working in parallel. These can be hard discs, but there is a trend to also use the technology for SSD (Solid State Drives).
- There are different RAID levels, each optimized for a specific situation. These are not standardized by an industry group or standardization committee.

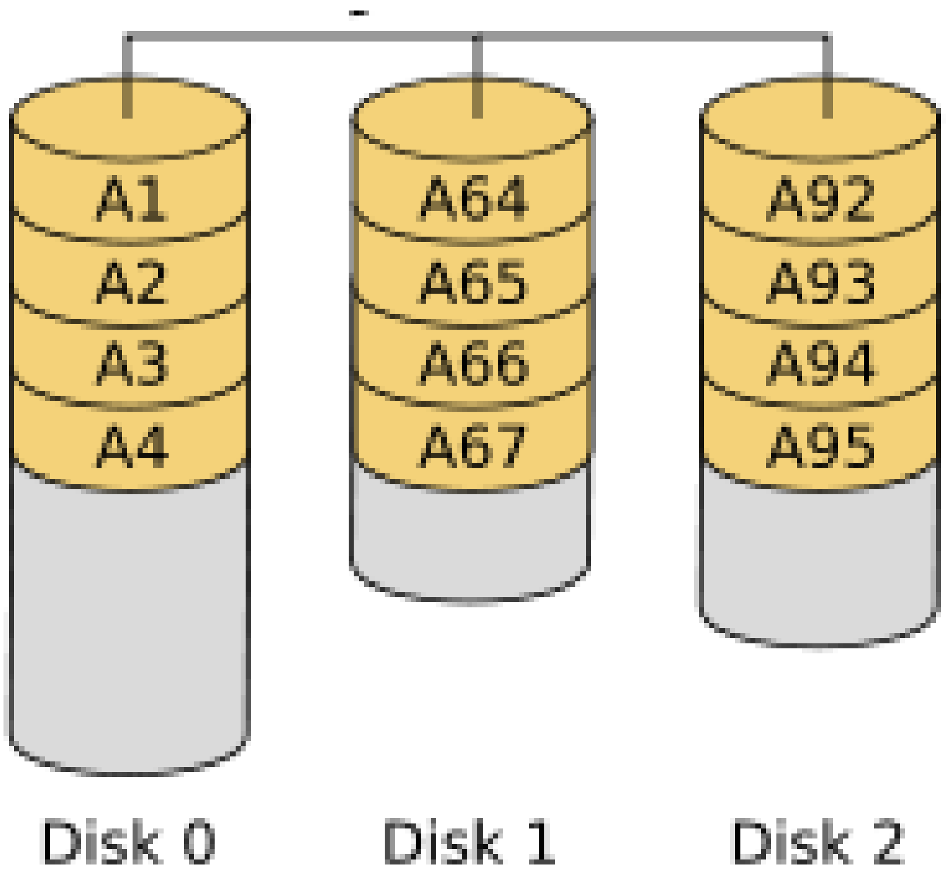
RAID Levels

- There are following RAID levels mostly used:
  - RAID 0 – concatenation
  - RAID 1 – striping
  - RAID 2 – Mirroring
  - RAID 4 – dedicated disk for parity
  - RAID 5 – stripping with parity across drives
  - RAID 6 – Multiple circular parity Disk
  - RAID 10- combining mirroring + stripping
- The software to perform the RAID-functionality and control the drives can either be located on a separate controller card (a hardware RAID controller) or it can simply be a driver.
- Some versions of Windows, such as Windows Server 2012 as well as Mac OS X, include software RAID functionality.
- Hardware RAID controllers cost more than pure software, but they also offer better performance, especially with RAID 5 and 6.

