Operating Systems

Youjip Won

KAIST

38. RAID

RAID (Redundant Array of Inexpensive Disks)

- **RAID** is to use multiple disks to build **faster, bigger, and more reliable** disk system.
- RAID is arranged into six different levels.
  - RAID Level 0: Striping multiple disks
  - RAID Level 1: Use mirroring
  - RAID Level 4, level 5: Parity based redundancy

Evalutation

- Capacity
  - N disks, B blocks per disk
  - N*B blocks in total ✉ How much useful capacity is available to the clients of RAID?
- Reliability
  - How many disk faults can the RAID tolerate
- Performance
  - Read
  - write

RAID Level 0

- RAID Level 0 is the simplest form as **striping** blocks.
  - Spread the blocks across the disks in a round-robin fashion.

Disk 0	Disk 1	Disk 2	Disk 3
0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

RAID-0: Simple Striping

RAID Level 0 (Cont.)

- Chunk size
  - Small chunk:
    - more intra-file parallelism
    - Larger positioning time: positioning time is the max positioning time of the disks
  - Large chunk:
    - Reduced intra-file parallelism
      - Smaller positioning time
- An example of RAID Label 0 with a bigger chunk size
  - Chunk size : 2 blocks (8KB)

Disk 0	Disk 1	Disk 2	Disk 3	
0	2	4	6	**chunk size: 2 blocks**
1	3	5	7	
5	10	12	14	
9	11	13	15	

Striping with a Bigger Chunk Size

RAID Level 0 Analysis

- Evaluate the capacity, reliability, performance of striping.
  - First way: single request latency
    - How much parallelism can exist during a single I/O operation.
  - Second way: steady-state throughput of the RAID:
    - Total bandwidth of many concurrent requests.

RAID Level 0 Analysis (Cont.)

- Single Disk
  - Average seek time: 7 ms
  - Average rotational delay: 3 ms
  - Transfer rate of disk: 50 MB/s
- Single Disk Performance
  - 10 Mbyte seq. IO, S = $\frac{Amount\ of\ Data}{Time\ to\ access} = \frac{10\ MB}{(7+3+200)=210\ ms}$ = 47.62 MB /s
  - 10 Kbyte Random IO, R = $\frac{Amount\ of\ Data}{Time\ to\ access} = \frac{10\ KB}{(7+3+0.195)=10.195\ ms}$ = 0.981 MB /s
- RAID 0
  - Random write, random read = N*R
  - Sequential write, sequential read = N*S

RAID Level 1

- RAID Level 1 is mirroring
  - Copy more than one of each block in the system.
  - Copy block places on a separate disk to tolerate the disk failures.

Disk 0	Disk 1	Disk 2	Disk 3
0	0	1	1
2	2	3	3
4	4	5	5
6	6	7	7

Simple RAID-1: Mirroring

RAID level 1

- Capacity N*B/2
- Reliability
  - From one to upto N/2 depending upon the failure disk
- Performance
  - Sequential write: N*S/2
  - Sequential read: N*S/2
  - Random write: N*R/2
  - Random Read: N*R

RAID Level 4

- RAID Level 4 is to add redundancy to a disk array as **parity**.

* P: Parity

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	1	2	3	P0
4	5	6	7	P1
8	9	10	11	P2
12	13	14	15	P3

Simple RAID-4 with parity

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	0	1	1	xor(0,0,1,1)
0	1	0	0	Xor(0,1,0,0)

RAID Level 4 (Cont.)

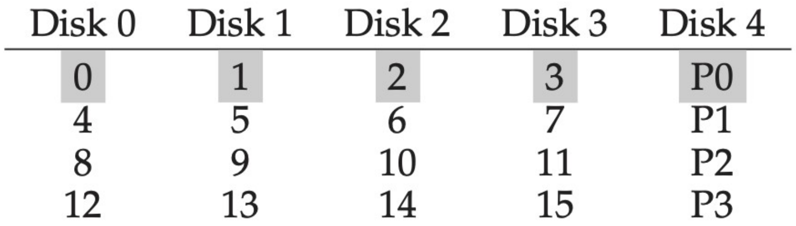
- The simple RAID Level 4 optimization known as a **Full-stripe write.**
  - Calculate the new value of P0 (Parity 0)
  - Write all of the blocks to the five disks above in parallel
  - Full-stripe writes are the most efficient way

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	1	2	3	P0
4	5	6	7	P1
8	9	10	11	P2
12	13	14	15	P3

Full-stripe Writes In RAID-4

Anlaysis

- Capacity: (N-1)*B
- Sequential read: (N-1)*S
- Sequential write: (N-1)*S for full stripe write



Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	1	2	3	P0
4	5	6	7	P1
8	9	10	11	P2
12	13	14	15	P3

- Random read: (N-1)*R

Analysis

- Random write:
  - Additive Parity update: read all blocks, update the block, compute the new parity and write the updated block and the updated parity.
  - Subtractive parity update: read the parity, write (new xor old) xor (old parity). (read on parity disk)
  - For each write, the RAID perform **4 physical I/O**. (two read and writes)

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	1	2	3	P0
*4	5	6	7	+P1
8	9	10	11	P2
12	*13	14	15	+P3

Random write performance: (R/2) MB/sec

Small write problem happens

RAID Level 5

- RAID Level 5 is solution of small write problem.
  - small write problem cause parity-disk bottleneck of RAID Level 4.
  - works almost identically to RAID-4, except that it rotates the parity blocks across drives.
- RAID Level 5's Each stripe is now rotated across the disks.

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
0	1	2	3	P0
5	6	7	P1	4
10	11	P2	8	9
15	P3	12	13	14
P4	16	17	18	19

RAID-5 with Rotated Parity

Analysis

- Capacity: (N-1)*B
- Reliability: 1
- Performance
  - Sequential read, sequential write: (N-1)S
  - Random read: N*R
  - Random write: single write can cause 4 IO's (two read, two write), All N disks can work in parallel: (N*R)/4

Summary

	RAID-0	RAID-1	RAID-4	RAID-5
Capacity	N	N/2	N-1	N-1
Reliability	0	1 (for sure) N/2 (if lucky)	1	1
Throughput				
Sequential Read	NS	(N/2)S	(N-1)S	(N-1)S
Sequential Write	NS	(N/2) S	(N-1)S	(N-1)S
Random Read	NR	NR	(N-1)R	NR
Random Write	NR	(N/2)R	R/2	(N/4)R
Latency				
Read	D	D	D	D
Write	D	D	2D	2D