**RAID (Redundant Arrays of Independent Disks)**

Last Updated : 04 Aug, 2025

RAID is a technique that combines multiple hard drives or SSDs into a single system to improve performance, data safety, or both. If one drive fails, data can still be recovered from the others.

RAID helps store data more reliably and efficiently by spreading or copying data across drives. Different RAID levels offer different combinations of speed, storage capacity, and fault tolerance.

**How RAID Works?**

Let’s understand how RAID (Redundant Array of Independent Disks) works through a relatable example:

Imagine you have a favorite book that you want to keep safe. Instead of giving the entire book to just one friend for safekeeping, you take a smart approach:

1. **Splitting the Book:** You divide the book into smaller pieces (like chapters or sections) and give each piece to a different friend.
2. **Making Copies:** For extra security, you might also create duplicate pieces and give them to multiple friends.

Now, if one friend misplaces their piece, you can still recreate the entire book using the pieces held by the others. This way, your book is safe even if someone loses their portion.

This is exactly how RAID works with hard drives! RAID splits your data across multiple drives (similar to dividing the book into pieces). Depending on the RAID configuration, it may also create duplicates (like making extra copies). If one drive fails, the remaining drives can help reconstruct the lost data.

**What is a RAID Controller?**

A RAID controller manages multiple hard drives, making them work together as one system. It helps improve speed and adds data protection by handling drive failures. Think of it as a smart manager that boosts performance and keeps your data safe.

**Types of RAID Controller**

There are three types of RAID controller:

**1. Hardware-Based RAID**

Uses a dedicated physical controller to manage hard drives.

* Offers high speed and reliability
* Can work independently from the computer's processor
* Often built into the motherboard or as a separate card

Think of it as a captain managing the drives smoothly.

**2. Software-Based RAID**

Uses the computer’s processor and memory to manage RAID.

* No special hardware needed
* Cost-effective, but may reduce overall system performance
* Slower than hardware RAID

Acts like a helpful assistant, but shares the load with other tasks.

**3. Firmware-Based RAID (Fake RAID)**

Built into the computer's BIOS/firmware, and works during boot-up.

* Needs a driver after the OS loads
* Cheaper than hardware RAID, but still uses CPU resources
* Also known as hybrid RAID or fake RAID

A startup helper that hands over the job to software once the system runs.

**Why Data Redundancy?**

[Data redundancy](https://www.geeksforgeeks.org/dbms/difference-between-data-redundancy-and-data-inconsistency/), although taking up extra space, adds to disk reliability. This means, that 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 multiple disks without the RAID technique, the loss of a single disk can affect the entire data.

**Key Evaluation Points for a RAID System**

When evaluating a RAID system, the following critical aspects should be considered:

**1. Reliability:** Refers to the system's ability to tolerate disk faults and prevent data loss.

**Example:**

* RAID 0 offers no fault tolerance; if one disk fails all data is lost.
* RAID 5 can tolerate one disk failure due to parity data.
* RAID 6 can handle two simultaneous disk failures.

**2. Availability:** The fraction of time the RAID system is operational and available for use.

**Example:**

* RAID 1 (Mirroring) allows immediate data access even during a single disk failure.
* RAID 5 and 6 may degrade performance during a rebuild, but data remains available.

**3. Performance:**Measures how efficiently the RAID system handles data processing tasks. This includes:

* **Response Time:** How quickly the system responds to data requests.
* **Throughput:** The rate at which the system processes data (e.g., MB/s or IOPS).

**Key Factors:**

* RAID levels affect performance differently:
* RAID 0 offers high throughput but no redundancy.
* RAID 1 improves read performance by serving data from either mirrored disk but may not improve write performance significantly.
* RAID 5/6 introduces overhead for parity calculations, affecting write speeds.
* Workload type (e.g., sequential vs. random read/write operations).

