Redundant Array of Inexpensive Disks (RAID)

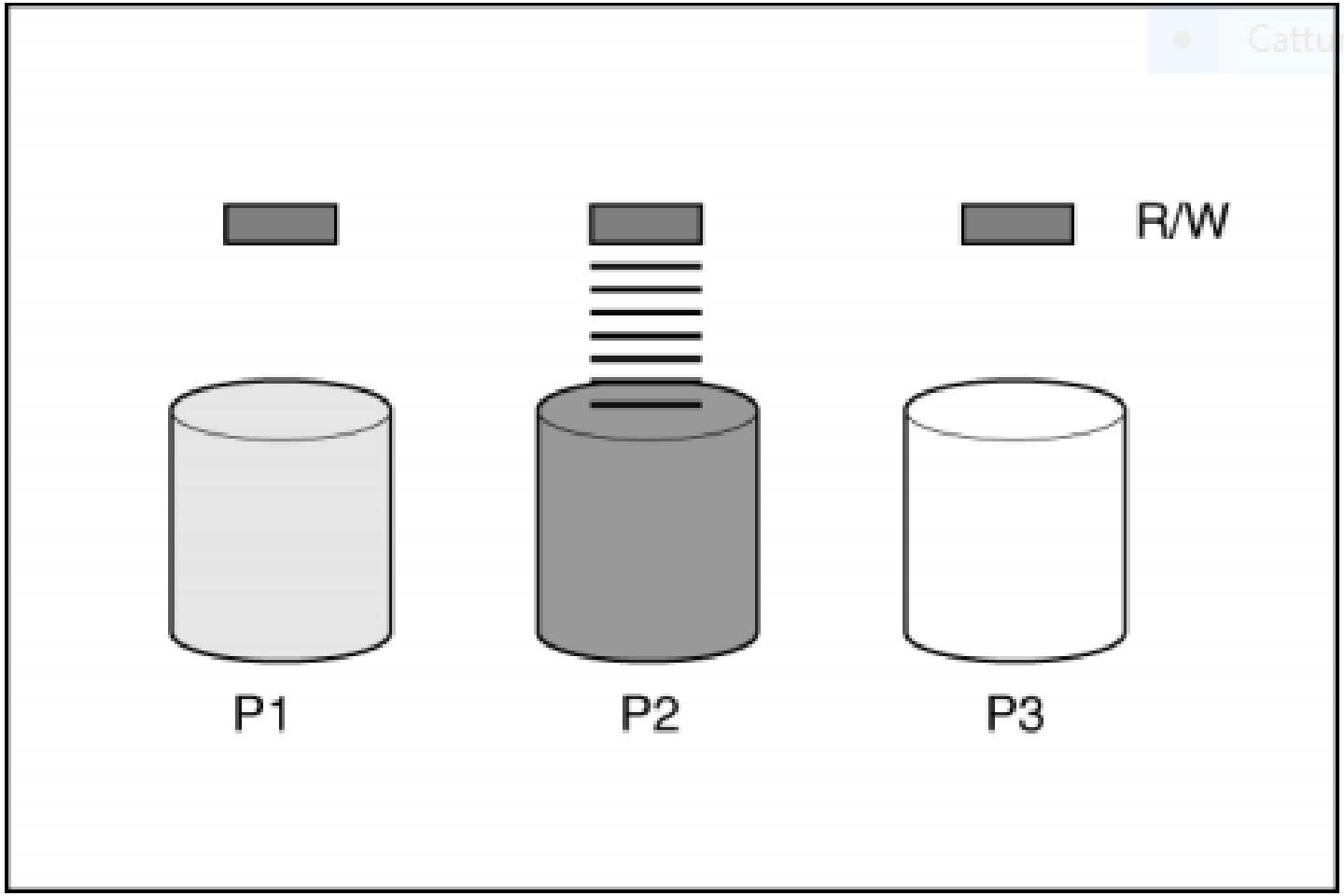
Drive arrays and fault-tolerance methods

