Storage

Introduction

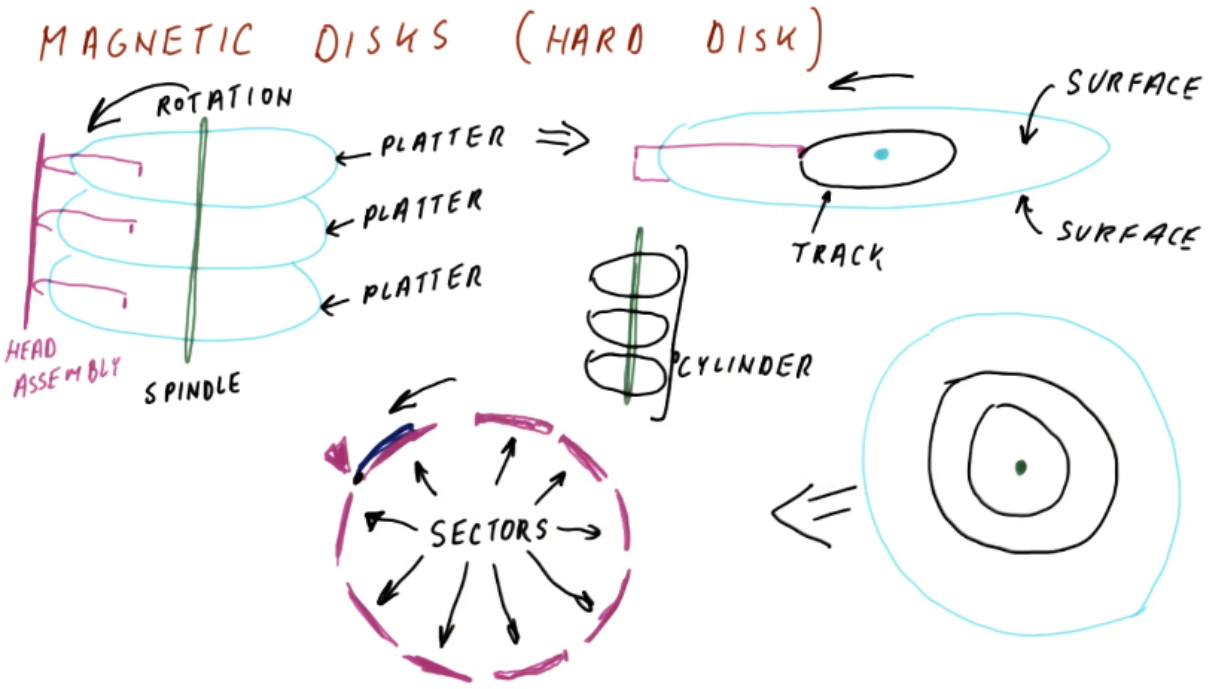
1. Explain how hard drives are connected to the rest of the computer
    - Understand why they're slower but can store so much more data

Storage

1. Files: Programs, data, settings
2. Virtual memory
3. Performance
    - Throughput: Improving, but not as quickly as processor speed
    - Latency: Improving very slowly, even slower than DRAM
4. Reliability
    - If a disk fails, we lose our data and program settings
    - Much worse than a processor failing
5. Types of storage
    - Magnetic disks
    - Optical disks
    - Tape
    - Flash

Magnetic Disks

1. Spindle with platters
    - Motor drives the spindle to rotate
    - Each platter has a surface covered with magnetic material on both sides
    - Magnetic head on both sides of platter
        - Attached to head assembly to move all heads to different points on the platters
        - All heads can access the same radius on the platters, called a track
    - Cylinder is all tracks across all platters
    - Tracks are stacked very close vertically, requiring that the head assembly be very precise
2. Data along one track is stored in sectors
    - Each sector has some bit pattern to indicate the beginning/end of sector
3. Disk capacity
    - num_platters * 2 * num_tracks / surface (num_cylinders) * num_sectors / track * num_bytes / sector
        - Number of platters: Typically 1-4
        - Number of tracks / surface: 1000s
        - Number of sectors per track: 10s - 100s
        - Number of bytes per sector: On the order of 1 kB



Storage Construction

Access Time for Magnetic Disks

1. Total access time is composed of the following:
   - Seek time: Move the head assembly to the correct cylinder
   - Rotational latency: Wait for start of our sector to get under the head
     - Heads are on the right track, but not above the correct sector
   - Data read: Read until end of sector seen by head
   - Controller time: Controller checks the checksum
   - I/O bus time: Once the controller has the data, how long does it take to get to main memory
2. Disk takes several seconds to begin spinning, so assume it is already spinning
3. Accesses to the disk are happening one at a time
   - This induces a significant queuing delay while we wait for our request to be serviced

Disk Access Time Quiz

1. Consider a disk with the following parameters:
   - 1000 cylinders, 10 sectors/track
   - Head assembly at cylinder 0 initially
   - Head moves 10 microseconds/cylinder
   - Disk rotates 100 times/second
   - Controller, bus perfect, no previous request
2. What is the average time to read a randomly chosen byte?
   - 10 microseconds/cylinder * 1000/2 cylinders = 5000 microseconds (seek)
   - 10 milliseconds/rotation = 1 millsecond per track
   - On average, must rotate 5 tracks = 5 milliseconds (rotational)
   - 5000 microseconds + 5 millseconds = 10 milliseconds
   - Read sector = 1/10 * 10ms = 1 ms
   - Total = 5 + 5 + 1 = 11 ms

