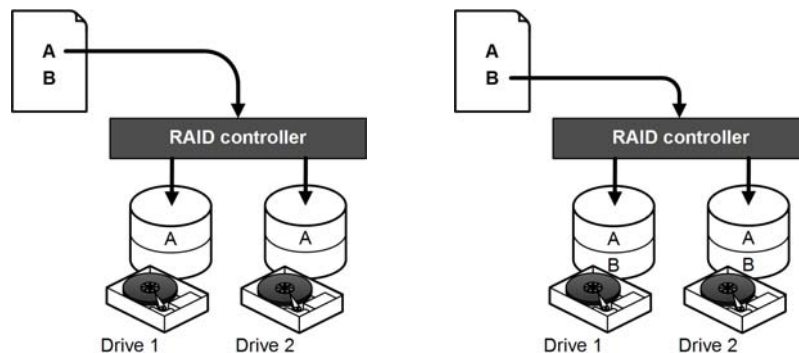


RAID

What is RAID?

RAID 1 — Mirrored Set

If your hard drive breaks then you could lose all the data on it. One solution to this problem is to create a live duplicate of the drive by saving the same data to a second disk drive.

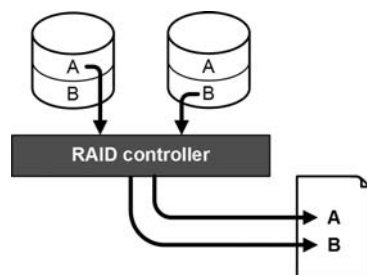


This is called **mirroring** and it adds **redundancy** to the system — if one drive breaks, you still have a full copy of the data.



This is a **Redundant Array of Independent Drives (RAID)** and it is called RAID type 1 (or RAID 1 for short).

The drives contain the same data so there is little point reading the same byte from both; the controller instead reads different bytes from each drive at the same time.



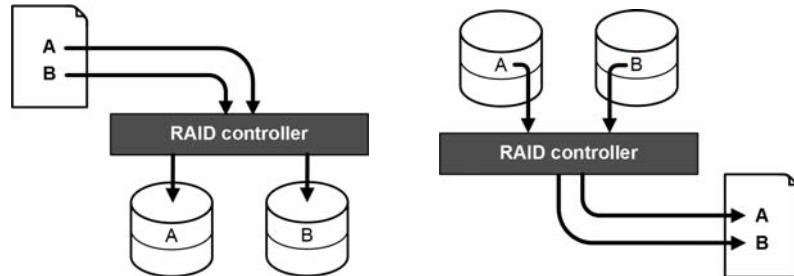
This means that it can read data twice as quickly as normal.

- Advantage: **greater fault tolerance**; **double-speed reading**
- Disadvantage: **cost** of second drive and RAID controller

RAID 0 — Striped Set

We have seen that redundant drives can increase speed.

RAID 0 splits data across different drives to give double-speed writing and reading. This is called **striping**.



However, RAID 0 has a major drawback. If one of the drives breaks then half the data is lost; if every other data byte is missing then the remaining data is useless.

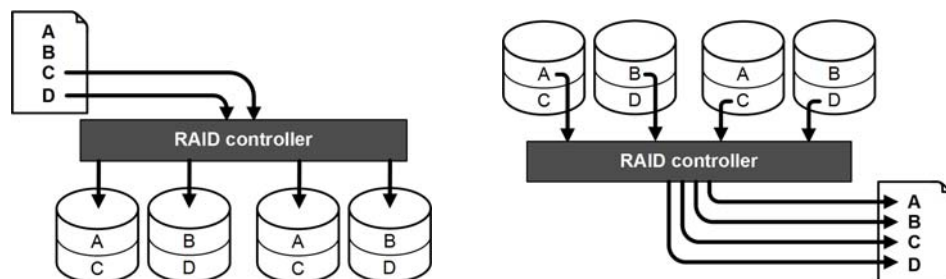


For this reason RAID 0 is rarely used.

- Advantage: **double-speed writing & reading**
- Disadvantage: **lower fault-tolerance** and **cost** of second drive and RAID controller

RAID 0+1 — Striped and Mirrored Set

To counter the problem with RAID 0 some designers create a mirror of the striped set.



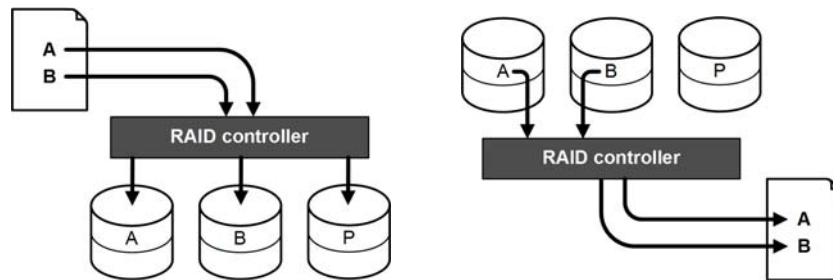
This requires at least four disk drives but gives **double-speed writing** and **quadruple-speed reading**.

- Advantage: **double-speed writing & 4-speed writing;**
greater fault tolerance
- Disadvantage: **high cost** of three extra drives and RAID controller

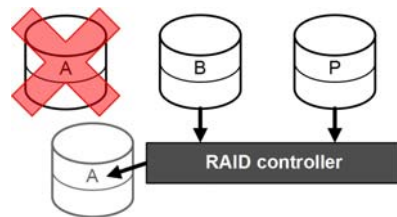
RAID 5 — Striped Set with Distributed Parity

Buying four drives to store one drive's worth of data is very expensive. RAID 5 offers a more cost-effective way to keep a balance of speed and redundancy.

It uses three drives and stores data byte A on one drive, B on another and calculates a **parity code** (essentially a measure of the difference between A and B), which is written to the third drive.



If one of the drives breaks then the data that would have been on it can be re-calculated — on-the-fly — by examining the remaining drives.



This is particularly useful because the broken drive can be removed and replaced without needing to take the whole storage system off-line. The RAID controller can automatically repopulate the drive once replaced. Performance will slow a little during this process but will not stop.

RAID 4 uses a dedicated drive for all parity codes.

RAID 5 distributes parity codes across all three drives, meaning that each drive contains only a third of the codes. This makes data recovery much quicker because fewer re-calculations are needed.

- Advantage: **double-speed writing & reading;**
greater fault tolerance; less expensive than RAID 0+1
- Disadvantage: **cost** of two extra drives and RAID controller

Getting RAID

Hardware RAID

Many cheap RAID cards support RAID 0, 1 and 0+1. These are also sometimes build into motherboards.

More expensive cards will support RAID 5 and other (less widely used) RAID configurations.

Software RAID

The Windows NT family (i.e. 2000 / XP / Vista) can organise data on disks to give RAID 0 and RAID 1.

Server versions (i.e. 2000 Server, Server 2003 and Server 2007) can do the same to provide RAID 5.

To do this, Windows **Disk Management** must be used to de-partition the disks and convert them to a different structure called **dynamic disks** (as opposed to normal disks which Microsoft call **basic disks**).

Software RAID is not as efficient as hardware RAID.