What is RAID (Redundant Array of Independent

Introduction to RAID

RAID is an acronym for **Redundant Array of Independent (or Inexpensive) Disks**. **RAID** is the way of combining several independent and relatively small disks into a single storage of a large size. The disks included into the array are called **Array Members**. The disks can be combined into the array in different ways which are known as **RAID levels.**

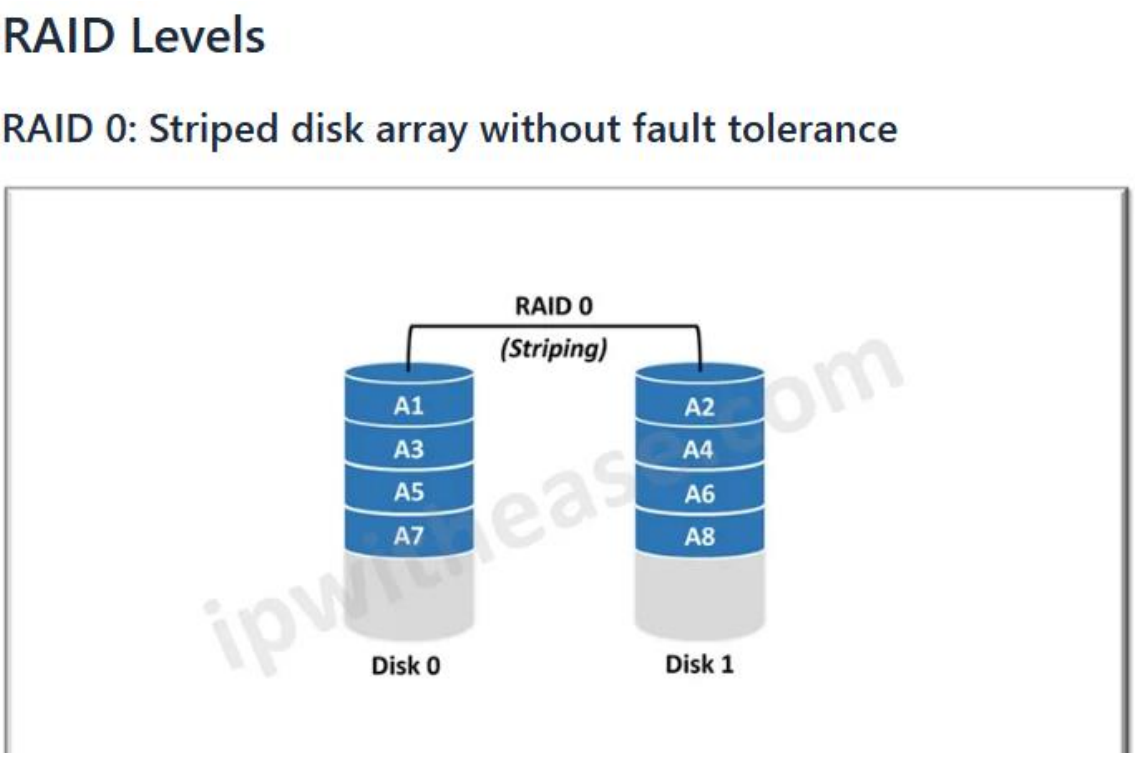
Advantages of RAID

- **When using RAID, you can have multiple hard drives working together to make one large volume. This allows you to increase your storage space.**
- **RAID can also help protect your data in the event one hard drive fails.**
- **It can also help minimize data loss due to hardware errors or power outages.**
- **RAID arrays are commonly used in tandem with replicated storage systems. Some replicated storage systems make use of more than one RAID array to ensure business continuity.**
- **Each of RAID levels has its own characteristics of:**
- **Fault-tolerance which is the ability to survive of one or several disk failures.**
- **Performance which shows the change in the read and write speed of the entire array as compared to a single disk.**
- **The capacity of the array which is determined by the amount of user data that can be written to the array. The array capacity depends on the RAID level and does not always match the sum of the sizes of the RAID member disks.**

