COMP 7970 Storage Systems

Dr. Xiao Qin

Department of Computer Science and Software Engineering
Auburn University
http://www.eng.auburn.edu/~xqin
xqin@auburn.edu
Redundant Arrays of Inexpensive Disks

RAID 4: High I/O Rate Parity

Example:
small read D0 & D5,
large write D12-D15

Block Interleaved Parity
Inspiration for RAID 5

• RAID 4 works well for small reads
• Small writes (write to one disk):
  – Option 1: read other data disks, create new sum and write to Parity Disk
  – Option 2: since P has old sum, compare old data to new data, add the difference to P
• Small writes are limited by Parity Disk: Write to D0, D5 both also write to P disk

COMP 7970, Auburn University
Redundant Arrays of Inexpensive Disks

RAID 5: High I/O Rate Interleaved Parity

Independent writes possible because of interleaved parity

Example: write to D0, D5 uses disks 0, 1, 3, 4

Increasing Logical Disk Addresses

Disk Columns

GOMP 7970, Auburn University
Problems of Disk Arrays
Small Writes

RAID-5: Small Write Algorithm

1 Logical Write = 2 Physical Reads + 2 Physical Writes

D0'  D0  D1  D2  D3  P

new  old
 data data

(1. Read)  (2. Read)

XOR

D0'  D1  D2  D3  P'

(3. Write)  (4. Write)

COMP 7970, Auburn University
RAID 6: P + Q Redundancy

- An extension to RAID 5 but with two-dimensional parity.
- Each row has P parity and each row has Q parity. (Reed-Solomon Codes)
- Has an extremely high data fault tolerance and can sustain multiple simultaneous drive failures
- Rarely implemented

More information, please see the paper:

*A tutorial on Reed-Solomon Coding for Fault Tolerance in RAID-like Systems*
Berkeley History: RAID-I

- RAID-I (1989)
  - Consisted of a Sun 4/280 workstation with 128 MB of DRAM, four dual-string SCSI controllers, 28 5.25-inch SCSI disks and specialized disk striping software
- Today RAID is $24 billion dollar industry, 80% nonPC disks sold in RAIDs
Summary: RAID Techniques: Goal was performance, popularity due to reliability of storage

- **Disk Mirroring, Shadowing (RAID 1)**
  
  Each disk is fully duplicated onto its "shadow"
  Logical write = two physical writes
  100% capacity overhead

- **Parity Data Bandwidth Array (RAID 3)**
  
  Parity computed horizontally
  Logically a single high data bw disk

- **High I/O Rate Parity Array (RAID 5)**
  
  Interleaved parity blocks
  Independent reads and writes
  Logical write = 2 reads + 2 writes