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

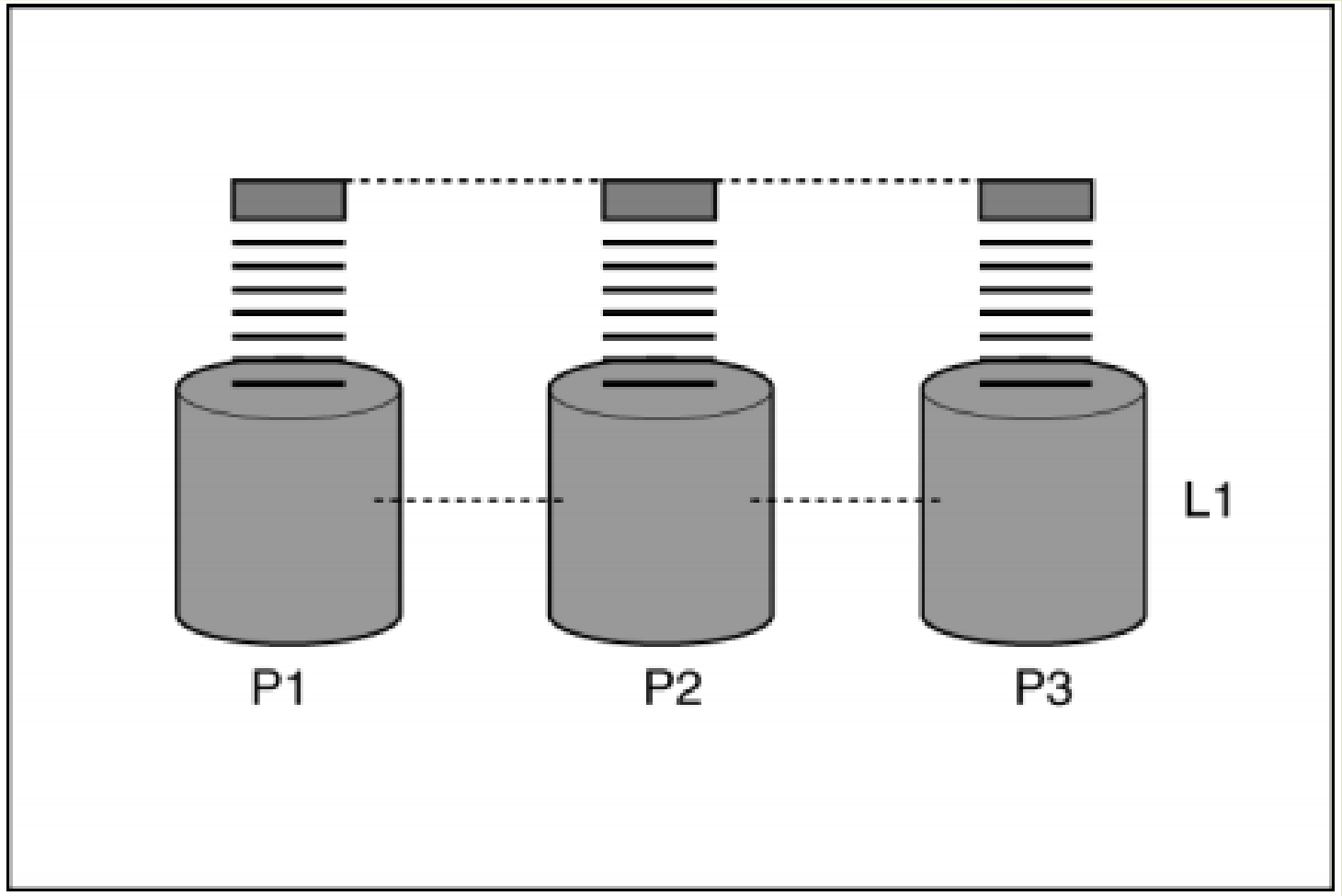
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

Drive arrays

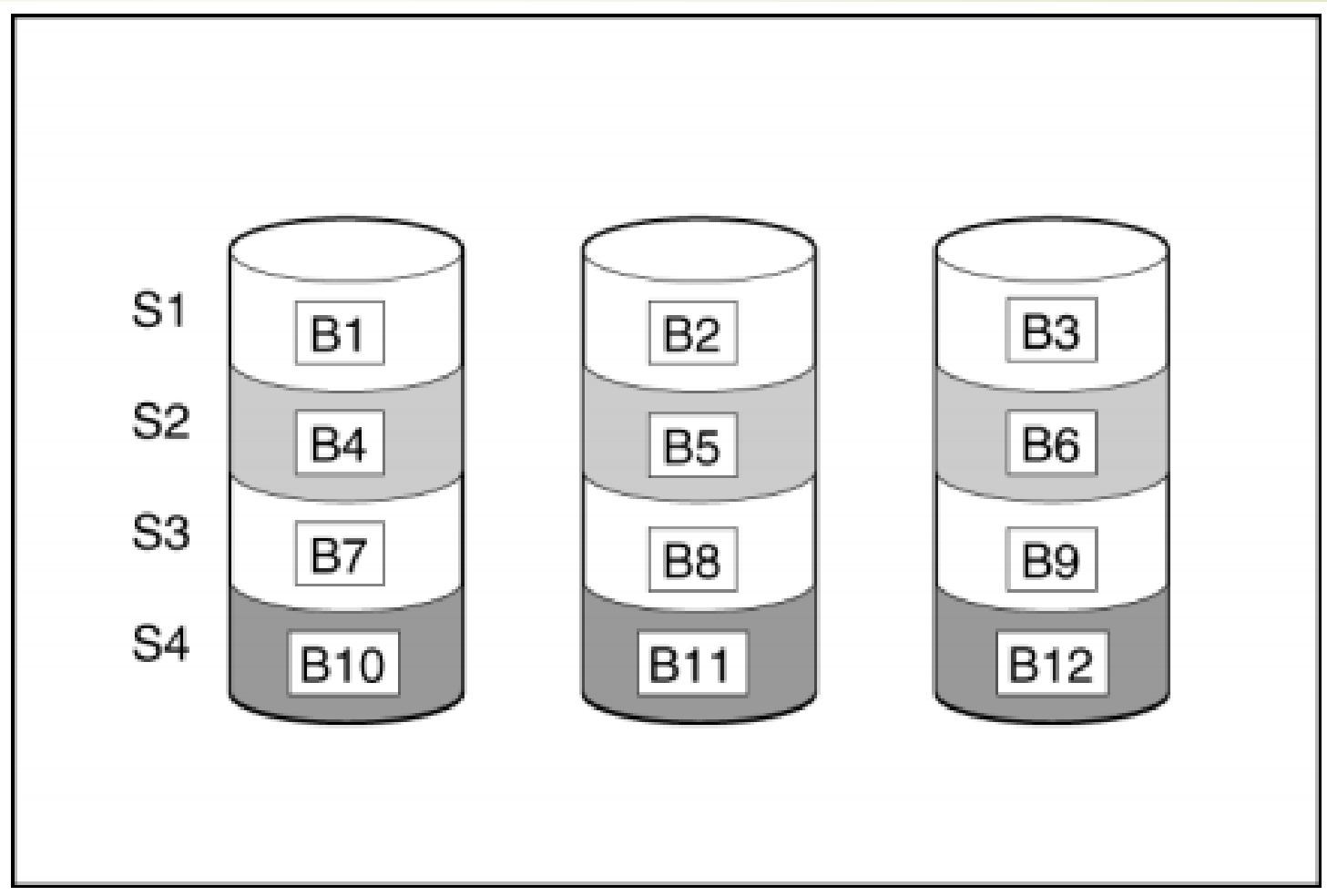
- The capacity and performance of a single physical (hard) drive is adequate for home users. However, business users demand higher storage capacities, higher data transfer rates, and greater protection against data loss when drives fail. Connecting extra physical drives (**Pn** in the figure) to a system increases the total storage capacity but has no effect on the efficiency of read/write (R/W) operations. Data can still be transferred to only one physical drive at a time.



