CSE 451: Operating Systems
Autumn 2005

Redundant Arrays of Inexpensive Disks (RAID)

Hank Levy

The challenge

- Disk transfer rates are improving, but much less fast than CPU performance

- We can use multiple disks to improve performance
  - by *striping* files across multiple disks (placing parts of each file on a different disk), we can use parallel I/O to improve access time

- Striping reduces reliability
  - 100 disks have 1/100th the MTBF (mean time between failures) of one disk

- So, we need striping for performance, but we need something to help with reliability / availability

- To improve reliability, we can add redundant data to the disks, in addition to striping
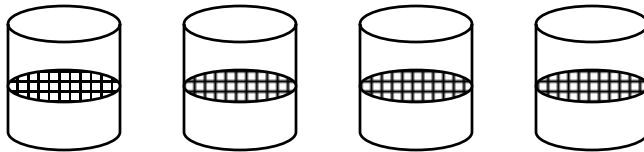
RAID

- A RAID is a Redundant Array of Inexpensive Disks
- Disks are small and cheap, so it's easy to put lots of disks (10s to 100s) in one box for increased storage, performance, and availability
- Data plus some redundant information is striped across the disks in some way
- How striping is done is key to performance and reliability

Some RAID tradeoffs

- Granularity
  - fine-grained: stripe each file over all disks
    - high throughput for the file
    - limits transfer to 1 file at a time
  - course-grained: stripe each file over only a few disks
    - limits throughput for 1 file
    - allows concurrent access to multiple files
- Redundancy
  - uniformly distribute redundancy information on disks
    - avoids load-balancing problems
  - concentrate redundancy information on a small number of disks
   - partition the disks into data disks and redundancy disks

RAID Level 0

- RAID Level 0 is a <u>non-redundant</u> disk array
- Files are striped across disks, no redundant info
- High read throughput
- Best write throughput (no redundant info to write)
- Any disk failure results in data loss

RAID Level 1

- RAID Level 1 is <u>mirrored disks</u>
- Files are striped across half the disks
- Data is written to two places – data disks and mirror disks
- On failure, just use the surviving disk
- 2x space expansion

data disks   **mirror copies**

© 2005 Gribble, Lazowska, Levy
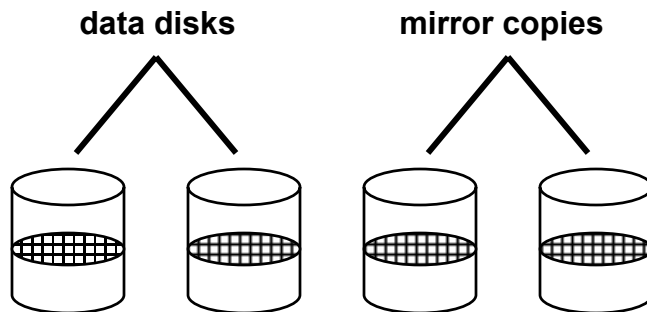
RAID Levels 2, 3, and 4

- RAID levels 2, 3, and 4 use <u>ECC</u> (error correcting code) or <u>parity</u> disks
  - E.g., each byte on the parity disk is a parity function of the corresponding bytes on all the other disks
- A read accesses all the data disks
- A write accesses all the data disks plus the parity disk
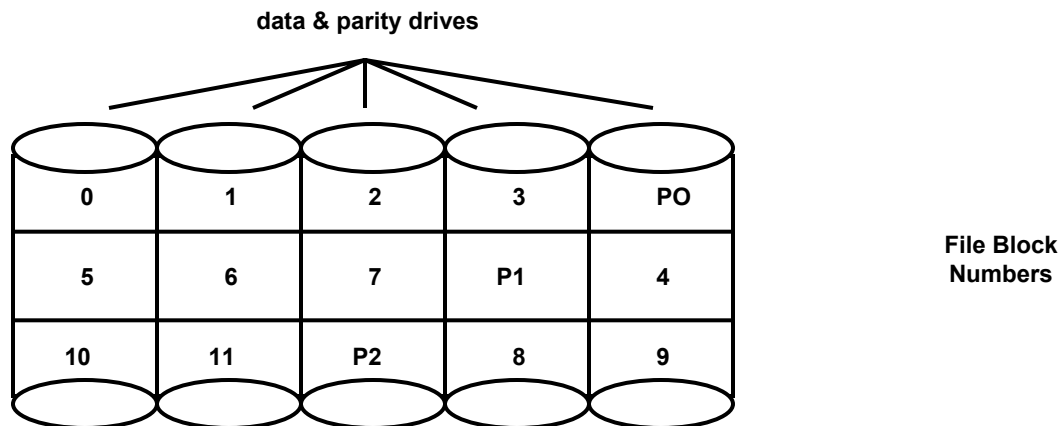- On disk failure, read the remaining disks plus the parity disk to compute the missing data

