W4118: RAID

Instructor: Junfeng Yang

References: Modern Operating Systems (3rd edition), Operating Systems Concepts (8th edition), previous W4118, and OS at MIT, Stanford, and UWisc

RAID motivation

- Performance
  - Disks are slow compared to CPU
  - Disk speed improves slowly compared to CPU
- Reliability
  - In single disk systems, one disk failure ➔ data loss
- Cost
  - A single fast, reliable disk is expensive

RAID idea

- RAID idea: use redundancy to improve performance and reliability
  - Redundant array of cheap disks as one storage unit
  - Fast: simultaneous read and write disks in the array
  - Reliable: use parity to detect and correct errors
- RAID can have different redundancy levels, achieving different performance and reliability
  - Seven different *RAID levels* (0-6)

Evaluating RAID

- Cost
  - *Storage utilization*: data capacity / total capacity
- Reliability
  - Tolerance of disk failures
- Performance
  - (Large) sequential read, write, read-modify-write
  - (Small) random read, write, read-modify-write
  - Speedup over a single disk

Computing cost

- *G* = number of data disks in a RAID group
- *C* = number of check disks in a RAID group

- Cost = *C*/(*G*+*C*)

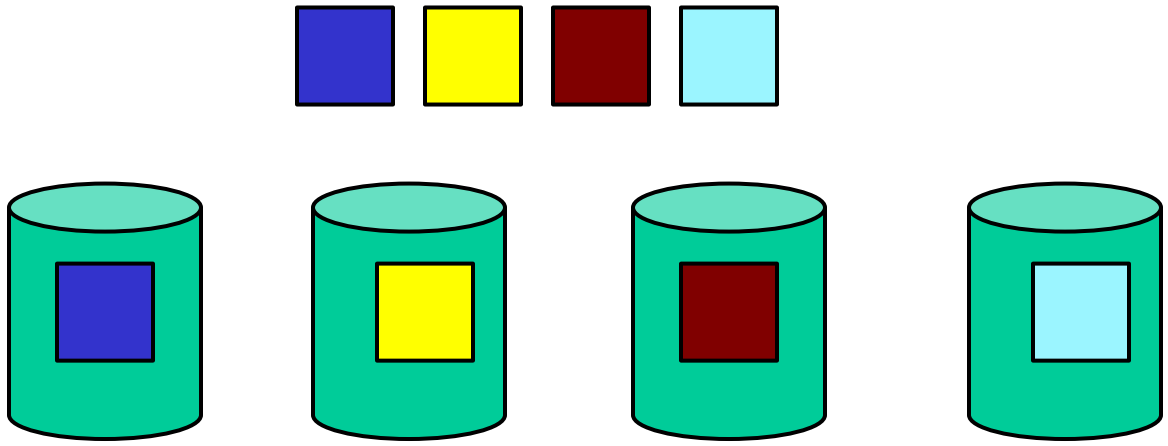
Computing reliability

- N = total number of disks
- G = number of data disks in a RAID group
- C = number of check/parity disks in a RAID group
- MTTFdisk = mean time to failure for a disk
- MTTR = mean time to repair for a failed disk

- MTTFraid = ?

RAID 0: non-redundant striping

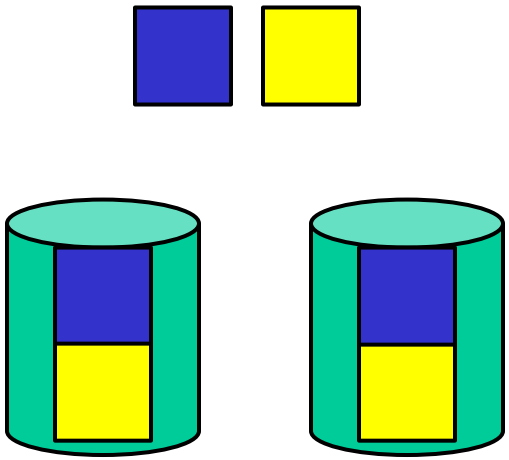
- Structure
  - Data striped across all disks in an array
  - No parity
- Advantages:
  - Good performance: with N disks, roughly N times speedup
- Disadvantages:
  - Poor reliability: one disk failure ➔ data loss



