Block devices - disk, RAID, flash

Where the data comes from?
How it behaves?

These devices have blocks of data which you store and retrieve.

Performance characteristics of a disk:

- move the head to the point of data (seek)
- the head has to move to position itself
  This is called rotational delay

Hence,

\[ \text{disk access time} = \text{seek} + \text{rotational delay} + \text{transfer time} \]

Seek = 1-10 ms
delay
Rotational = 6-10 ms (@ 7200 RPM)
delay
Transfer = 80 MB/s
time

Average seek delay : 8 ms

Rotational delay : 5 ms

Avg. transfer time : 0.015 ms/1k

To decrease access time:

- increase rpm (more power is required)
- making disk drives lot bigger (expensive)
Seagate Disk Drive has

Size: 2x10^8 bytes.

120 MB/s (transfer time)

4 platters / 8 heads

16,666 s to read all data of the disk drive [20G]

\[ \frac{16,666}{60} = 278 \text{ minutes} \]

\[ = 2 \times 10^9 \text{ Rs} @ 7200 \text{ rpm} \]

\[ \div 8 \text{ surfaces} = 250,000 \text{ tracks/inch} \]

How to address data on disk?

→ Old way used

Cylinder/head/sector addressing

Why they gave up this addressing?

→ Nowadays we have

Block number [for each block on disk]

0...N-1 [512-byte blocks]

2 primary interfaces used for disk drive

1. SCSI
2. IDE

How IDE works: [old days]

Read

1. put the addr
2. Specify the Command
3. Read the data

Write

1. write the data
2. Specify addr
3. Specify the Strokes

[Diagram of disk drive and cylinder read]
SCSI

Small computer systems interface

Works like network protocol

Commands: read/write

Parameters: Count

addy [block number]

CPU

Controller

SCSI bus

SCSI - 3

68 pin ribbon cable

[LSA, LSCSI]

USB, Firewire

SAS (Serial Attached SCSI)

Fibre Channel

Disk 2

Read Starting

@ block xyz

Count = 5 blocks

data

data

data

OK (status)

Logical block numbering so that each block is next to the other block.

Single Surface

Multiple Surface

Fibre Channel
SCSI addressing

```
device + LUN (logical unit no.)
[ disk 2 ]
@ block XYZ
Count : 5 blocks
```

1. Identify the device which has blocks on it.
2. Specify the block no.
3. Count

RAID: Redundant Array of Inexpensive Disks

To prevent failing of disks
1. Make the disks robust
2. Have Redundant disks.

1. Write
   => write it on both mirrors
2. Read
   => Read from any mirror (parity)
3. If one of them dies, we have other one functional.

RAID 1

```
   logical
exposed to
host  \{ 1 2 3 4 \}

drive 1
\[ \begin{array}{c}
1 \\
2 \\
3 \\
4 \\
\end{array} \]

drive 2
\[ \begin{array}{c}
1 \\
2 \\
3 \\
4 \\
\end{array} \]
```

Mirroring gives fault tolerance through redundancy.
Striping to give performance
Striping

RAID 0

logical
\[
\begin{align*}
\begin{array}{c}
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
\end{array} & \Rightarrow \\
\begin{array}{c}
\text{drive 1} \\
\text{drive 2} \\
\text{drive 3} \\
\text{drive 4} \\
\text{drive 5} \\
\text{drive 6} \\
\end{array}
\end{align*}
\]

This gives twice the throughput of single disk.

→ If one fails volume goes down (lost blocks)

RAID 4

drive 1

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

Any failure can be recovered using the parity.

Use parity to recover from failure

\[
\text{Parity} (10110) = 1 \quad \text{even 1'3)}
\]

\[
(100010) = 0 \quad \text{even 1'3)}
\]

\[
\text{Calculate parity}
\]

byte 1

\[
\begin{array}{c}
\text{byte 1} \\
\text{byte 2} \\
\text{...}
\end{array}
\]
RAID 4 has significant performance disadvantage

Parity drive becomes a bottleneck when servicing any cluster during random writes.

RAID 5