RAID 0 - Concatenation

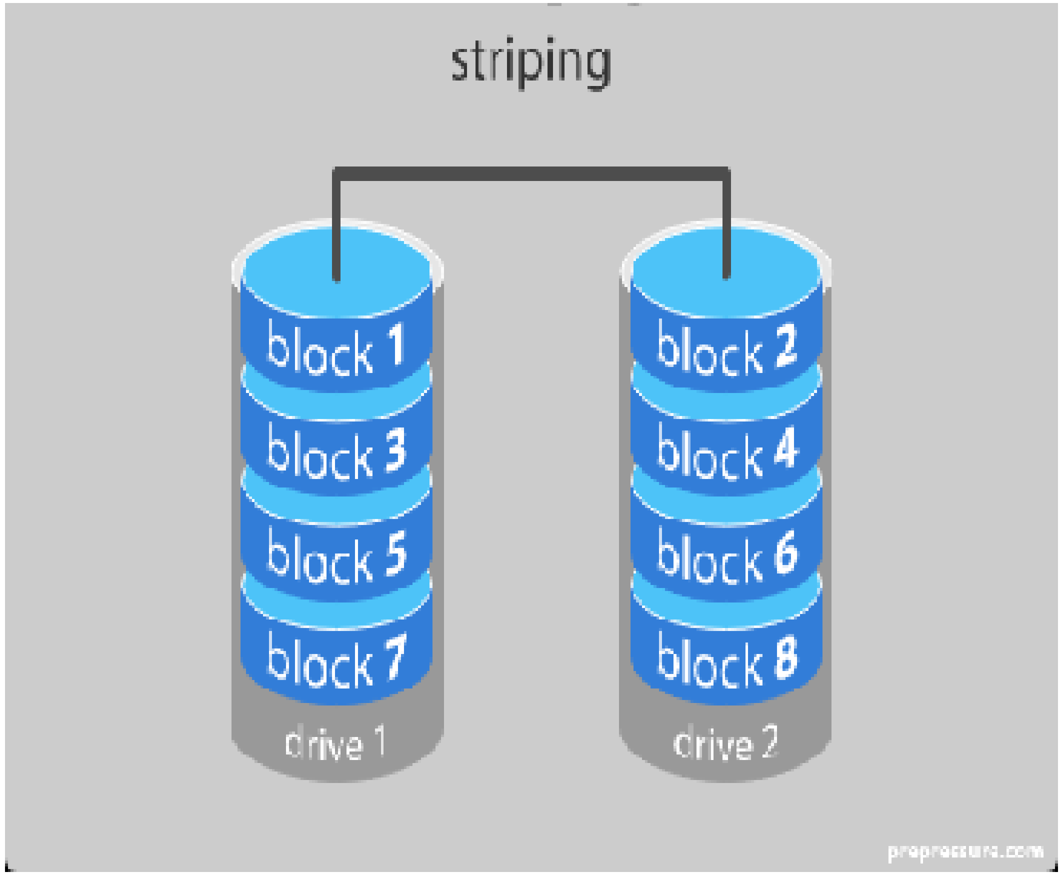
- Concatenation or spanning of drives is not one of the numbered RAID levels, but it is a popular method for combining multiple physical disk drives into a single logical disk.

It provides no data redundancy. Drives are merely concatenated (linking things together in a series) together, end to beginning, so they appear to be a single large disk.



RAID level 1 – Striping

- In a RAID 1 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.

