Block devices - disk, RAID, Flash

Disk access time = seek delay + rotational delay + transfer time

Example: seek: 1-10 ms average: 8 ms
         rotational delay: 0 - 10 ms: 7200 RPM: 6 ms
         transfer time: ≈ 80 MB/s: ≈ 0.016 ms/k

         ≈ 13 ms

Disk access time can be reduced by making disk rotate faster and increasing the swing of the disk drive pin.

Seagate

size: 2 x 10^12 bytes
     = 120 MB/s
     18 platters / 8 heads

16,666 x = 248 ms

1 / 2 x 10^9 Rs @ 7200 RPM

Time per sector = 8 surfaces: 250,000 tracks/inch
cylindrical block size addressing used earlier
Recently used: 512-byte blocks
Block number: 0 ... N-1
if 2.1Tb disk = 4x10^12 blocks

2 primary interfaces used for disk drives
SCSI
IDE

Teycomputer: [read command/status]

address (disk: c/h/l/s)

data

SCSI: Small computer systems interface

<table>
<thead>
<tr>
<th>CPU</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controller</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

doesn't access memory directly
CPU sends out a packet named (cmd)

controller

CPU

Logical Unit Number

data packets

OK

status code
- Uses parity drive to recover from failure.

\[ \text{XOR} \]

\[ \text{parity } (1010) = 1 \]
\[ (10010) = 0 \]

- Can compute the value of the failed bit by computing its value from the rest of the bits with the parity bit.

- Block 1, byte 0 bits, block 2, block 3, parity bit.

Performance problem:
- Updating: Eg: update block 5

\[ \text{read 5 (old)} \]
\[ \text{read parity} \]
\[ \text{write 5} \]
\[ \text{write P} \]

- Bottleneck: Access to disk having parity is more
- RAID5 arrangement: RAID-5

\[ \begin{array}{cccc}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & P \\
9 & 10 & P & 11 \\
13 & P & 14 & 15 \\
\end{array} \]
SCSI
- 68-pin ribbon cable
USB, Firewire
- use SCSI commands
SAS - Serial Access SCSI

Fibre channel - new protocol
- designed to allow switches

Two components to SCSI addressing
1. Block number
2. Way of addressing the disk - device LUN

Logical Unit Number

RAID - Redundant Array of Inexpensive Disks
Thinning: fault tolerance through redundancy

System

\[ \text{block device} \quad \text{interface} \quad \text{RAID 1} \]

write read: write to both disk
read - can do from either
when one disk fails, and read would come through
use the other for the read
RAID 1

\[
\begin{array}{ccc}
1 & 1 \\
2 & 2 \\
3 & 3 \\
4 & 4 \\
\end{array}
\]

logical

\[
\begin{array}{ccc}
disk 1 & disk 2 \\
\end{array}
\]

Striping: efficiency

RAID 0

get performance: 2 drives with appearance of 1 drive

\[
\begin{array}{ccc}
1 & 1 & 2 \\
2 & 3 & 4 \\
3 & 5 & 6 \\
4 & disk 1 & disk 2 \\
\end{array}
\]

write: write to both disks

reliability: failure of one drive can cause failure of the whole volume

RAID 4

mirroring: 100% overload in keeping another disk

parity

\[
\begin{array}{ccc}
1 & 2 & 3 & P \\
4 & 5 & 6 & P \\
7 & 8 & 9 & P \\
\end{array}
\]