**Performance Trade-offs:** Higher redundancy often comes at the cost of slower writes (due to parity calculations).

**4. Capacity:** The amount of usable storage available to the user after accounting for redundancy mechanisms.

**Key Calculation:** For a set of N disks, each with B blocks, the available capacity depends on the RAID level:

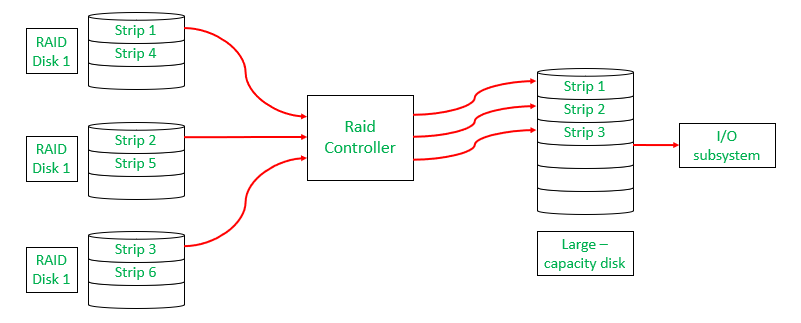
* **RAID 0:** All N×B blocks are usable (no redundancy).
* **RAID 1:** Usable capacity is B (only one disk's capacity due to mirroring).
* **RAID 5:** Usable capacity is (N−1)×B (one disk's worth of capacity used for parity).
* **RAID 6:** Usable capacity is (N−2)×B (two disks’ worth used for parity).

**Trade-offs:** Higher redundancy (RAID 5/6) reduces available capacity compared to non-redundant setups (RAID 0).

RAID is very transparent to the underlying system. This means, that to the host system, it appears as a single big disk presenting itself as a linear array of blocks. This allows older technologies to be replaced by RAID without making too many changes to the existing code.

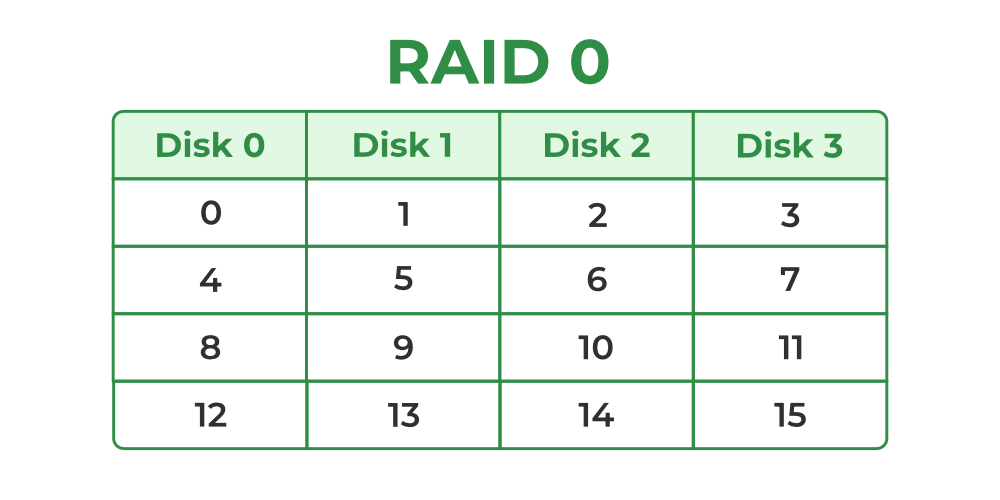
**Different RAID Levels**

* [RAID-0 (Stripping)](https://www.geeksforgeeks.org/operating-systems/disk-striping-raid-0/)
* [RAID-1 (Mirroring)](https://www.geeksforgeeks.org/operating-systems/disk-mirroring-raid-1/)
* RAID-2 (Bit-Level Stripping with Dedicated Parity)
* RAID-3 (Byte-Level Stripping with Dedicated Parity)
* RAID-4 (Block-Level Stripping with Dedicated Parity)
* RAID-5 (Block-Level Stripping with Distributed Parity)
* RAID-6 (Block-Level Stripping with two Parity Bits)

Raid Controller

**1. RAID-0 (Stripping)**

* RAID-0 improves system performance by splitting data into smaller "blocks" and spreading them across multiple disks. This process is called "striping." It enhances data access speed by enabling parallel read/write operations but provides no redundancy or fault tolerance.