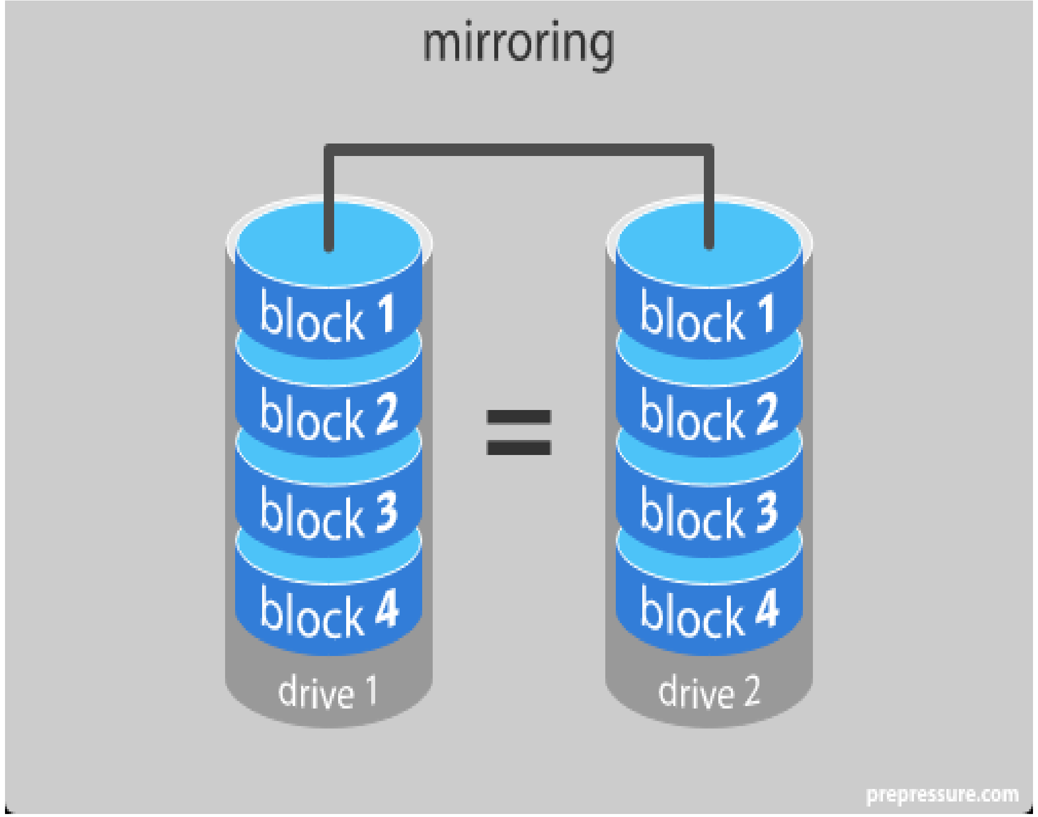


RAID level 1 – Striping

- **Advantages**
  - RAID 1 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 1 is not fault-tolerant. If one drive fails, all data in the RAID 1 array are lost. It should not be used for mission-critical systems.
- **Ideal use**
  - RAID 1 is ideal for non-critical storage of data that have to be read/written at a high speed, such as on an image retouching or video editing station.
  - If you want to use RAID 1 purely to combine the storage capacity of tween drives in a single volume, consider mounting one drive in the folder path of the other drive. This is supported in Linux, OS X as well as Windows and has the advantage that a single drive failure has no impact on the data of the second disk or SSD drive.

RAID level 2 – 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 2 array.