Terminology of RAID Data Storage

- **Striping: splitting the flow of data into blocks of a certain size (called "block size") then writing of these blocks across the RAID one by one. This way of data storage affects on the performance.**
- **Mirroring is a storage technique in which the identical copies of data are stored on the RAID members simultaneously. This type of data placement affects the fault tolerance as well as the performance.**
- **Parity is a storage technique which is utilized striping and checksum methods. In parity technique, a certain parity function is calculated for the data blocks. If a drive fails, the missing block are recalculated from the checksum, providing the RAID fault tolerance.**

- **Block is the physical location on the disk where data is written. The amount of space on a RAID volume is determined by the RAID controller.**
- **Left/right symmetry: The way in which data and parity are distributed across a RAID array is determined by symmetry. There are four common types of symmetry (depending on the RAID vendor) and some companies develop their own proprietary styles to suit their unique requirements.**
- **Degraded mode: A degraded RAID occurs when a drive becomes unreadable and is removed from the array. The new data and parity are then written to the remaining drives. If a data request comes from the failed drive, it is solved with the parity of the others. A degraded RAID is a RAID with fewer drives, which consequently decreases the performance of the RAID.**
- **Hot spare: Its a spare disk that can replace a failed disk within a RAID array.**



Provides data striping (spreading out blocks of each file across multiple disk drives) but *no redundancy.* This improves performance but does not deliver fault tolerance. If one drive fails then all data in the array is lost. RAID 0: Striped Disk Arrays

RAID 0 is a group of hard drives configured to store data in parallel. It distributes data across multiple drives in a striped array (hence the name RAID 0), which allows for higher data transfer rates and increased storage capacity.

RAID 0 is used mostly for performance-intensive tasks like video editing and stock trading, where data must be accessed quickly. The benefit of RAID 0 is increased

performance due to the parallel access of data. This configuration is capable of reading and writing data at the same rate as a single drive.

Therefore, RAID 0 is great for situations where you need to quickly transfer large amounts of data, such as when editing video.