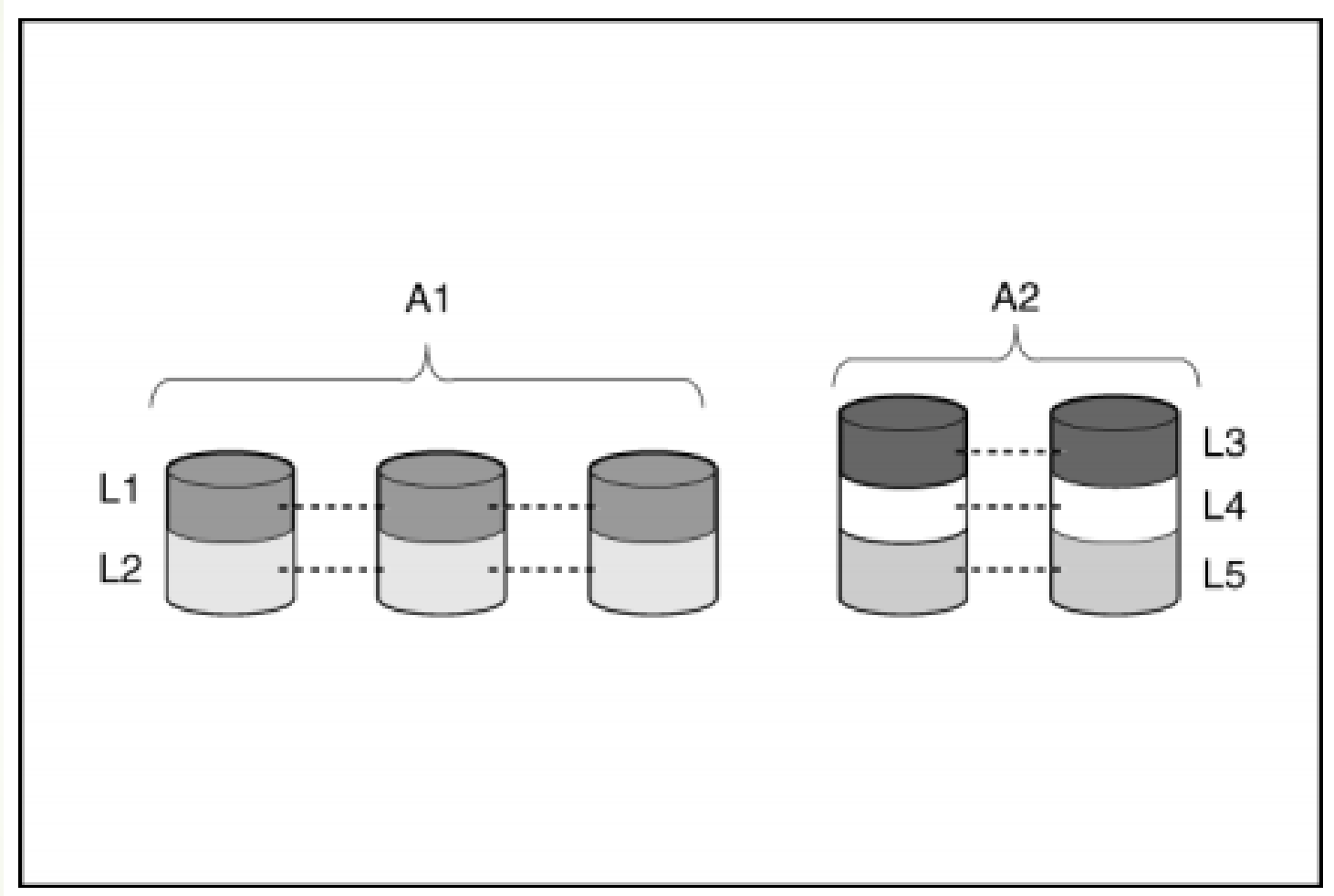
- With an array controller installed in the system, the capacity of several physical drives can be combined into one or more virtual units called **logical drives** (also called **logical volumes** and denoted by Ln in the figures in this section). Then, the read/write heads of all the constituent physical drives are active simultaneously, reducing the total time required for data transfer.



- Because the read/write heads are active simultaneously, the same amount of data is written to each drive during any given time interval. Each unit of data is called a **block** (denoted by **Bn** in the figure), and adjacent blocks form a set of data **stripes** (**Sn**) across all the physical drives that comprise the logical drive.