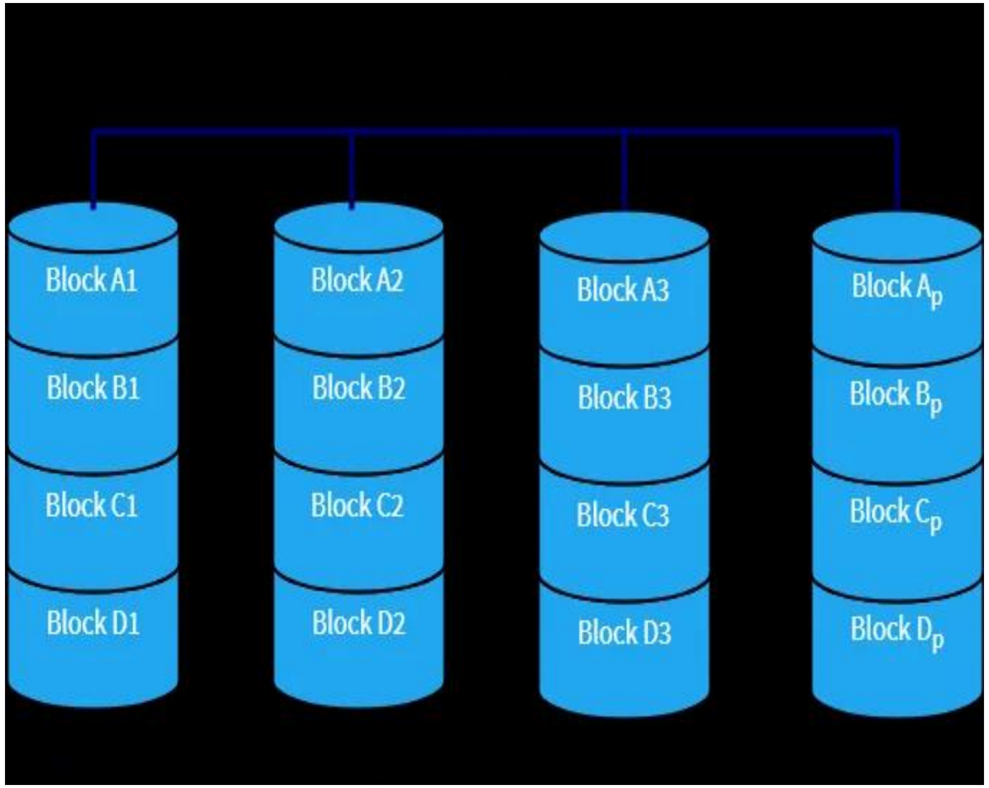
RAID level 2 – Mirroring

SOFTENGER
Strive to Excel

- **Advantages**
  - RAID 2 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 rebuild, they just have to be copied to the replacement drive.
  - RAID 2 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.
  - Software RAID 2 solutions do not always allow a hot swap of a failed drive. That means the failed drive can only be replaced after powering down the computer it is attached to. For servers that are used simultaneously by many people, this may not be acceptable. Such systems typically use hardware controllers that do support hot swapping.
- **Ideal use**
  - RAID-2 is ideal for mission critical storage, for instance for accounting systems. It is also suitable for small servers in which only two data drives will be used.

RAID level 4(Dedicated disk for parity)

- RAID 4 stripes the data across multiple disks just like RAID 0. In addition to that, it also stores parity information of all the disks **in a separate dedicated disk** to achieve redundancy.
- In the diagram, Disk 4 serves as the parity disk having parity blocks Ap, Bp, Cp and Dp. So, if one of the disks fails, the data can be reconstructed using the parity information of that disk.
- Space is more efficiently used here when compared to RAID 2 since parity information uses way less space than mirroring the disk. The write performance becomes slow because all the parity information is written on a single disk which is a bottleneck. This problem is solved in RAID 5 as we will see next.