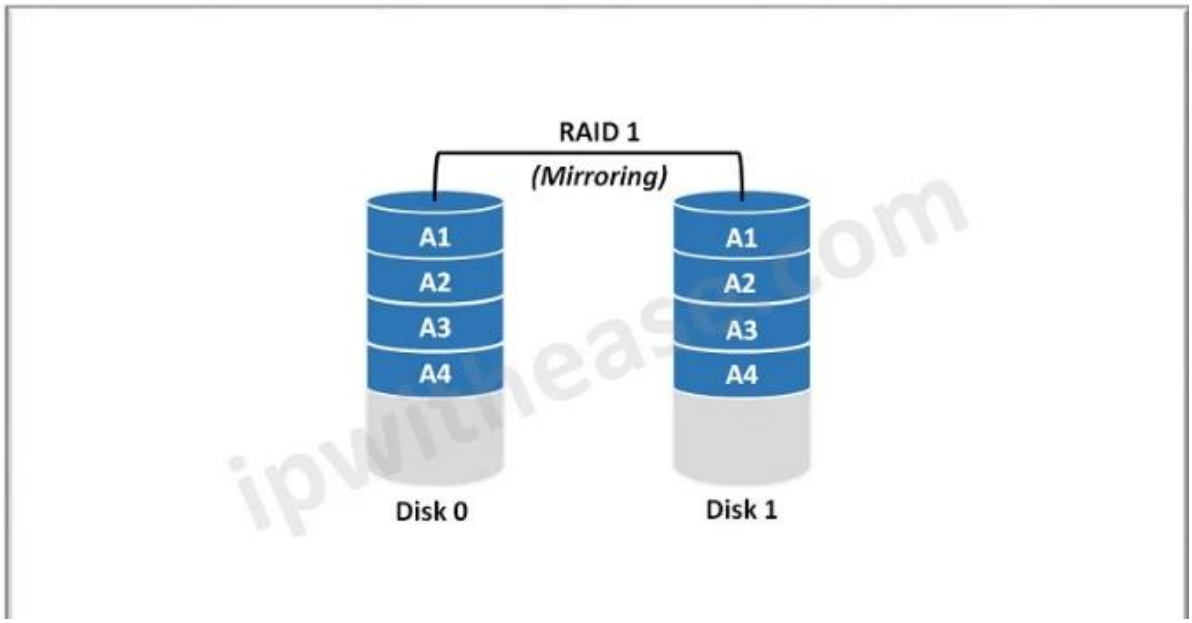
Benefits of RAID 0

- **Increased performance –** RAID 0 offers a significant performance boost thanks to parallel access of data.
- **Increased storage capacity –** By distributing data across multiple drives, RAID 0 can use the 100% storage capacity of a system.
- **Consistency –** Since the data is distributed across the drives in the array, RAID 0 provides consistent read and write speeds.
- **Simplicity –** Since RAID 0 is designed for end users, it's an easy configuration to set up.
- **Low maintenance –** RAID 0 has low maintenance requirements since it doesn't require any special monitoring or management.

Drawbacks of RAID 0

- **Loss of data –** RAID 0 isn't designed for data protection, if any one drive in the array fails, the data on the entire array is lost.
- **Increased risk of data corruption –** RAID 0 increases the risk of data corruption due to the increased likelihood of the drives reading and writing data at the same time.
- **Lower durability –** RAID 0 has a shorter lifespan than other configurations due to the increased number of read/write cycles. This can reduce the lifespan of the drives within the array.
- **Loss of speed when replacing a failed drive –** When a drive fails within a RAID 0 configuration, you must replace the failed drive before you can use the RAID 0 again. This can cause a noticeable loss in speed until the new drive is fully integrated into the array.

RAID 1: Mirroring and duplexing



Provides disk mirroring. Level 1 provides twice the read transaction rate of single disks and the same write transaction rate as single disks.

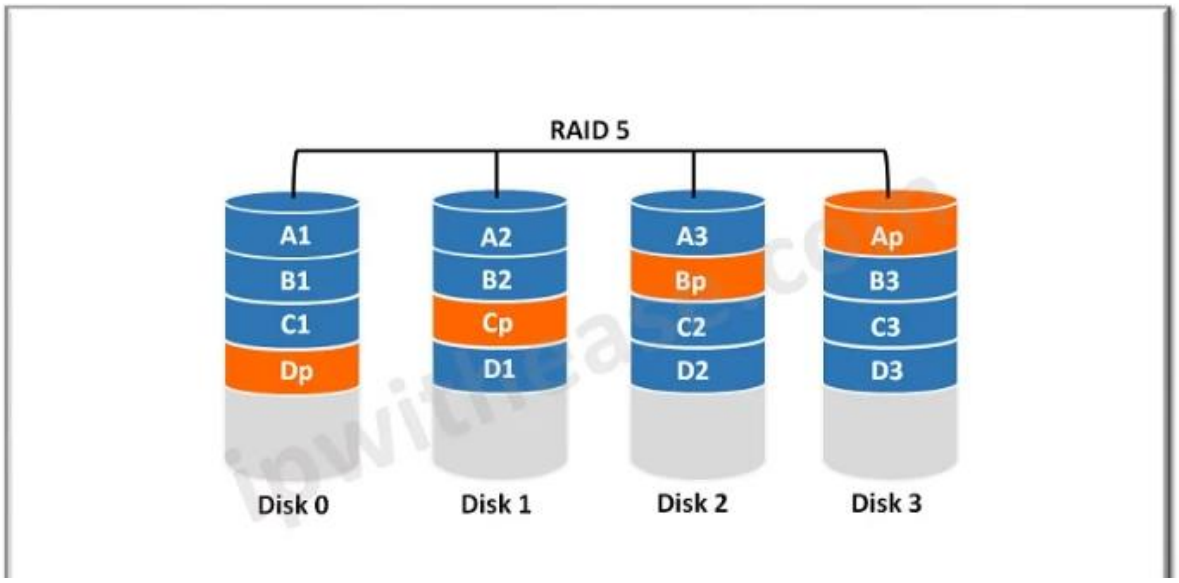
Benefits of RAID 1

- **Data redundancy –** In the event of a drive failure, the data on the other drive in the array will remain unaffected.
- **Fault tolerance:** This type of data storage is most-suited for the most mission-critical applications. When one drive fails, another drive takes over the primary duty. Since both drives include identical data, the users are unaffected.
- **High availability**: In the event of an emergency or data loss, data can be retrieved from two or more disks available due to disk mirroring. Therefore, chances to lose data are less.
- **High security:** RAID 1 provides data security by copying data at multiple places. In the event that one of your systems is hacked and you lose data, you can still access another.