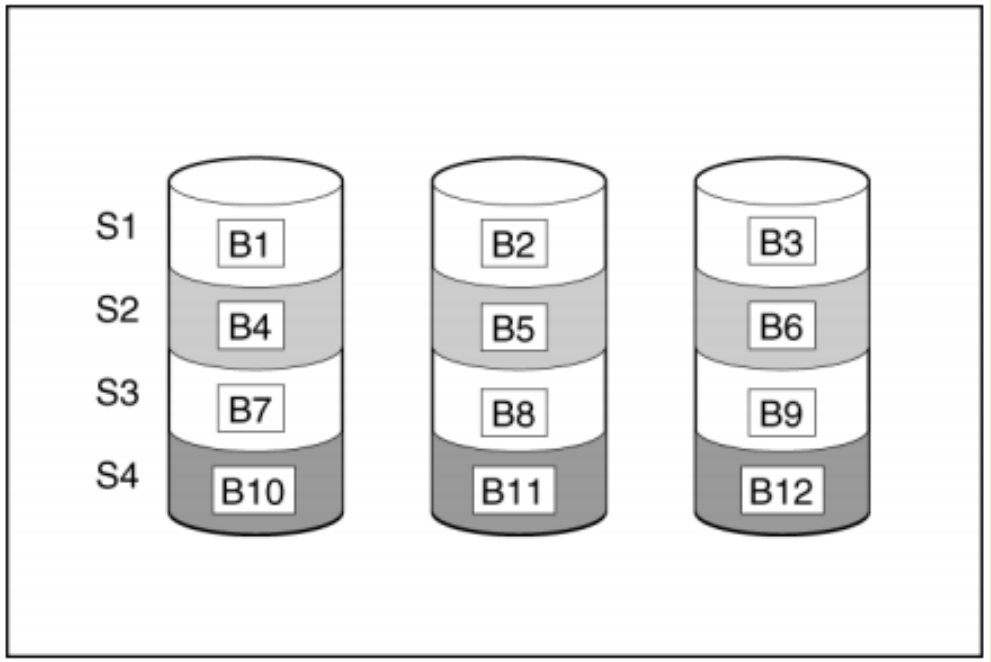
- For data in the logical drive to be readable, the data block sequence must be the same in every stripe. This sequencing process is performed by the array controller, which sends the data blocks to the drive write heads in the correct order. A natural consequence of the striping process is that each physical drive in a given logical drive will contain the same amount of data. If one physical drive has a larger capacity than other physical drives in the same logical drive, the extra capacity is wasted because it cannot be used by the logical drive.
- The group of physical drives containing the logical drive is called a **drive array**, or just **array** (denoted by **An** in the figure). Because all the physical drives in an array are commonly configured into just one logical drive, the term array is often used as a synonym for logical drive. However, an array can contain several logical drives, each of a different size.



- Each logical drive in an array is distributed across all of the physical drives within the array. A logical drive can also extend across more than one port on the same controller, but it cannot extend across more than one controller. Drive failure, although rare, is potentially catastrophic. For arrays that are configured as shown in the previous figure, failure of any physical drive in the array causes every logical drive in the array to suffer irretrievable data loss. To protect against data loss due to physical drive failure, logical drives are configured with **fault tolerance**.
- For any configuration except RAID 0, further protection against data loss can be achieved by assigning a drive as an **online spare** (or **hot spare**). This drive contains no data and is connected to the same controller as the array. When any other physical drive in the array fails, the controller automatically rebuilds information that was originally on the failed drive to the online spare. The system is thus restored to full RAID-level data protection, although it now no longer has an online spare. (However, in the unlikely event that another drive in the array fails while data is being rewritten to the spare, the logical drive will still fail.) When you configure an online spare, it is automatically assigned to all logical drives in the same array. Additionally, you do not need to assign a separate online spare to each array. Instead, you can configure one hard drive to be the online spare for several arrays if the arrays are all on the same controller.

Fault-tolerance methods

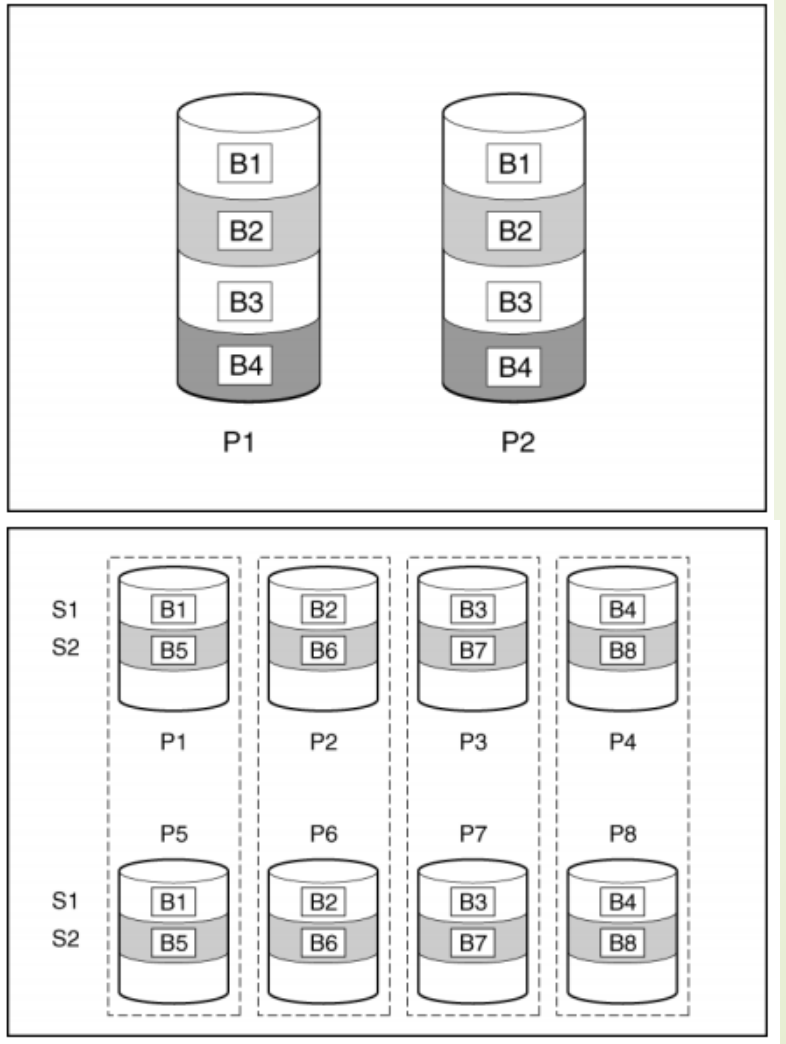
- **RAID 0 - No fault tolerance**
  - A RAID 0 configuration provides data striping, but there is no protection against data loss when a drive fails. However, it is useful for rapid storage of large amounts of noncritical data (for printing or image editing, for example) or when cost is the most important consideration.