RAID-0

* A set of blocks distributed across the disks forms a "stripe." For instance, "0, 1, 2, 3" is one stripe.
* Instead of placing just one block into a disk at a time, we can work with two (or more) blocks placed into a disk before moving on to the next one.

**Evaluation**

* **Reliability:** 0   
  There is no duplication of data. Hence, a block once lost cannot be recovered.
* **Capacity:** N\*B   
  The entire space is being used to store data. Since there is no duplication, N disks each having B blocks are fully utilized.

**Advantages**

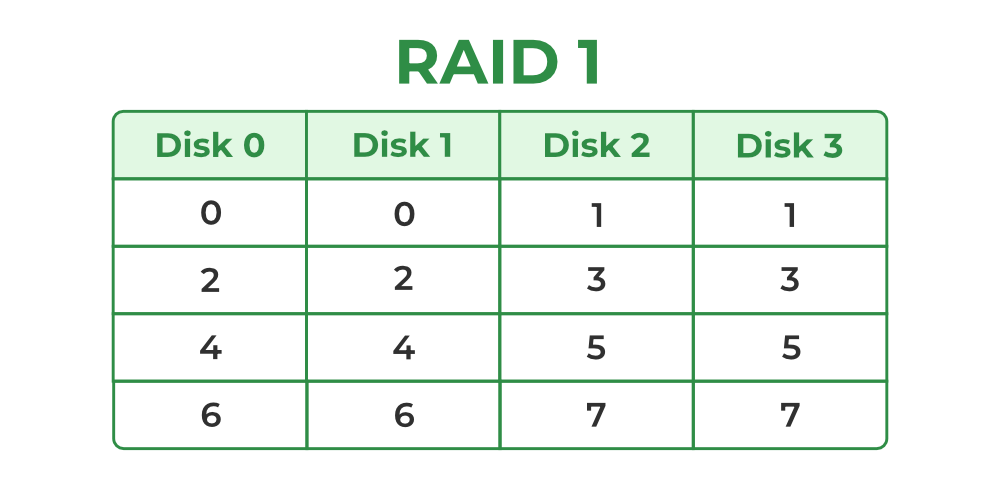
* It is easy to implement.
* It utilizes the storage capacity in a better way.

**Disadvantages**

* A single drive loss can result in the complete failure of the system.
* It's not a good choice for a critical system.

**2. RAID-1 (Mirroring)**

* RAID-1 enhances reliability by creating an identical copy (mirror) of each data block on separate disks. This ensures that even if one disk fails, the data remains accessible from its duplicate. While this configuration is highly reliable, it requires significant storage overhead.



Raid-1

**Mirroring:**  
Each block of data is written to two (or more) disks.  
  
**For example:**

* Block 0 is stored on Disk 0 and its duplicate on Disk 1.
* Block 1 is stored on Disk 2 and its duplicate on Disk 3.

**Mirroring Level 2:**  
In the figure, "mirroring level 2" indicates that there are two copies of each block, distributed across different disks.  
RAID 0 was unable to tolerate any disk failure. But RAID 1 is capable of reliability.

Read about [Difference Between RAID 0 and RAID 1.](https://www.geeksforgeeks.org/dbms/difference-between-raid-0-and-raid-1/)

**Evaluation**

Assume a RAID system with mirroring level 2.

* **Reliability:**1 to N/2   
  1 disk failure can be handled for certain because blocks of that disk would have duplicates on some other disk. If we are lucky enough and disks 0 and 2 fail, then again this can be handled as the blocks of these disks have duplicates on disks 1 and 3. So, in the best case, N/2 disk failures can be handled.
* **Capacity:**N\*B/2   
  Only half the space is being used to store data. The other half is just a mirror of the already stored data.

**Advantages**

* It covers complete redundancy.
* It can increase data security and speed.

**Disadvantages**

* It is highly expensive.
* Storage capacity is less.

