

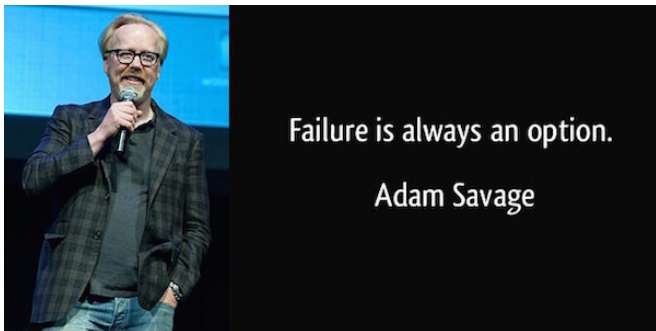
# Lecture 25 — Reliability: RAID

Jeff Zarnett  
jzarnett@uwaterloo.ca

Department of Electrical and Computer Engineering  
University of Waterloo

March 9, 2024

# Failure is Always an Option



Sometimes hard drives die – so now what?

Backups help but they are not the only way.

A backup may be a bit out of date?

Are you sure they work?



Maybe loss of data, maybe system is down, maybe reduced performance?

# Too Good To Be True?

What if I told you there exists a solution to this where you not only get higher reliability and availability, but also maybe better performance?



It does cost money... But it's not magic.

The Solution is Called...



# The Solution is Called...



RAID: Redundant Array of Independent Disks.

Multiple different independent disks that work together but appear to the user as if they are one drive/volume.

Can be managed in software or hardware.



# How Much Redundancy?

Choose level of redundancy based on the data in question.

Is losing data inconvenient? A tragedy? Does it affect uptime?

For any device we can talk about the **mean time to failure**.

This is always an estimate!



Let's say you bought a disk that says its mean time to failure is 100 000 hours (11.4 years)

You might think that this is perfectly fine, even if it lives only half its expected life it's still more than five years and you'll replace it by then!

... but if you are running a data centre and you have 100 of them, the MttF of one of these disks is really only  $100\,000 / 100 = 1\,000$  hours (41.66 days).

The textbook example is misleading – hard drive deaths aren't evenly distributed!

It's more likely that everything is fine for a while and then drives start dying in quick succession.

So if we want to have improved reliability, it's obvious this is going to have a cost.

The cost is in the redundancy; having more than one copy of the data and those copies are stored on different physical storage media.

Either: more money for more disks, or less total capacity for more redundancy.

Obvious approach is **mirroring**.

If one of the disks dies, then the data is on the other disk and it's not lost.

How fast do we have to replace it?

If the mean time to failure is 100 000 hours and the mean time to repair is 10 hours?

The mean time to data loss is  $(100\,000)^2 / (2 \times 10) = 500 \times 10^6$  hours (about 57 000 years).

Is that realistic?

This version means there's additional cost to every write...

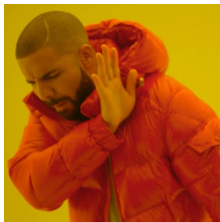
What happens if one succeeds and one fails?



RAID is describes as having different “levels”.

Each has a different configuration; higher number isn't always better.

There are seven basic levels that have broad acceptance.



**RAID**

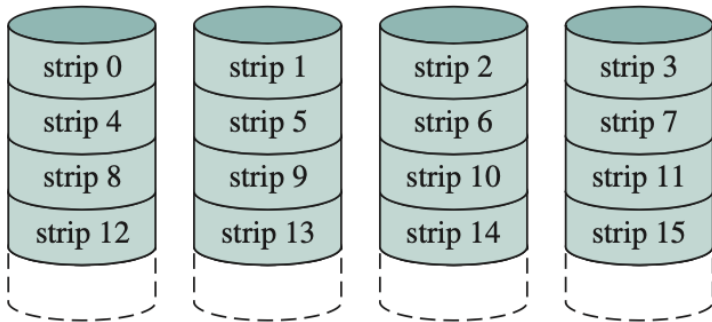


**JBOD**

This is the do-nothing approach.

Manual backups?

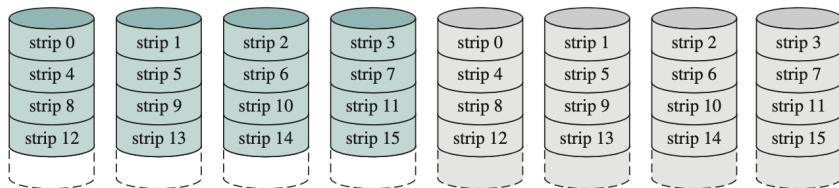
Disk striping; no redundancy.