- Advantages:
  - Has the highest write performance of all RAID methods.
  - Has the lowest cost per unit of stored data of all RAID methods.
  - All drive capacity is used to store data (none is needed for fault tolerance).
- Disadvantages:
  - All data on the logical drive is lost if a physical drive fails.
  - Cannot use an online spare.
  - Can only preserve data by backing it up to external drives.

RAID 10

- RAID 1+0 (RAID 10) – MIRRORING
- (same physical HDs)
  - In a RAID 1+0 (RAID 10) configuration, data is duplicated to a second drive.
  - When the array has more than two physical drives, drives are mirrored in pairs.



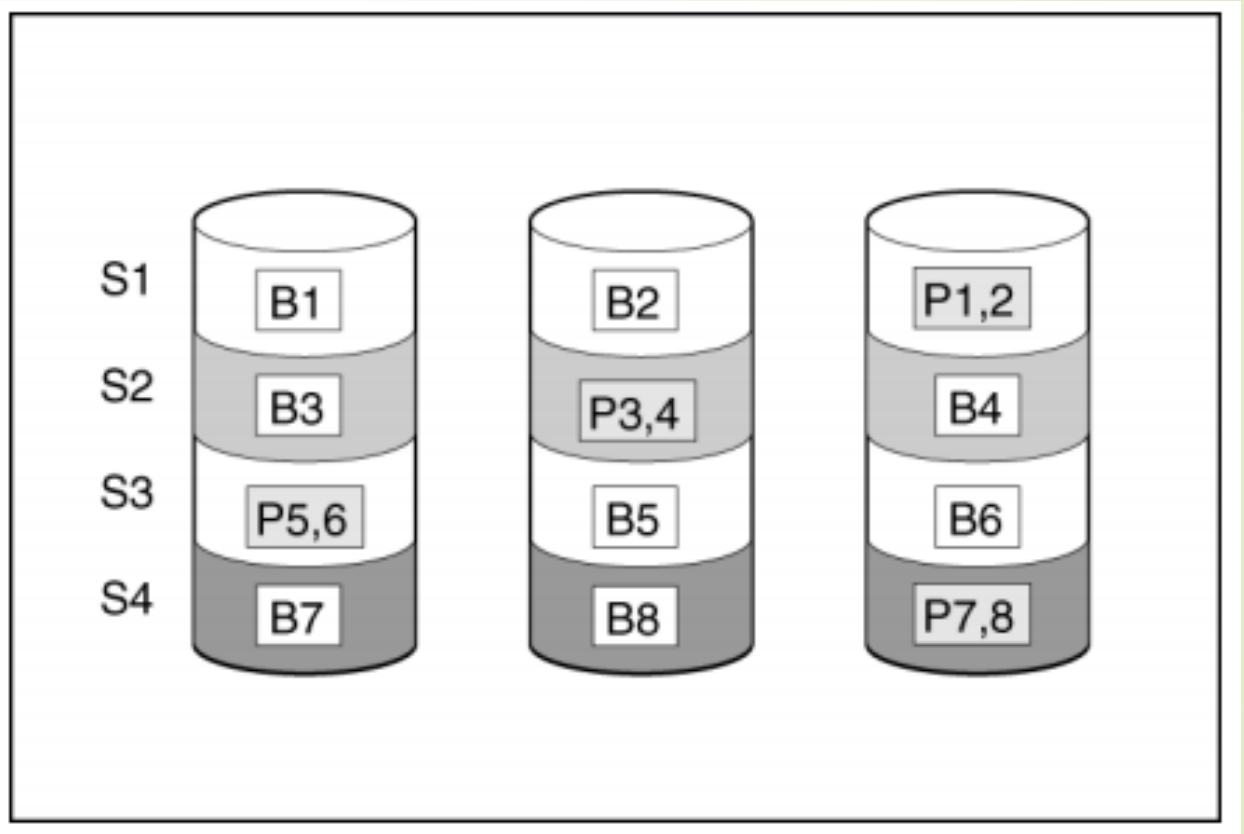
RAID 10

- In each mirrored pair, the physical drive that is not busy answering other requests answers any read requests that are sent to the array. This behavior is called load balancing. If a physical drive fails, the remaining drive in the mirrored pair can still provide all the necessary data. Several drives in the array can fail without incurring data loss, as long as no two failed drives belong to the same mirrored pair.
- This fault-tolerance method is useful when high performance and data protection are more important than the cost of physical drives.
- **NOTE:** When there are only two physical drives in the array, this fault-tolerance method is often referred to as RAID 1.

- Advantages:
  - This method has the highest read performance of any fault-tolerant configuration.
  - No data is lost when a drive fails, as long as no failed drive is mirrored to another failed drive.
  - Up to half of the physical drives in the array can fail.
- Disadvantages:
  - This method is expensive, because many drives are needed for fault tolerance.
  - Only half of the total drive capacity is usable for data storage.