**3. RAID-2 (Bit-Level Stripping with Dedicated Parity)**

* RAID-2 is a specialized RAID level that uses bit-level striping combined with error correction using [Hamming Code](https://www.geeksforgeeks.org/computer-networks/hamming-code-in-computer-network/). In this configuration, data is distributed at the bit level across multiple drives, and a dedicated parity drive is used for error detection and correction. While it offers strong fault tolerance, its complexity and cost make it rarely used in practice.

**Advantages**

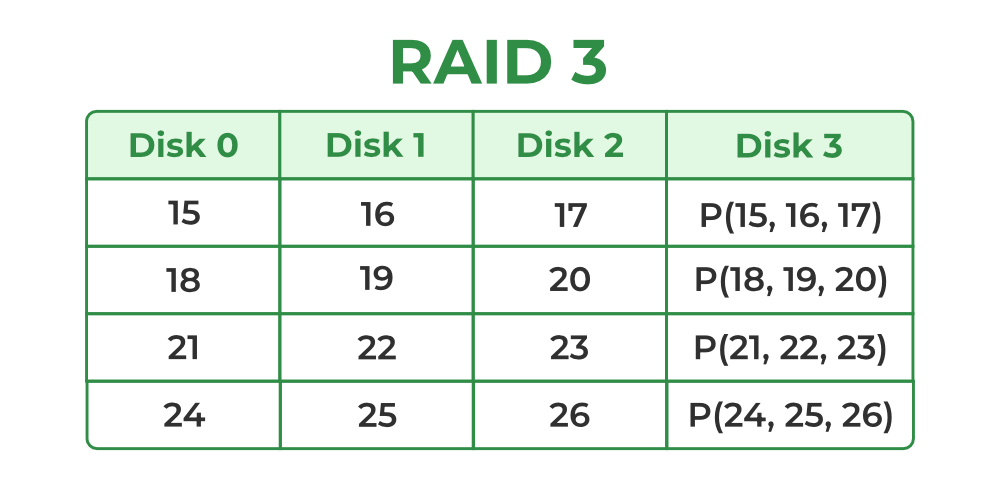
* In case of Error Correction, it uses hamming code.
* It Uses one designated drive to store parity.

**Disadvantages**

* It has a complex structure and high cost due to extra drive.
* It requires an extra drive for error detection.

**4. RAID-3 (Byte-Level Stripping with Dedicated Parity)**

* RAID-3 enhances fault tolerance by employing byte-level striping across multiple drives and storing parity information on a dedicated parity drive. The dedicated parity drive allows for the reconstruction of lost data if a single drive fails. This configuration is suitable for workloads requiring high throughput for sequential data but is less efficient for random I/O operations.



Raid-3

* Here Disk 3 contains the Parity bits for Disk 0, Disk 1, and Disk 2. If data loss occurs, we can construct it with Disk 3.

**Evaluation**

**Reliability**: RAID-3 can tolerate the failure of one disk. The lost data can be reconstructed using the parity drive and the remaining data drives.  
**Capacity:**Usable Capacity=(N−1)×B where N is the total number of drives, and B is the number of blocks per drive. The capacity of one drive is reserved for storing parity information.

**Advantages**

* Data can be transferred in bulk.
* Data can be accessed in parallel.