RAID 0 performance

RAID 1: mirroring

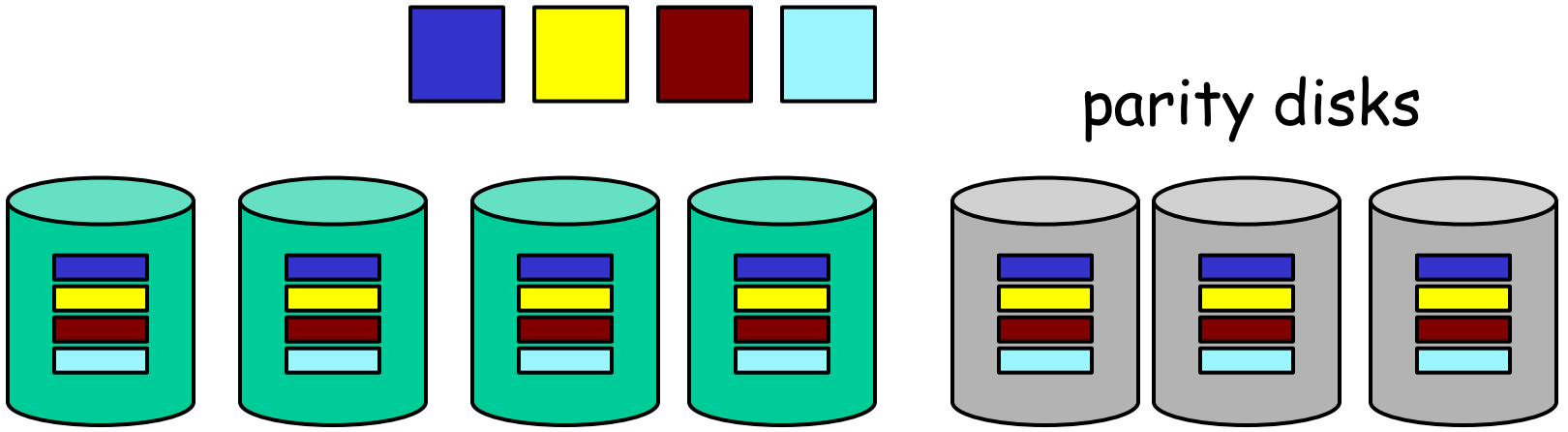
- Structure
  - Keep a *mirrored* (shadow) copy of data
- Advantages
  - Good reliability: one disk failure OK
  - Good read performance
- Disadvantage
  - High cost: one data disk requires one parity disk



RAID 1 performance

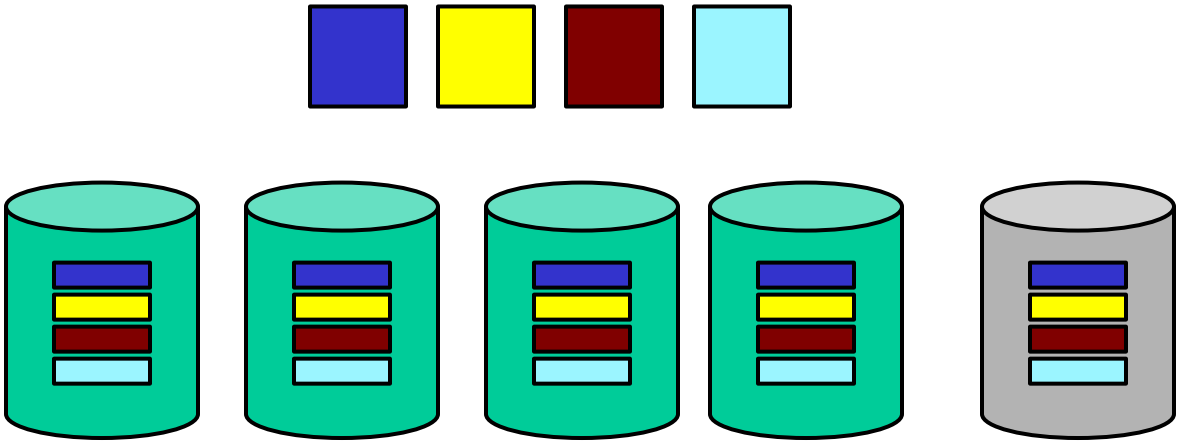
RAID 2: error-correction parity

- Structure
  - A data sector striped across data disks
  - Compute *error-correcting parity* and store in parity disks
- Advantages
  - Good reliability with higher storage utilization than mirroring
- Disadvantages
  - Unnecessary cost: disk can already detect failure
  - Poor random performance



