Storage: Hard Disk Drives

- Non-volatile magnetic storage
- Rapidly rotating platters with two surfaces
- Disk head reads from and writes to the magnetic surface

Hard Disks

- Rotation typically measured in RPM = rotations per minute
- Heads are attached to disk arms
- Arms can be moved through a (single) arm assembly
- Data is stored in fixed-sized sectors
- A circle of sectors is a track

Total capacity of a disk drive

- (bytes per sector) * (sectors per track) * (tracks per surface) * (surfaces per platter) * (platters per disk)

Geometry

- Early disks: fixed number of sectors per track
- Modern disks: multiple zones
Hard Disk Access Time

- **Seek time**
  - Move arm over target track
  - Minimum seek time
    - move to adjacent track
  - Maximum seek time
    - Move from innermost to outermost track or vice versa
  - Average seek time
    - Often approximated through time to seek one third of the way across the disk
    - “Geometric” average, not representative for any workload

- **Rotation time**
  - Once the head has settled on the target track, this is the time until the disk has rotated under the head to the target sector
  - Typical rotation: 4200 RPM to 15000 RPM
  - Average rotation latency: time for ½ rotation
    - Typical latencies: 2 ms - 7.5 ms

- **Transfer time**
  - Time to transfer data from/to disk
    - From/to buffer

  \[ \text{Disk access time} = \text{Seek time} + \text{Rotation time} + \text{Transfer time} \]

Disk Scheduling

- Disk access can be costly
  - Moving the disk arm
  - Waiting for the platters to rotate

- Order in which requests are served can make a significant difference
  - CPU is much faster than a disk access

- Scheduling problem
  - Can be done by OS, disk firmware, or both

  \[ \text{SSTF} \]
  - Shortest seek time first
  - Pick the request closest to the current head position
  - Akin to Shortest Job First in task scheduling
  - Can lead to starvation…
Disk Scheduling

- **SCAN**
  - Sometimes called the elevator algorithm
  - Idea: “preserve momentum”
  - E.g., disk arm first sweeps from inner towards outer track
  - Serving all requests between current and outermost track

- **C-SCAN**
  - Head only moves in one direction
  - When reaching border, return to start position
  - More uniform access time

**Question!**

- What would be the average read time of a disk that has 5400 RPM and average seek time of 20ms?
- 5400 RPM = 90 RPS = 90 Hz
- So a rotation happens every 11.11.. ms
- On average we will have to wait a half of a rotation to get the data
- \(20 + \frac{11.11}{2} = 25.55\) ms

**RAID**

- **Redundant Array of Inexpensive Disks**
  - **Striping**
    - Improve performance
    - Does not improve resilience
  - **Mirroring**
    - Can improve performance
    - If read from disk with least seek time
    - Improves resilience
RAID

- Dedicated parity disk
- All spindles need to be synchronized
- Data is striped on a byte granularity
- With three disks, can handle the loss of one disk
- Stripe the parity
- Stripe the data on a block granularity

Hybrid RAID

- RAID 1+0
- Mirroring a pair of striped disks

Solid State Drive (SDD)

- No moving parts, stores data in “non-volatile” memory
- Usually NAND-Flash
  - Can read and write data at once but only at page granularity
  - Typical page size: 2kB to 4kB
- Random access is cheap
  - Read access latency is low
    - Typically < 0.1ms for NAND-Flash SSD
    - A fast HDD has ~2.9 ms, a typical laptop HDD has 12 ms
  - Write access more costly, requires full erase first (order of ms)
- Problem: cost
- Problem: wear

Flash Translation Layer

- Indirection between logical blocks and physical blocks
- Count write accesses and dynamically remap blocks (ideally to such already erased)
- Wear-leveling
  - Implemented in the device firmware
  - Use a map of logical block addresses to physical block addresses
- Dynamic wear-leveling
  - Every write goes to a new physical block
- Static wear-leveling
  - Also periodically remap blocks that do not change
Hybrid drives

- Can’t decide between the quick access of SSD and the durability of HDD? Get a hybrid drive.

**Manual Configuration**
- Put your Operating System and any programs you know you will want quick access but also won’t change too often

**Automatic configuration**
- Use disk scheduling to dynamically move frequently used block on to the SSD. Great disk cache.