RAID

Last class
Disks, Disk Scheduling

One disk is NOT enough!

1. capacity
2. performance
3. reliability

Redundant Array of Inexpensive Disks

RAID-O (Striping)

```
Disk 0
0  4  8  12

D1  1  5  9 13

D2  2  6 10 14

D3  3  7 11 15
```

"transparently"

SCSI

HDD

CPU

Memory

RAID

...
disk = Addr % NUM_DISKS

offset = Addr / NUM_DISKS

---

<table>
<thead>
<tr>
<th>Chunk</th>
<th>D0</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>size of chunk</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>= 2 blocks</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>= 8 KB</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

1 block = 4 KB

---

FAULT MODEL

1. FAIL-STOP
   - Working
   - Failed

2. Disk Corruption

3. Latent Sector error (1 sector - inaccessible)

---

RAID can detect disk failure easily

---

Workloads

Seq. (min: 1MB)  Rand, (small, 4KB)

Read  Write  R  W
Perf. Analysis

1. Latency $\leq$ 1 Write
   $\quad$ (T - time to R/W 1 block in a disk)

2. Throughput $\leq$
   $\quad$ Seq $\leq$ R
   $\quad$ Rand $\leq$ W

Seq. R/W - S MB/s (1 disk)
Rand. R/W - R MB/s (1 disk)

S $\gg$ R

RAID-0 Analysis

capacity
$\quad= N \cdot B$ blocks

N - Disks
B - blocks/disk

Reliability = 0

Perf.

Latency $\leq$ 1 Read (T)

I Read (T)

I Write (T)
Throughput

Seq.:
\[ R = \frac{3}{N} \ N \cdot S \text{ MB/s} \]

\[ W = N \cdot S \text{ MB/s} \]

Rand.:
\[ R = N \cdot R \text{ MB/s} \]

\[ W = N \cdot R \text{ MB/s} \]

Time = \( \frac{1}{S} \) sees

Rate = S MB/s
Time = \( 2(\frac{1}{S}) \) sees

<table>
<thead>
<tr>
<th>RAID - D</th>
<th>Rate = 2 \times S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

RAID - 1 (Mirroring)

<table>
<thead>
<tr>
<th>Do</th>
<th>D₀</th>
<th>D₁</th>
<th>D₂</th>
<th>D₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
RAID-1 Analysis

\[ \text{capacity} = \frac{N \cdot B}{2} \]

\[ \text{reliability} = \left( \frac{N}{2} \right) \quad \text{lucky!} \]

\[ = 1 \quad \text{(always)} \]

\[ \text{perf. latency} < R = T \]

\[ < W = T \]

\[ \text{R} = \frac{2 \cdot N}{S} \cdot \left( \frac{S}{2} \right) \]

\[ \text{W} = \left( \frac{N}{2} \right) \cdot S \]

\[ R = N \cdot R \quad \text{MB/s} \]

\[ W = \frac{N \cdot R}{2} \quad \text{MB/s} \]

Read: 0, 1, 2, 3, 4, 5, 6, 7.

D0 D2 D1 D4  D0 D2 D1 D3
### RAID-4 (Parity)

<table>
<thead>
<tr>
<th>Do</th>
<th>D₁</th>
<th>D₂</th>
<th>D₃</th>
<th>Parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>P₀⁻³</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>P₄⁻⁷</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>P₈⁻¹¹</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>P₁₂⁻¹⁵</td>
</tr>
</tbody>
</table>
RAID-4 Analysis

capacity = (N-1) \cdot B

Reliability = 1

\[ \begin{array}{cccc}
  C_0 & C_1 & C_2 & C_3 \\
  0 & 1 & * & 1 & 0 & 0 \\
  1 & 0 & x & 0 & 1 & 1
\end{array} \]

Write

How can writes happen in RAID-4?

Additive parity

Write \( C_3 \):

Read \( C_0, C_1, C_2 \)

\[ C_3 = C_0^1 \oplus C_0 + C_1 + C_2 \]

Sub. parity

1. read \( C_3 \) old
2. read \( C_4 \) old (parity)
3. \( P_{new} = (C_3^{old} \oplus C_4) \oplus P_{old} \)
4. Write \( C_3 \) new
5. Write \( P_{new} \)
Throughput

\[ R = (N-1) \cdot S \text{ MB/s} \]

\[ W = (N-1) \cdot S \text{ MB/s (full-stripe write).} \]

\[ R = (N-1) \cdot R \text{ MB/s.} \]

\[ W = \left( \frac{R}{2} \right) \text{ MB/s.} \]

\[ p_{0-3}' = c_0' \oplus g' \oplus c_2' \oplus c_3' \]

WRITE (in parallel).

Write (4, 13) in parallel.

1. Read 4 from D0
1. Read 13 from D1
2. Write 4' to D0
2. Write 13' to D1
2. Write P_{4-7} to D4
2. Write P_{12-15} to D4.