RAID 5 - distributed data guarding

- In a RAID 5 configuration, data protection is provided by parity data (denoted by Px,y in the figure). This parity data is calculated stripe by stripe from the user data that is written to all other blocks within that stripe. The blocks of parity data are distributed evenly over every physical drive within the logical drive.



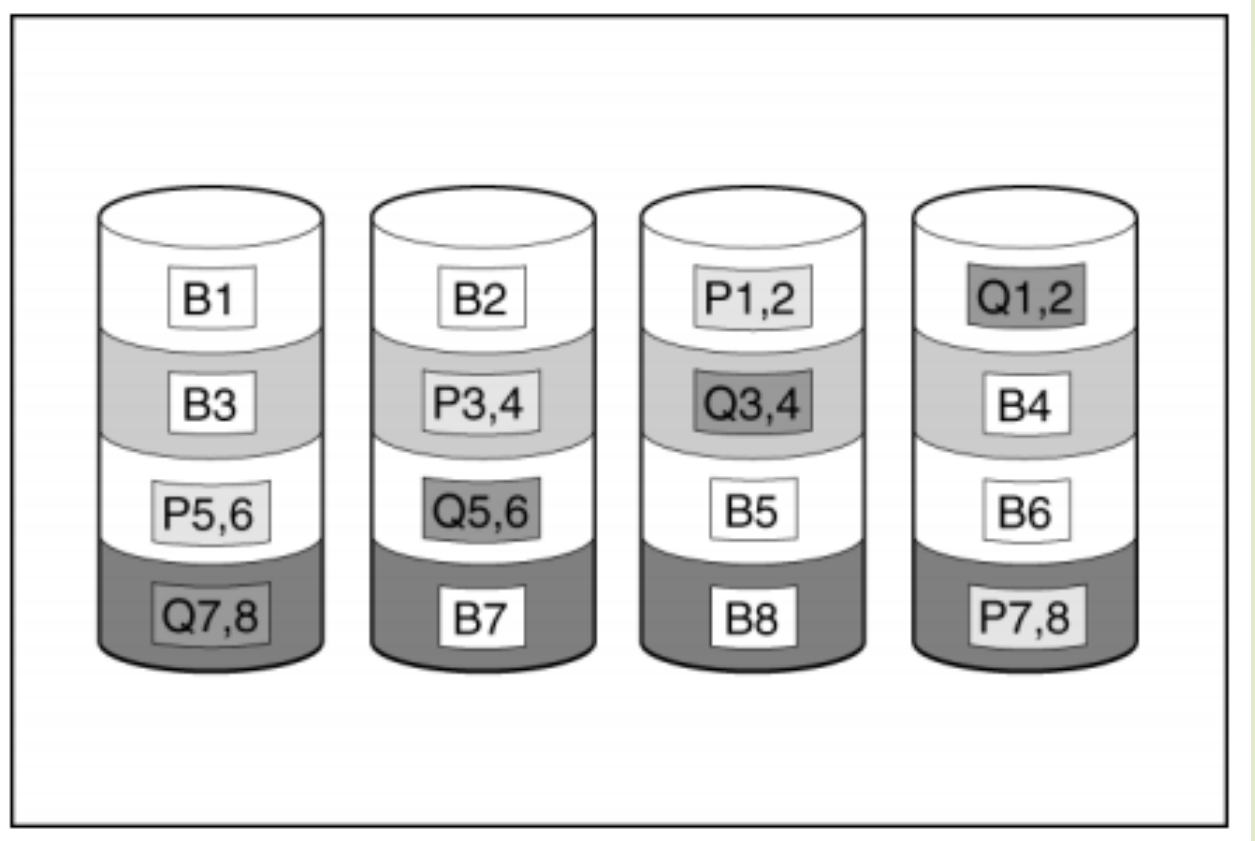
RAID 5 - distributed data guarding

- When a physical drive fails, data that was on the failed drive can be calculated from the remaining parity data and user data on the other drives in the array. This recovered data is usually written to an online spare in a process called a rebuild.
- **NOTE:** This configuration is useful when cost, performance, and data availability are equally important.

- Advantages:
  - Has high read performance.
  - Data is not lost if one physical drive fails.
  - More drive capacity is usable than with RAID 1+0—parity information requires only the storage space equivalent to one physical drive.
- Disadvantages:
  - Has relatively low write performance.
  - Data is lost if a second drive fails before data from the first failed drive is rebuilt

RAID 6 (ADG) - Advanced Data Guarding

- RAID 6 (ADG), like RAID 5, generates and stores parity information to protect against data loss caused by drive failure. With RAID 6 (ADG), however, two different sets of parity data are used (denoted by Px,y and Qx,y in the figure), allowing data to still be preserved if two drives fail. Each set of parity data uses a capacity equivalent to that of one of the constituent drives.