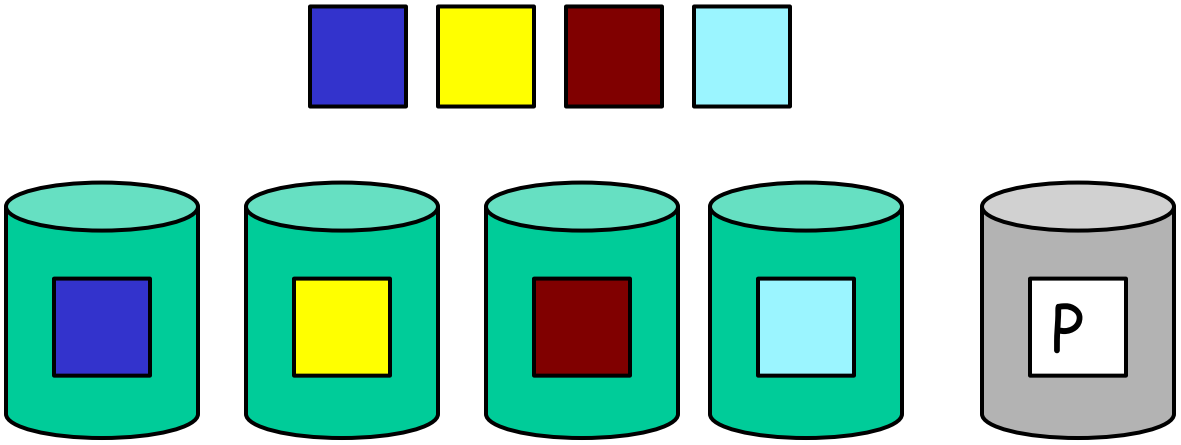
RAID 3: bit-interleaved parity

- Structure
  - Single parity disk (XOR of each stripe of a data sector)
- Advantages
  - Same reliability with one disk failure as RAID2 since disk controller can determine what disk fails
  - Higher storage utilization
- Disadvantages
  - Poor random performance



RAID 4: block-interleaved parity

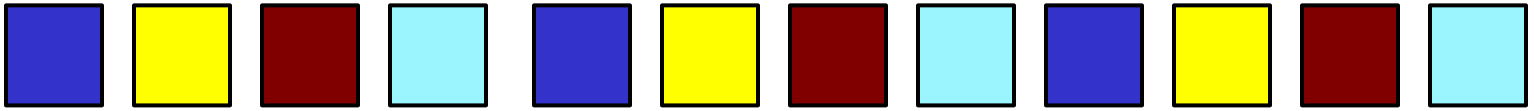
- Structure
  - A set of data sectors (*parity group*) striped across data disks
- Advantages
  - Same reliability as RAID3
  - Good random read performance
- Disadvantages
  - Poor random write and read-modify-write performance