Drawbacks of RAID 1

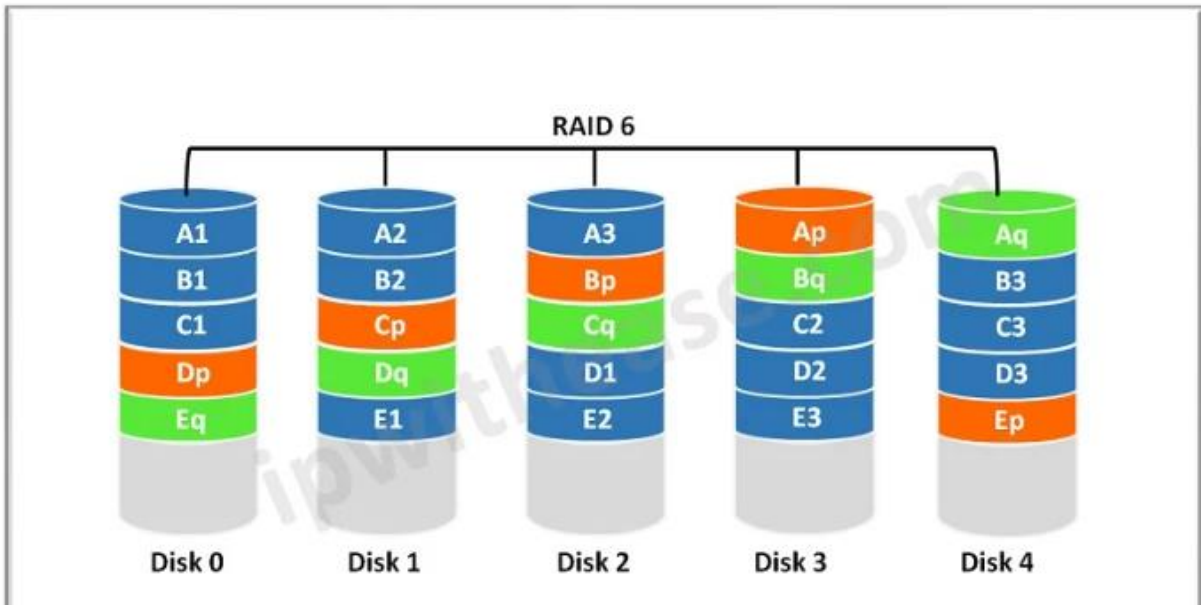
- **Lower performance –** RAID 1 is designed to provide consistent data across both drives in the array, which means it's slower than a single drive.
- **Increased cost –** Due to the need for two drives, RAID 1 is one of the more expensive configurations available.
- **Reduced Storage capacity –** As mirrored disks with same data are present, the actual storage capacity is 50%
- **Lag while swapping –** In case of failure of first disk, the second disk doesn't start automatically and need to be restarted.