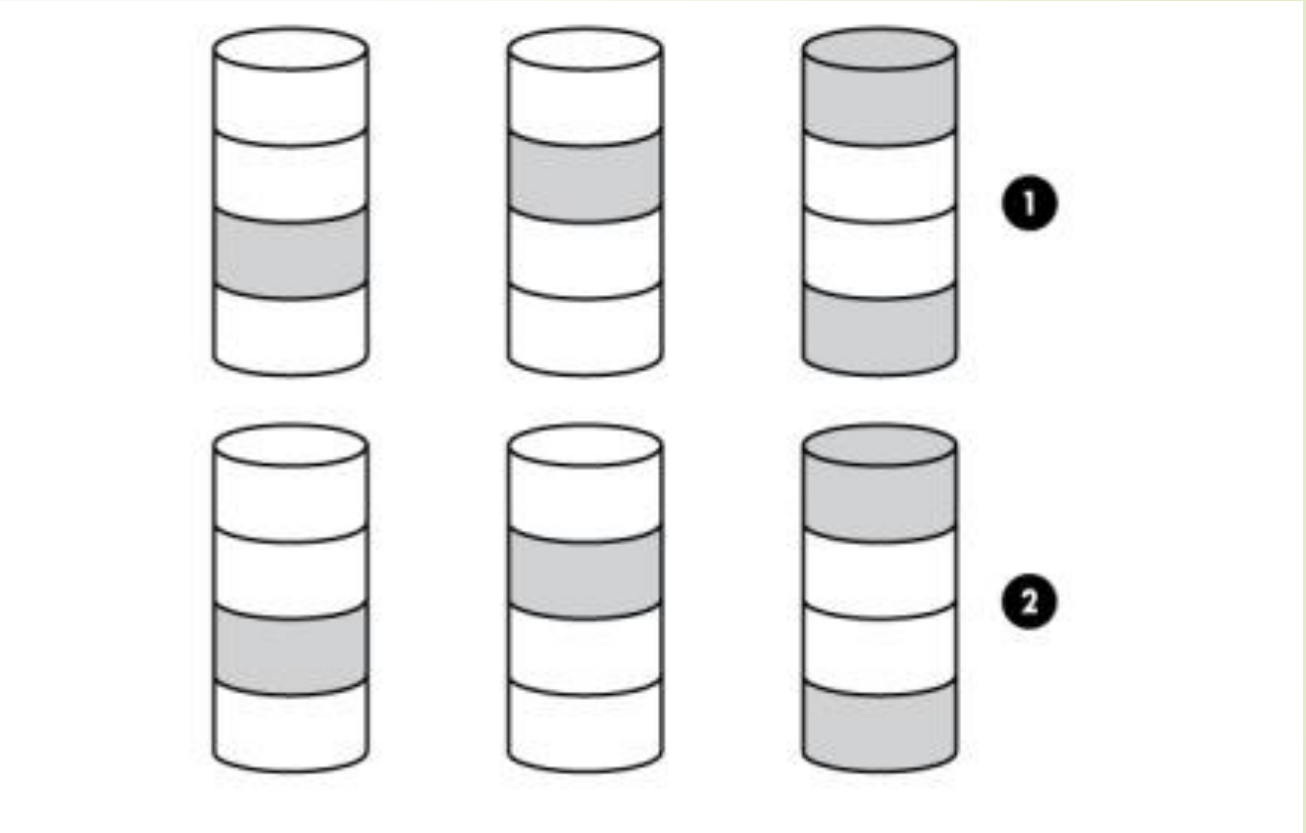
RAID 6 (ADG) - Advanced Data Guarding

- This method is most useful when data loss is unacceptable but cost is also an important factor. The probability that data loss will occur when an array is configured with RAID 6 (ADG) is less than it would be if it was configured with RAID 5.

- Advantages:
  - This method has a high read performance.
  - This method allows high data availability—Any two drives can fail without loss of critical data.
  - More drive capacity is usable than with RAID 1+0—Parity information requires only the storage space equivalent to two physical drives.
- Disadvantages:
  - The main disadvantage of RAID 6 (ADG) is a relatively low write performance (lower than RAID 5) because of the need for two sets of parity data.

RAID 50

- RAID 50 is a nested RAID method in which the constituent hard drives are organized into several identical RAID 5 logical drive sets (parity groups). The smallest possible RAID 50 configuration has six drives organized into two parity groups of three drives each.