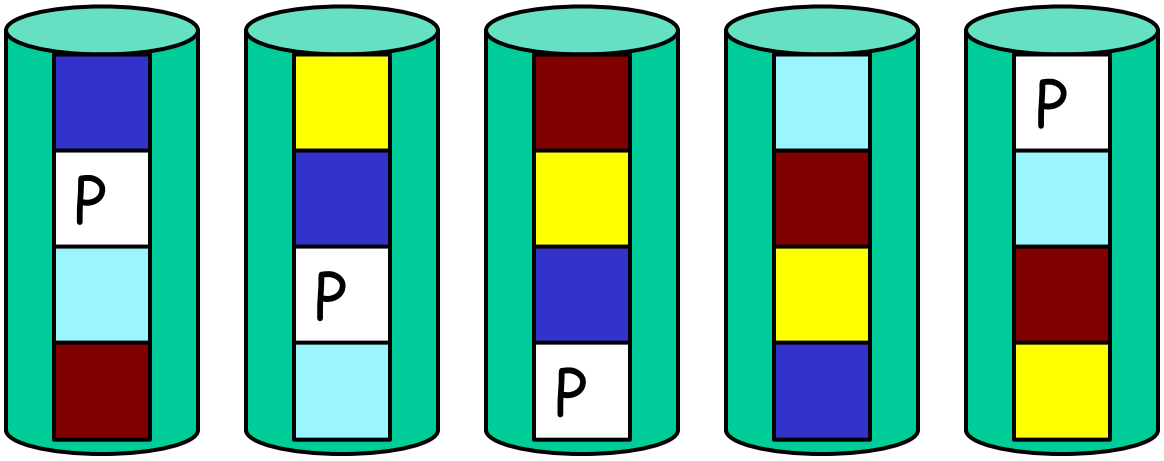


RAID 4 performance

RAID 5: block-interleaved distributed parity

- Structure
  - Parity sectors distributed across all disks
- Advantages
  - Good performance

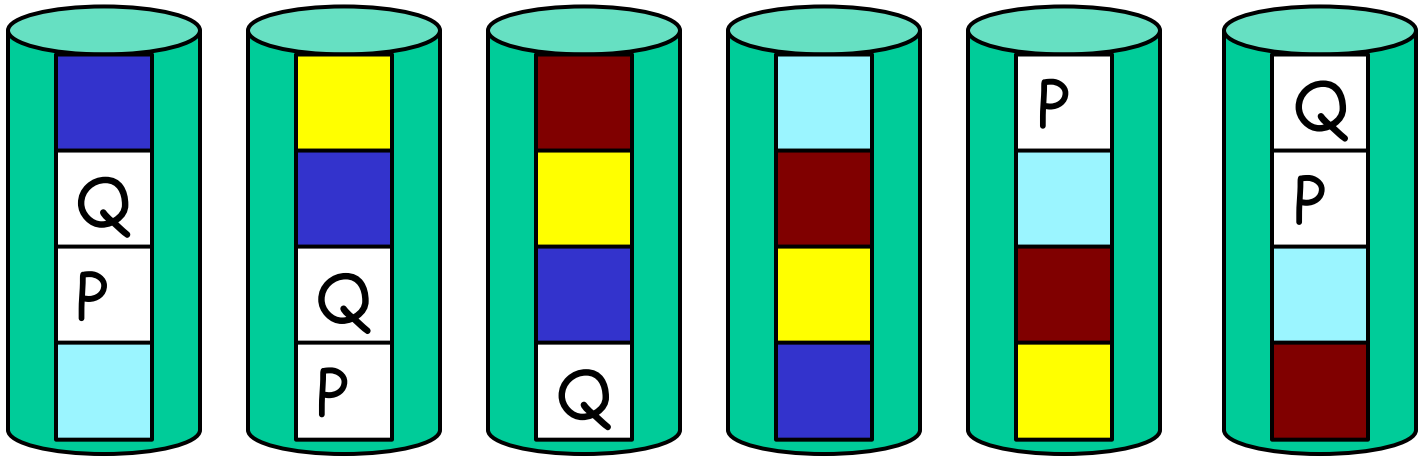




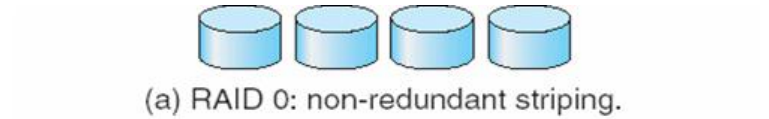
RAID 5 performance

RAID6: P+Q redundancy

- Structure
  - Same as RAID 5 except using two parity sectors per parity group
- Advantages
  - Can tolerate two disk failures



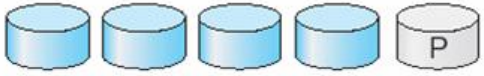
RAID levels



(a) RAID 0: non-redundant striping.

(b) RAID 1: mirrored disks.

(c) RAID 2: memory-style error-correcting codes.



(d) RAID 3: bit-interleaved parity.

(e) RAID 4: block-interleaved parity.

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

(g) RAID 6: P + Q redundancy.