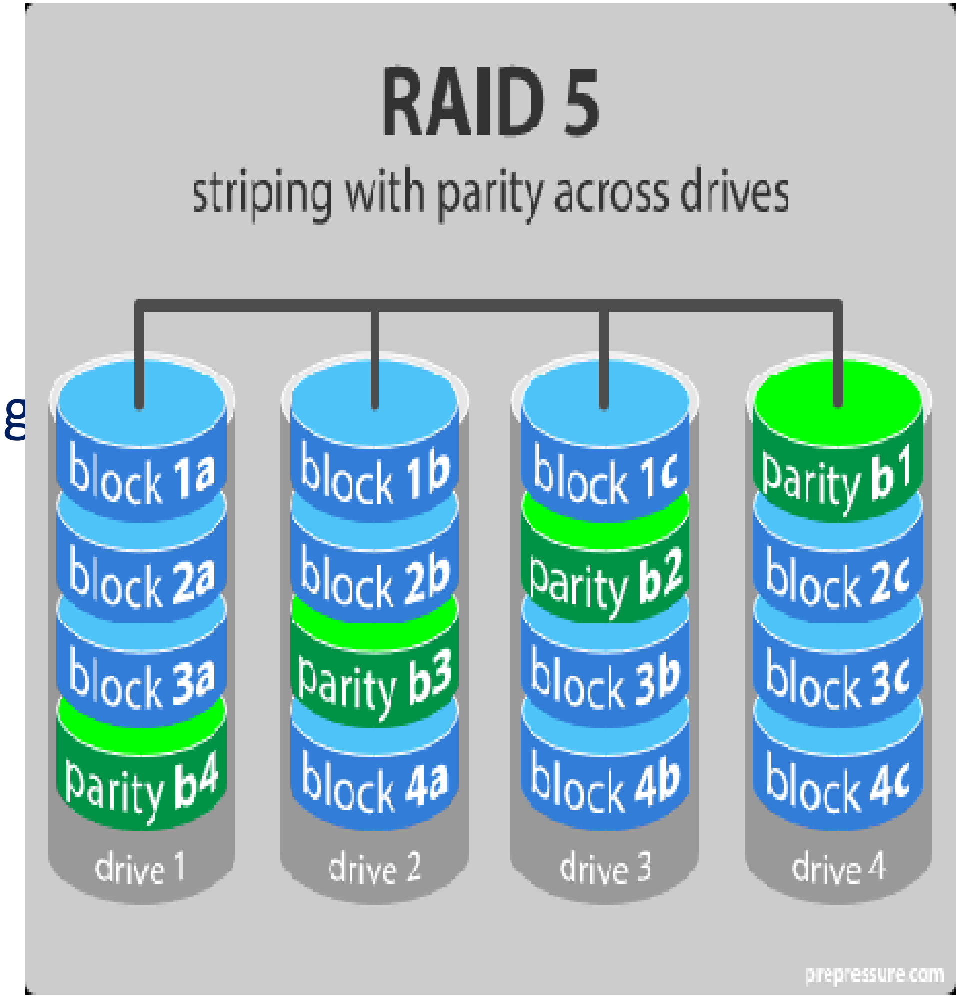


RAID 4

- **Advantages**
  - Efficient data redundancy in terms of cost per unit memory
  - Performance boost for read operations due to data stripping
- **Disadvantages**
  - Write operation is slow
  - If the dedicated parity disk fails, data redundancy is lost.

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.

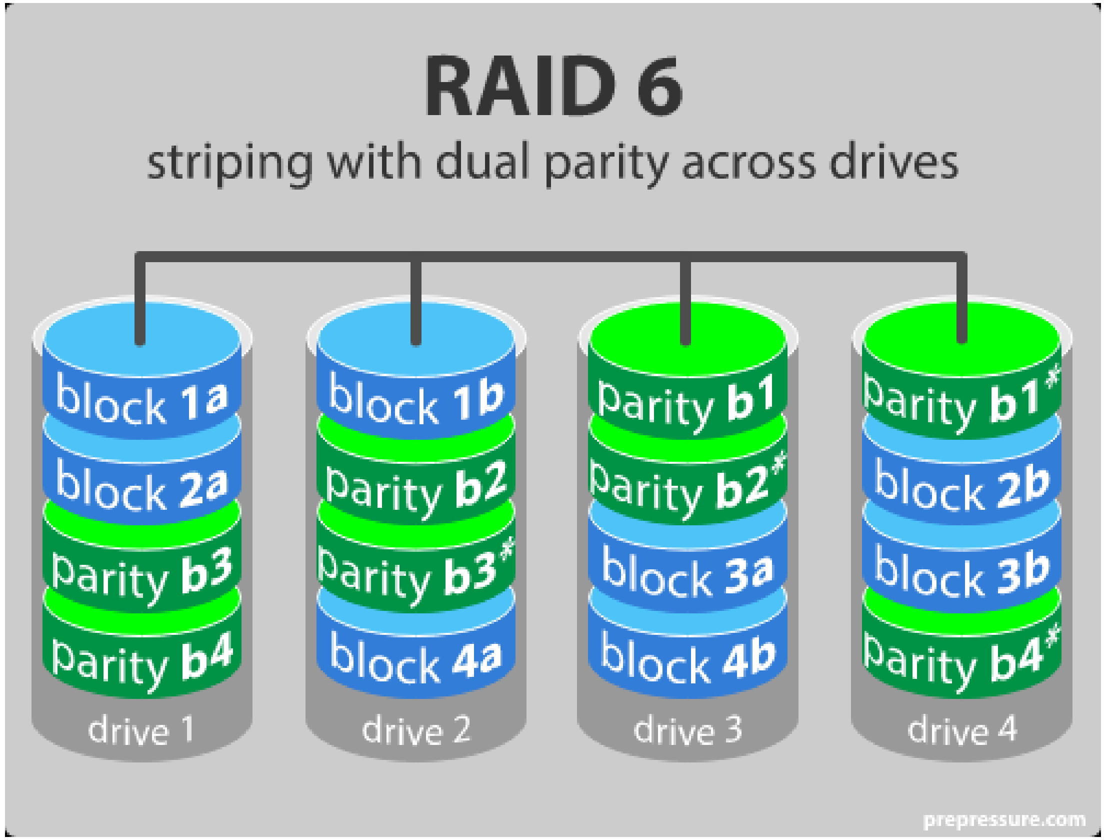


RAID 5

- 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.
- Ideal use
  - RAID 5 is a good all-round system that combines efficient storage with excellent security and decent performance. It is ideal for file and application servers that have a limited number of data drives.