RAID 5: Block interleaved distributed parity



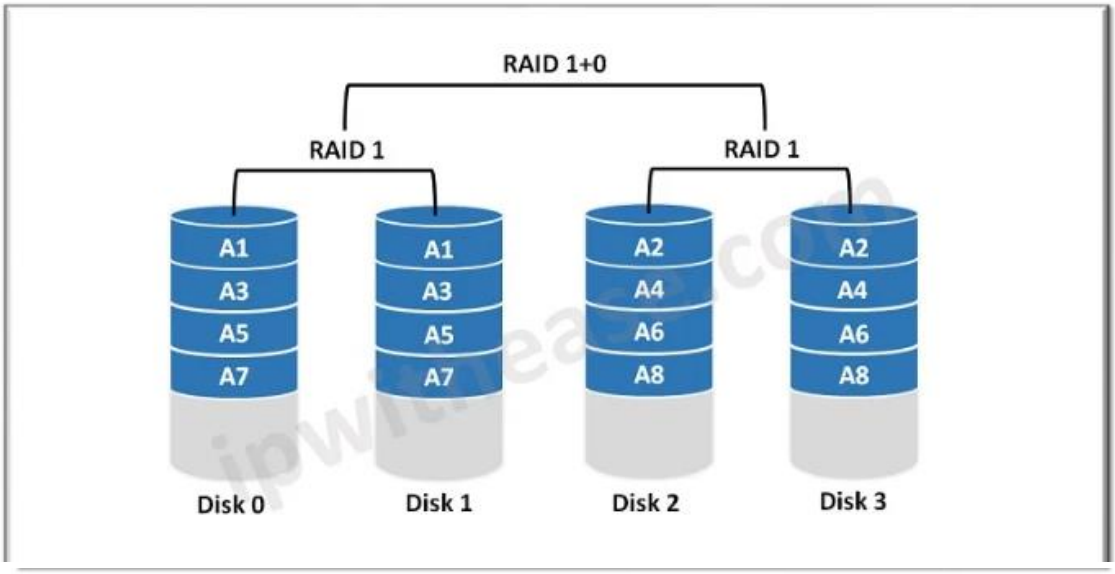
Provides data striping at the byte level and also stripe error correction information. This results in excellent performance and good fault tolerance. Level 5 is one of the most popular implementations of RAID.

RAID 6: Independent data disks with double parity



Provides block-level striping with parity data distributed across all disks.

RAID 10: A stripe of mirrors



Not one of the original RAID levels, multiple RAID 1 mirrors are created, and a RAID 0 stripe is created over these.

PARAMETER	RAID 0	RAID1 1	RAID 5	RAID 6	RAID 10
Minimum Drives	2	2	3	4	4
Data Protection	No Protection	Single-drive failure	Single-drive failure	Two-drive failure	Up to 1 disk failure in each sub-array
Capacity Utilization	100%	50%	67% – 94%	50% – 88%	50%
Read Performance	High	High	High	High	High
Write Performance	High	Medium	Low	Low	Medium
Read Performance (Degraded)	NA	Medium	Low	Low	High
Write Performance (Degraded)	NA	High	Low	Low	High
Use Cases	High end workstations, real-time rendering, data logging, very transitory data, etc.	Transaction databases, operating System, etc.	Web Serving, Data Warehousing, Archiving, etc.	Data archive backup to disk, server with larger capacity requirements, HA solutions, etc.	Application servers, fast databases, etc.

Difference: RAID 0 vs RAID 1

Below table explains the main differences between the 2 major raid levels **RAID 0** and **RAID 1:**

PARAMETER	RAID 0	RAID 1
Primary Operation	Disk Striping	Disk Mirroring
Mirroring	No	Yes
Redundancy	No	Yes
Fault Tolerance	No	Yes
Relative Storage Capacity	100%	50%
Performance	RAID 0 offers faster read and write speeds compared with RAID 1.	RAID 1 offers slower write speeds but could offer the same read performance as RAID 0 if the RAID controller uses multiplexing to read data from disks.
Data Recovery	Not possible	Can be recovered in a disaster recovery program
Data Protection	No	Mirror protection
Target applications	Where data reliability is less of a concern and speed is important.	Where data is very critical and data loss is unacceptable
Minimum number of physical disks required	2	2

How to choose between RAID 0 and RAID 1?

When it comes to choosing between RAID 0 and RAID 1, there are several factors to consider.

- RAID 0 is best suited for environments where faster data throughput is critical, such as video editing or big data analytics.
- RAID 1 is best suited for environments that require high levels of data redundancy, like medical facilities, financial institutions, and government agencies.
- It's also important to consider the lifespan of the drives you're using in your system. If you're using high-end drives that have a long lifespan, then RAID 1 is a good option.

Conclusion

Ultimately, the decision between RAID 0 and RAID 1 comes down to how critical the data is that's stored on your hard drives. If the data is critical and cannot be interrupted by hardware failure, then RAID 1 is the better option. If, however, you need faster data throughput, then RAID 0 is the better option.