(a) RAID 0 (nonredundant)

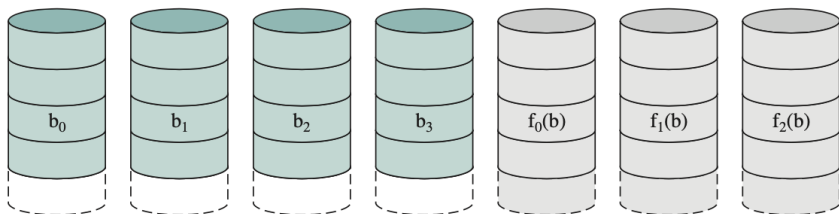
Some use cases for it...

## Mirroring



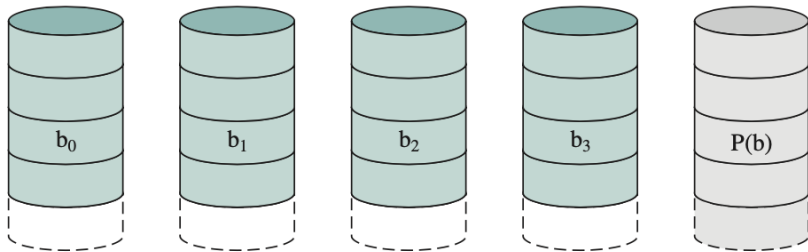
(b) RAID 1 (mirrored)

## Bit Parity

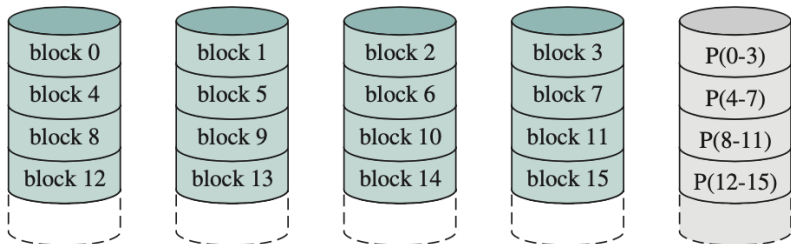


(c) RAID 2 (redundancy through Hamming code)

## Byte Parity

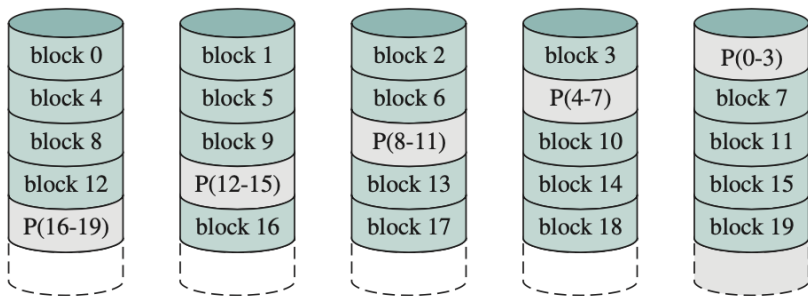


## Block Parity



(e) RAID 4 (block-interleaved parity)

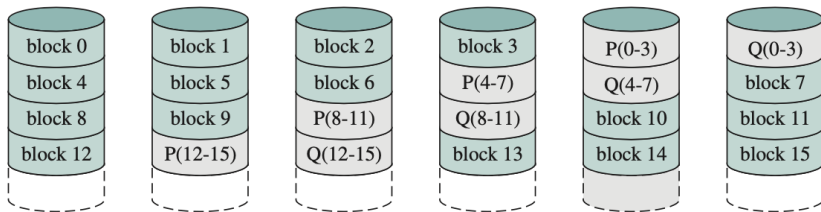
## Distributed Block Parity



(f) RAID 5 (block-interleaved distributed parity)



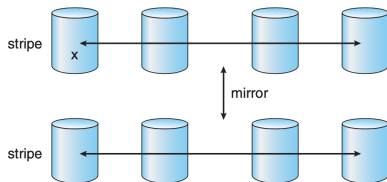
Like RAID 5 but can survive more disk failures.



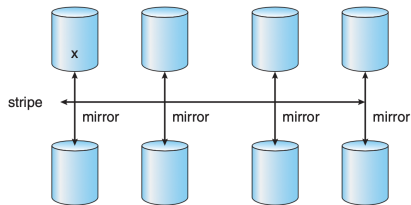
(g) RAID 6 (block-interleaved dual distributed parity)

here are a number of ways that RAID may be combined, such as RAID 01, 10, 50...

These are best thought of as being something like  $1 + 0$  rather than ten, because what we're doing is combining both these things at two different levels.



a) RAID 0 + 1 with a single disk failure.



b) RAID 1 + 0 with a single disk failure.

- How critical is it that data is not lost?
- What is the budget?
- How important is performance?
- Do we need to be able to carry on in the event a disk dies?
- Are rebuild times important?

No one-size-fits-all answers.

RAID 2, 3, 4 not very popular...

Choose what makes sense!