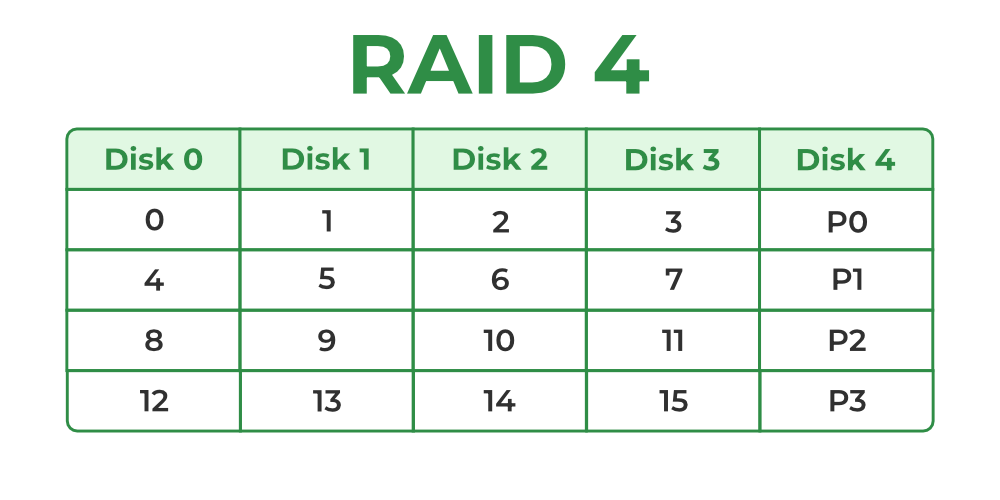
**Disadvantages**

* It requires an additional drive for parity.
* In the case of small-size files, it performs slowly.

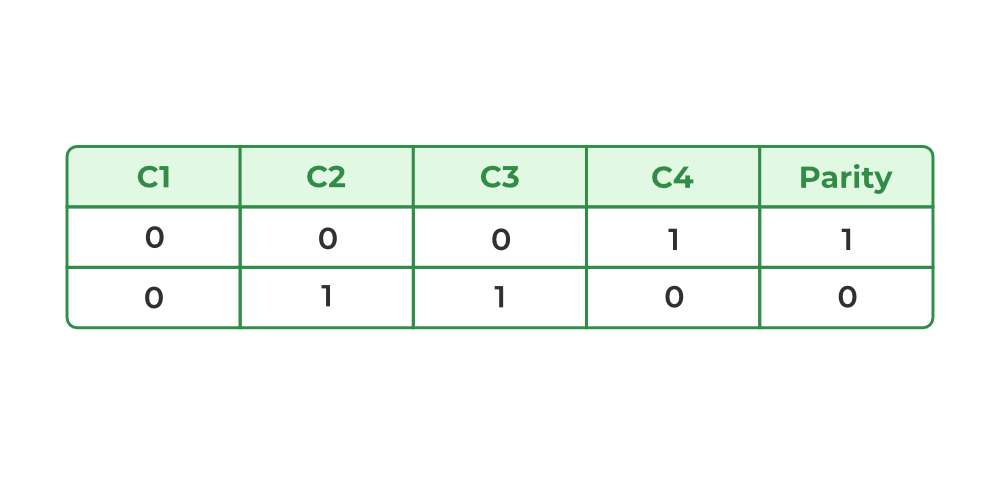
Read about [Difference Between RAID 2 and RAID 3](https://www.geeksforgeeks.org/dbms/difference-between-raid-2-and-raid-3/).

**5. RAID-4 (Block-Level Stripping with Dedicated Parity)**

* RAID-4 introduces block-level striping across multiple disks, combined with a dedicated parity disk to provide fault tolerance. Data is written in blocks, and a separate disk stores parity information calculated using the XOR function. This setup allows for data recovery in case of a single disk failure, making RAID-4 more reliable than RAID-0 but less efficient in write-intensive scenarios due to reliance on a dedicated parity disk.

Raid-4

* In the figure, we can observe one column (disk) dedicated to parity.
* Parity is calculated using a simple XOR function. If the data bits are 0,0,0,1 the parity bit is XOR(0,0,0,1) = 1. If the data bits are 0,1,1,0 the parity bit is XOR(0,1,1,0) = 0. A simple approach is that an even number of ones results in parity 0, and an odd number of ones results in parity 1.

Raid-4

* Assume that in the above figure, C3 is lost due to some disk failure. Then, we can recompute the data bit stored in C3 by looking at the values of all the other columns and the parity bit. This allows us to recover lost data.

Read about [Difference Between RAID 3 and RAID 4.](https://www.geeksforgeeks.org/dbms/difference-between-raid-3-and-raid-4/)

**Evaluation**

* **Reliability:** 1   
  RAID-4 allows recovery of at most 1 disk failure (because of the way parity works). If more than one disk fails, there is no way to recover the data.
* **Capacity:** (N-1)\*B   
  One disk in the system is reserved for storing the parity. Hence, (N-1) disks are made available for data storage, each disk having B blocks.