data disks   **parity disk**

© 2005 Gribble, Lazowska, Levy
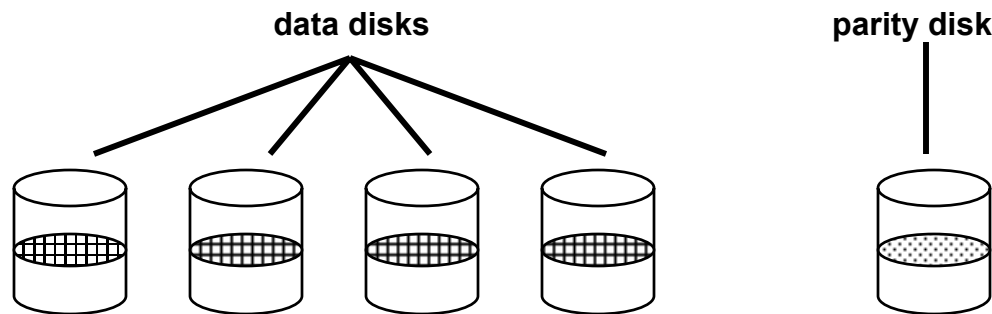
Refresher: What's parity?

1	0	1	1	0	1	1	0	1
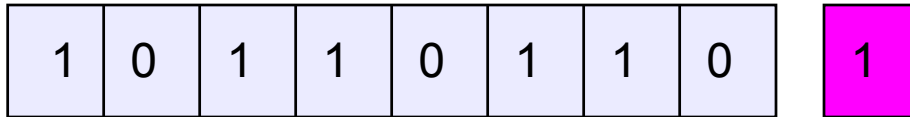
- To each byte, add a bit set so that the total number of 1's is even

- Any single missing bit can be reconstructed

- (Why does memory parity not work quite this way?)

RAID Level 5

- RAID Level 5 uses <u>block interleaved distributed parity</u>
- Like parity scheme, but distribute the parity info (as well as data) over all disks
  - for each block, one disk holds the parity, and the other disks hold the data
- Significantly better performance
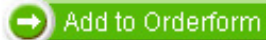 - parity disk is not a hot spot

data & parity drives

0	1	2	3	PO
5	6	7	P1	4
10	11	P2	8	9

File Block Numbers

DELL ™ USA
Dell Higher Education Store

Full Catalog	Software, Peripherals & Par

🛒 Current Orderform | 📋 Retrieve E-Quote | ☑

BACK TO: USA > Dell Higher Education Store > Software & Peripherals   ✉ E-

Product Details

PROMISE TECHNOLOGY
VTrak 15100 RAID Storag
$5,652.95
Usually Ships: 5-7 Days
➜ Add to Orderform

⊕ View Larger Image

Overview	Tech Specs	Highlights
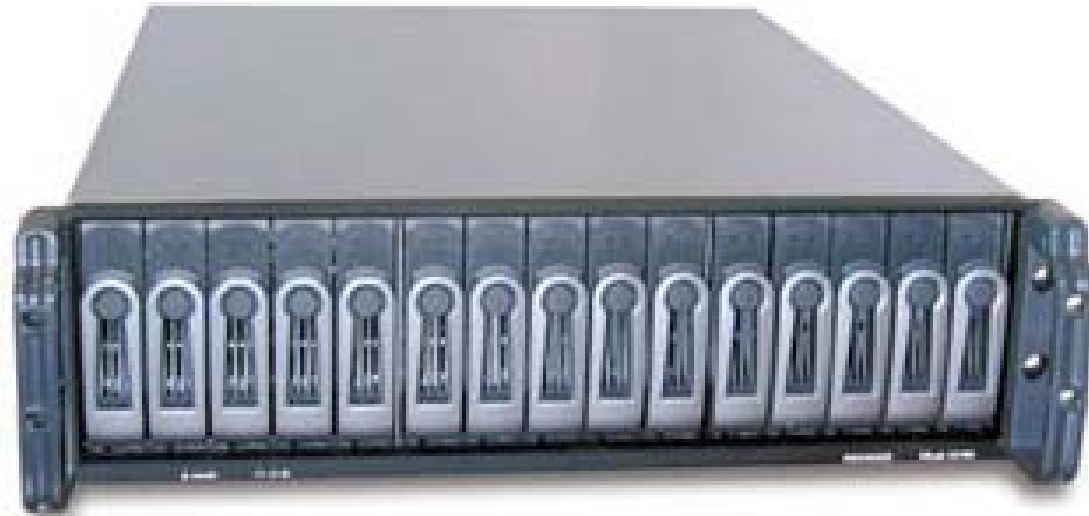
Cache / Buffer Size: 256 MB

Data Transfer Rate: Up to 200 MBps (aggregate using both SCSI channels)

Device Type: RAID Storage System

Dimensions (WxDxH) / Weight: 17.6" x 26" x 5" / 65 lbs (without drives)

Interface Type: SCSI

Port(s) Total (Free) / Connector Type: 2 x External Ultra160 SCSI (VHDCI)

Power: Dual 500 W, 100-240 VAC auto-ranging, 50-60 Hz, dual hot swap and redundant with PFC, N+1 design

Power Consumption Operational: 440 Watts (under load)

RAID Level: RAID 0,1,3,5 or 10 (mirrored stripes), and 50 (striped RAID 5 arrays)

Channel Qty: 2

11/27/2005   © 2005 Gribble, Lazowska, Levy   10