UNIVERSITY of WISCONSIN-MADISON
Computer Sciences Department

CS 537
Introduction to Operating Systems

Andrea C. Arpaci-Dusseau
Remzi H. Arpaci-Dusseau

Storage System: RAID

Questions answered in this lecture:

What is RAID?

How does one trade-off between:
performance, capacity, and reliability?

What is RAID-0, RAID-1, RAID-4, and RAID-5?

Motivation: Why use multiple disks?

Capacity
- More disks allows us to store more data

Performance
- Access multiple disks in parallel
- Each disk can be working on independent read or write
- Overlap seek and rotational positioning time for all

Reliability
- Recover from disk (or single sector) failures
- Will need to store multiple copies of data to recover

RAID: Redundant Array of Inexpensive/Independent Disks

Hardware vs. Software RAID

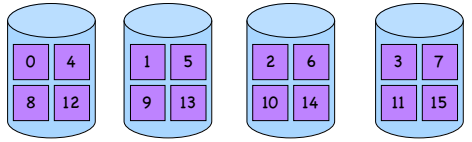
Hardware RAID
- Storage box you attach to computer
- Same interface as single disk, but internally much more
  - Multiple disks
  - More complex controller
  - NVRAM (holding parity blocks)

Software RAID
- OS (device driver layer) treats multiple disks like a single disk
- Software does all extra work

Interface for both
- Linear array of bytes, just like a single disk (but larger)

RAID-0: Striping

Stripe blocks across disks in a "chunk" size
- How to pick a reasonable chunk size?

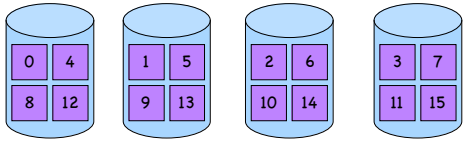


How to calculate where chunk # lives?

Disk:

Offset within disk:

RAID-0: Striping



Evaluate for D disks

Capacity: How much space is wasted?

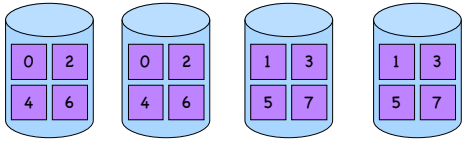
Performance: How much faster than 1 disk?

Reliability: More or less reliable than 1 disk?

RAID-1: Mirroring

Motivation: Handle disk failures

Put copy (mirror or replica) of each chunk on another disk



Capacity:

Reliability:

Performance:

RAID-4: Parity

Motivation: Improve capacity

Idea: Allocate parity block to encode info about blocks

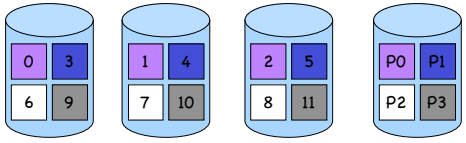
- Parity checks all other blocks in stripe across other disks

Parity block = XOR over others (gives "even" parity)

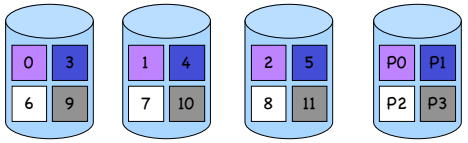
- Example: 0 1 0 --> Parity value?

How do you recover from a failed disk?

- Example: x 0 0 and parity of 1
- What is the failed value?



RAID-4: Parity



Capacity:

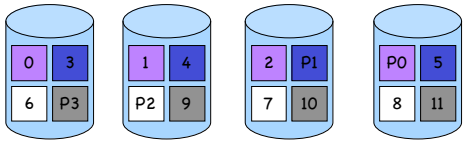
Reliability:

Performance:

- Reads
- Writes: How to update parity block?
  - Two different approaches
    - Small number of disks (or large write):
    - Large number of disks (or small write):
  - Parity disk is the bottleneck

RAID-5: Rotated Parity

Rotate location of parity across all disks



Capacity:

Reliability:

Performance:

- Reads:
- Writes:
- Still requires 4 I/Os per write, but not always to same parity disk

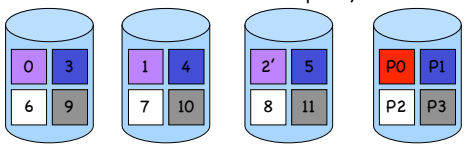
Advanced Issues

What happens if more than one fault?

- Example: One disk fails plus "latent sector error" on another
- RAID-5 cannot handle two faults
- Solution: RAID-6 (e.g., RDP) Add multiple parity blocks

Why is NVRAM useful?

- Example: What if update 2, don't update P0 before power failure (or crash), and then disk 1 fails?
- NVRAM solution: Use to store blocks updated in same stripe
  - If power failure, can replay all writes in NVRAM
- Software RAID solution: Perform parity scrub over entire disk



Conclusions

RAID turns multiple disks into a larger, faster, more reliable disk

RAID-0: Striping
Good when performance and capacity really matter, but reliability doesn't

RAID-1: Mirroring
Good when reliability and write performance matter, but capacity (cost) doesn't

RAID-5: Rotating Parity
Good when capacity and cost matter or workload is read-mostly

Good compromise choice