RAID 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.
- If another drive dies during that time, you still lose all of your data. With RAID 6, the RAID array will even survive that second failure.

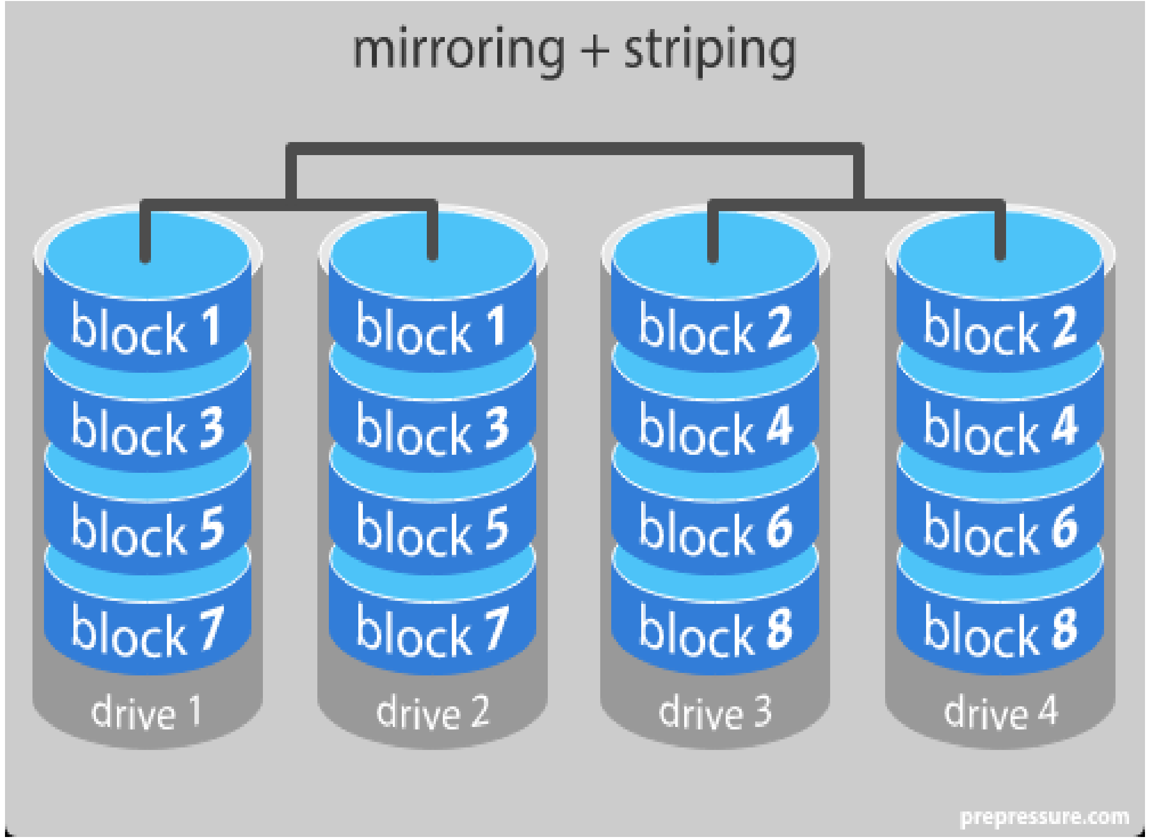


RAID 6

- Advantages
  - Like with RAID 5, read data transactions are very fast.
  - 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. The write performance can drop as much as 20% lower.
  - Drive failures have an effect on throughput, although this is still acceptable.
  - This is complex technology. Rebuilding an array in which one drive failed can take a long time.
- Ideal use
  - RAID 6 is a good all-round system that combines efficient storage with excellent security and decent performance. It is preferable over RAID 5 in file and application servers that use many large drives for data storage.

RAID level 10 – combining mirroring + stripping

- It is possible to combine the advantages (and disadvantages) of mirroring and stripping 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.



RAID 10

- **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.

RAID 2, 3, 4 and 7

- What about RAID levels 2, 3, 4 and 7?
- These levels do exist but are not that common (RAID 3 is essentially like RAID 5 but with the parity data always written to the same drive).