**Advantages**

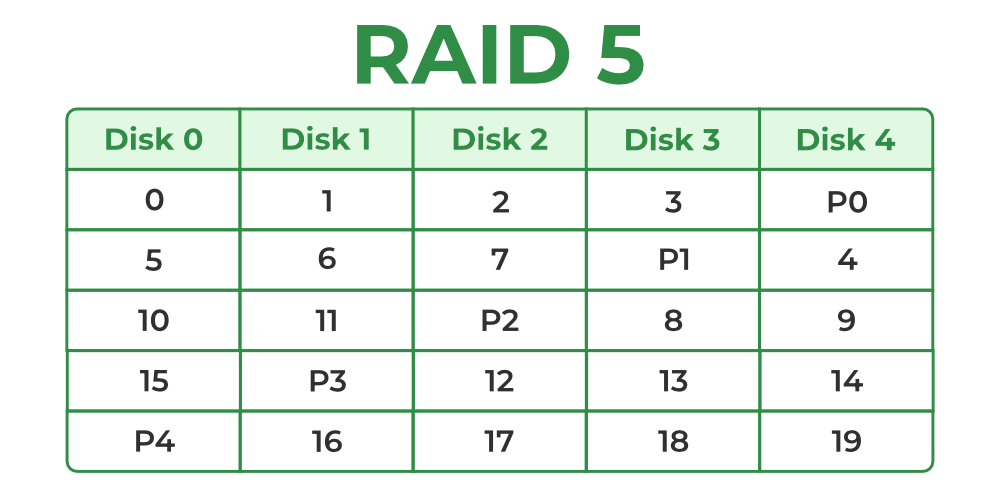
* It helps in reconstructing the data if at most one data is lost.

**Disadvantages**

* It can't help reconstructing data when more than one is lost.

**6. RAID-5 (Block-Level Stripping with Distributed Parity)**

* RAID-5 builds on RAID-4 by distributing parity information across all disks instead of storing it on a dedicated parity drive. This distributed parity significantly improves write performance, especially for random write operations, while maintaining fault tolerance for single disk failures. RAID-5 is one of the most commonly used RAID configurations due to its balance between reliability, performance, and storage efficiency.

Raid-5

**How RAID-5 Works?**

**Block-Level Striping:** Data is divided into blocks and striped across all drives in the array.

**Distributed Parity:** Parity bits, calculated using the XOR function, are distributed across all drives in a rotating pattern.

For example:

* Disk 0: Data (D1, D2), Parity (P1)
* Disk 1: Data (D3, D4), Parity (P2)
* Disk 2: Data (D5, D6), Parity (P3)

This rotation ensures no single disk is burdened with all parity operations, reducing bottlenecks.

**Data Recovery:** In case of a single disk failure, the missing data can be reconstructed by XOR-ing the remaining data blocks and parity information.

**Evaluation**

* **Reliability:** 1   
  RAID-5 allows recovery of at most 1 disk failure (because of the way parity works). If more than one disk fails, there is no way to recover the data. This is identical to RAID-4.
* **Capacity:** (N-1)\*B   
  Overall, space equivalent to one disk is utilized in storing the parity. Hence, (N-1) disks are made available for data storage, each disk having B blocks.