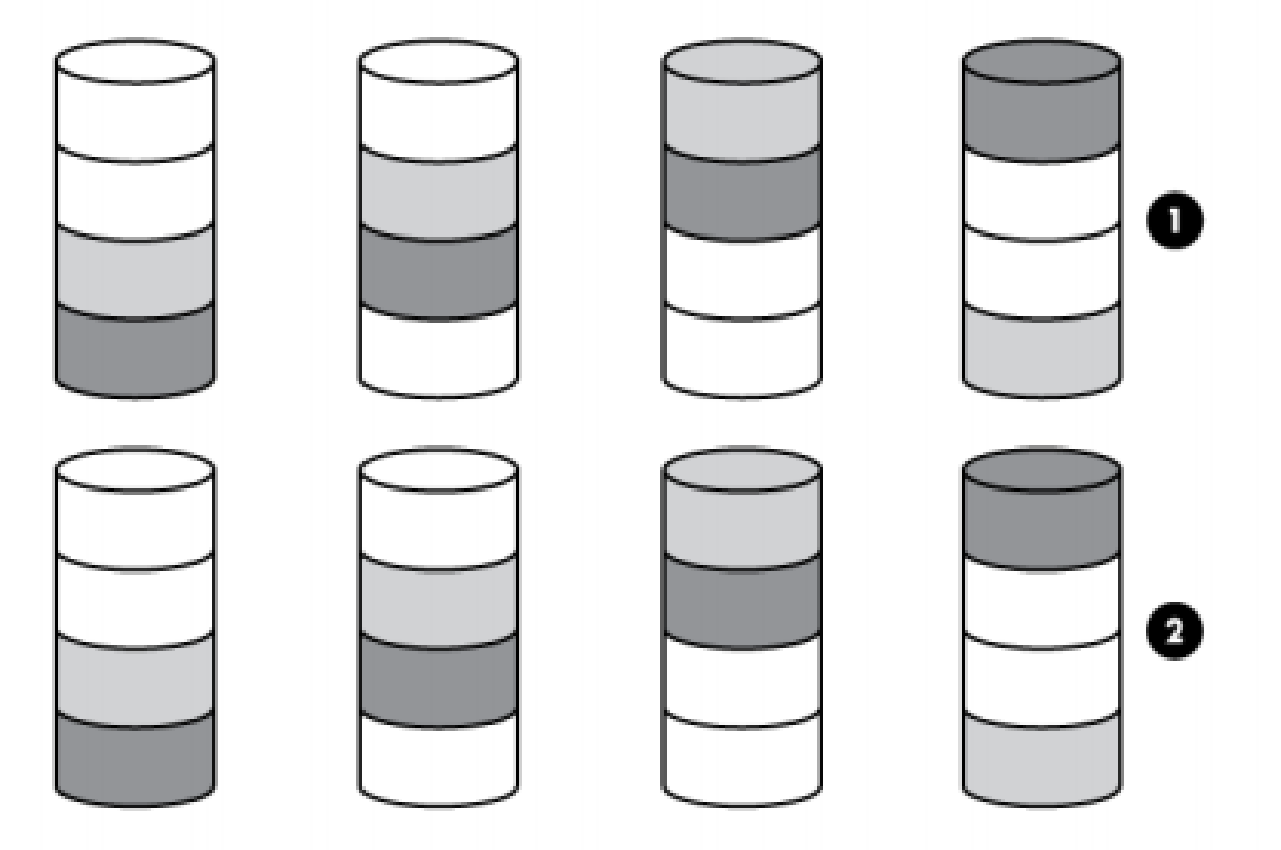
RAID 50

- For any given number of hard drives, data loss is least likely to occur when the drives are arranged into the configuration that has the largest possible number of parity groups. For example, four parity groups of three drives are more secure than three parity groups of four drives. However, less data can be stored on the array with the larger number of parity groups.
- RAID 50 is particularly useful for large databases, file servers, and application servers.

- Advantages:
  - Higher performance than for RAID 5, especially during writes.
  - Better fault tolerance than either RAID 0 or RAID 5.
  - Up to n physical drives can fail (where n is the number of parity groups) without loss of data, as long as the failed drives are in different parity groups.
- Disadvantages:
  - All data is lost if a second drive fails in the same parity group before data from the first failed drive has finished rebuilding.
  - A greater percentage of array capacity is used to store redundant or parity data than with nonnested RAID methods.

RAID 60

- RAID 50 is a nested RAID method in which the constituent hard drives are organized into several identical RAID 5 logical drive sets (parity groups). The smallest possible RAID 50 configuration has six drives organized into two parity groups of three drives each.



RAID 60

- For any given number of hard drives, data loss is least likely to occur when the drives are arranged into the configuration that has the largest possible number of parity groups. For example, five parity groups of four drives are more secure than four parity groups of five drives. However, less data can be stored on the array with the larger number of parity groups.
- RAID 60 is particularly useful for data archives and high-availability solutions.

- Advantages:
  - Higher performance than for RAID 6, especially during writes.
  - Better fault tolerance than either RAID 0 or RAID 6.
  - Up to 2n physical drives can fail (where n is the number of parity groups) without loss of data, as long as no more than two failed drives are in the same parity group.
- Disadvantages:
  - All data is lost if a third drive in a parity group fails before one of the other failed drives in the parity group has finished rebuilding.
  - A greater percentage of array capacity is used to store redundant or parity data than with non-nested RAID methods

Comparing the hardware-based RAID methods

Item	RAID 0	RAID 1+0	RAID 5	RAID 6 (ADG)
Alternative name	Striping (no fault tolerance)	Mirroring	Distributed Data Guarding	Advanced Data Guarding
Formula for number of drives usable for data (*n* = total number of drives in array)	*n*	*n*/2	*n*-1	*n*-2
Fraction of drive space usable*	100%	50%	67% to 93%	50% to 96%
Minimum number of physical drives	1	2	3	4
Tolerates failure of one physical drive	No	Yes	Yes	Yes
Tolerates simultaneous failure of more than one physical drive	No	Only if no two failed drives are in the same mirrored pair	No	Yes
Read performance	High	High	High	High
Write performance	High	Medium	Low	Low
Relative cost	Low	High	Medium	Medium

*Values for the fraction of drive space usable are calculated with these assumptions: (1) all physical drives in the array have the same capacity; (2) online spares are not used; (3) no more than 14 physical drives are used per array for RAID 5; and (4) no more than 56 drives are used with RAID 6 (ADG).

Factors involved in logical drive failure

- The probability that a logical drive will fail depends on the RAID-level setting and on the number and type of physical drives in the array. If the logical drive does not have an online spare, the following results apply:
  - A RAID 0 logical drive fails if only one physical drive fails.
  - A RAID 1+0 logical drive fails if any two failed physical drives are mirrored to each other.
    - The maximum number of physical drives that can fail without causing failure of the logical drive is n/2, where n is the number of hard drives in the array. In practice, a logical drive usually fails before this maximum is reached. As the number of failed physical drives increases, it becomes increasingly likely that the newly failed drive is mirrored to a previously failed drive.
    - The minimum number of physical drive failures that can cause the logical drive to fail is two. This situation occurs when the two failed drives are mirrored to each other. As the total number of drives in the array increases, the probability that the only two failed drives in an array are mirrored to each other decreases.
  - A RAID 5 logical drive fails if two physical drives fail.
  - A RAID 6 (ADG) logical drive fails when three physical drives fail.

Factors involved in logical drive failure

- At any given RAID level, the probability of logical drive failure increases as the number of physical drives in the logical drive increases.
- This principle is illustrated more quantitatively in the graph. The data for this graph is calculated from the MTBF (Mean Time Between Failures) value for a typical physical drive, assuming that no online spares are present. If an online spare is added to any