**Advantages**

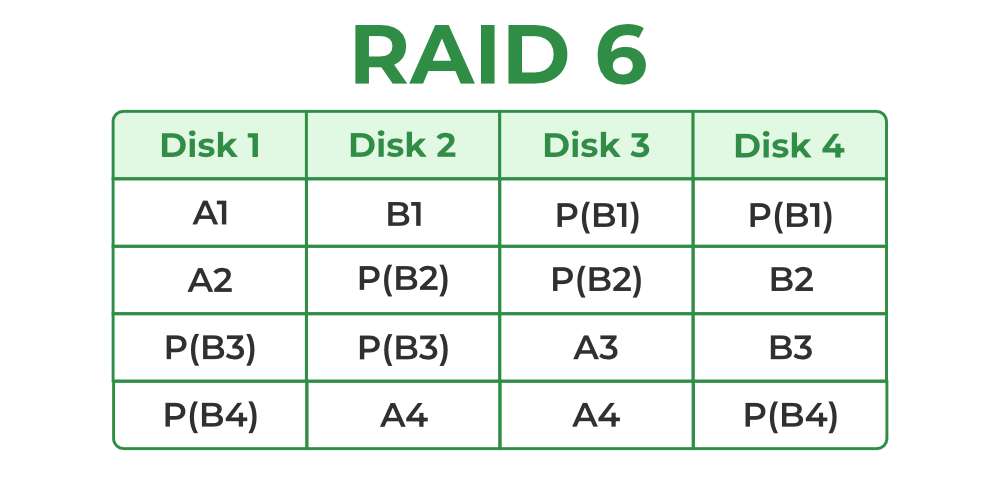
* Data can be reconstructed using [parity bits](https://www.geeksforgeeks.org/digital-logic/error-detection-codes-parity-bit-method/).
* It makes the performance better.

**Disadvantages**

* Its technology is complex and extra space is required.
* If both discs get damaged, data will be lost forever.

**7. RAID-6 (Block-Level Stripping with two Parity Bits)**

* RAID-6 is an advanced version of RAID-5 that provides enhanced fault tolerance by introducing double distributed parity. This allows RAID-6 to recover from the failure of up to two disks simultaneously, making it more reliable for critical systems with larger arrays. However, the added parity calculations can impact write performance.

Raid-6

**How RAID-6 Works ?**

**Block-Level Striping:** Data is divided into blocks and striped across all disks in the array.

**Double Distributed Parity:** Two sets of parity information are calculated for every block and distributed across all disks in the array in a rotating pattern.

Example:

* Disk 0: Data (D1, D2), Parity (P1)
* Disk 1: Data (D3, D4), Parity (P2)
* Disk 2: Data (D5, P1), Parity (P3)
* Disk 3: Parity (P2, P3), Data (D6)

**Data Recovery:** If one or two disks fail, the missing data can be reconstructed using the remaining data and parity information.

**Evaluation**

**1. Reliability:** RAID-6 can tolerate the simultaneous failure of two disks, providing greater fault tolerance than RAID-5.

**2. Capacity:**Usable Capacity=(N−2)×B where N is the total number of disks and B is the number of blocks per disk.

**Advantages**

* Very high data Accessibility.
* Fast read data transactions.

**Disadvantages**

* Due to double parity, it has slow write [data transactions](https://www.geeksforgeeks.org/dbms/acid-properties-in-dbms/).
* Extra space is required.

**Advantages of RAID**

* **Data redundancy:** By keeping numerous copies of the data on many disks, RAID can shield data from disk failures.
* **Performance enhancement:** RAID can enhance performance by distributing data over several drives, enabling the simultaneous execution of several [read/write operations](https://www.geeksforgeeks.org/computer-organization-architecture/read-and-write-operations-in-memory/).
* **Scalability:**RAID is scalable, therefore by adding more disks to the array, the storage capacity may be expanded.
* **Versatility:**RAID is applicable to a wide range of devices, such as workstations, servers, and personal PCs

**Disadvantages of RAID**

* **Cost:** RAID implementation can be costly, particularly for [arrays](https://www.geeksforgeeks.org/dsa/what-is-array/) with large capacities.
* **Complexity:** The setup and management of RAID might be challenging.
* **Decreased performance:** The[parity](https://www.geeksforgeeks.org/digital-logic/error-detection-codes-parity-bit-method/)calculations necessary for some RAID configurations, including RAID 5 and RAID 6, may result in a decrease in speed.
* **Single point of failure:** RAID is not a comprehensive backup solution while offering data redundancy. The array's whole contents could be lost if the RAID controller malfunctions.

Screenshots: A screen shot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screen shot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screen shot of a computer

AI-generated content may be incorrect.A screen shot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A computer chip on a black background

AI-generated content may be incorrect.A computer chip on a black background

AI-generated content may be incorrect.A computer screen shot of a computer

AI-generated content may be incorrect.A computer screen shot of a computer

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A table with text and images

AI-generated content may be incorrect.