Trends for Magnetic Disks

1. Capacity: 2x per 1-2 years
2. Seek time: 5-10 ms with very slow improvement
    - Faster motor that can still move head assembly very precisely
    - Make platter smaller so we have less distance to traverse
        - 5 in -> 3.5 in -> 2.5 in
3. Rotation: 5000 rpm -> 10000 rpm -> 15000 rpm
    - Changes in material of platter
    - Makes more noise
    - Improves slowly
4. Controller and bus
    - Improves at a reasonable rate
5. Disk access time is dominated by seek time and, to a lesser extent, rotational latency
    - Not subject to something like Moore's law for processors

Optical Disks

1. Very similar to a hard disk in that it has a platter and bits are stored on the surface
    - Shoot a laser to the surface and the reflection tells us if it's a 0 or 1
2. Smudges and dust are less of a problem
    - Dirt can ruin a hard drive head, so they must be enclosed
    - CD or DVD are much more portable
3. Standardization helps portability
    - Limits rate of improvement
    - Technology improves -> Standards process -> Products
    - With hard drives, the whole enclosure needs to obey a standard, but what happens inside the enclosure can evolve more rapidly

Magnetic Tape

1. Typically used for backups (secondary storage)
    - First search the hard drive, then go to secondary storage
    - Large capacity, replaceable
2. Access is fundamentally sequential
    - Need to seek along the tape until we find the correct data
3. Slowly dying out -> Cost not dropping as rapidly as disks
    - Cheaper to use disks with all of the built-in electronics than it is to buy the tapes and readers
        - USB drives are an entire disk with a USB interface

Disks and Tape Quiz

1. Read 1GB file from start to end
    - Disk or tape
2. Read only first and last byte
    - Disk only
3. Make cat happy
    - Tape only

Using RAM for Storage

1. Hard disks don't benefit from Moore's law the same way processors or DRAM do
    - DRAM capacity increases at roughly the same rate as hard disks
    - DRAM is much faster (1 microsecond vs 10 milliseconds for hard drive)
2. Disk is about 100x cheaper / GB

3. DRAM has about 100,000x better latency
4. Solid-state disk (SSD)
    - Not a disk at all, just electronic storage
    - DRAM + battery: Very fast, reliable, but very expensive
        - Not good for archiving (battery dies)
    - Flash: Uses transistors to store data
        - Low power, don't need to move heads, very fast (slower than DRAM)
        - Smaller capacity than disks (GBs vs TBs)
        - Keeps data without power

Hybrid Magnetic Flash

1. Magnetic disk
    - Low $/GB
    - Huge capacity
    - Power hungry
    - Slow (mechanical movement)
    - Sensitive to impacts while spinning
2. Flash
    - Fast
    - Power efficient
    - No moving parts
3. Combine the benefits of each
    - Use flash as a cache for disk
    - Both can keep data without power, so data will persist during an outage

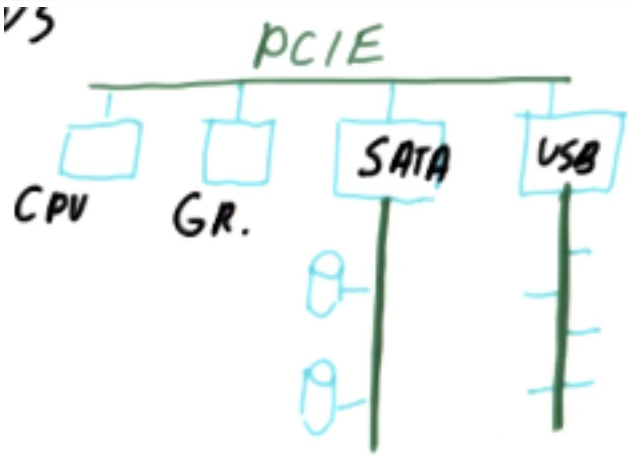
Flash vs Disk vs Both Quiz

1. Consider the following use of storage:
    - Play game for 2 hours (read 2GB, write another 10 MB)
    - Watch movie for 2 hours (read 1GB sequentially)
    - Repeat a total of four times
2. Disk reads 100 MB/s sequentially or 1 MB/s
3. Flash reads at 1GB/s
4. Total access time with disk
    - Game: 20 for read + 10 for write = 30 seconds
    - Movie: 10 seconds
    - Total = 4 * 40 = 160 seconds
5. Total access time with flash
    - Game: 2 for read + 0.01 for write
    - Movie: 1 second
    - Total = 4 * 3.01 = 12.04
6. Total access time with disk + 4GB flash cache
    - 40 seconds reading from disk + 3.01 seconds copying to flash
    - Subsequent reads are 3.01 seconds
    - Total = 40 + 4 * 3.01 = 52.05

Connection IO Devices

1. Connect I/O devices through a standardized bus
    - Needs to be standardized to be able to swap between manufacturers easily
        - Standards limit the rate of improvement
2. Typically have a hierarchy of buses
    - Mezzanine bus (PCI Express)

  - Pretty fast, short
  - Directly connect fast devices (graphics)
- SATA/SCSI controllers are connected through PCI Express
  - These controllers connect to a SATA bus
  - Slower and less change
  - Can update PCI Express without changing underlying hard drive
- USB hub connects to a separate USB bus



Bus Hierarchy

Conclusion

1. How do we handle the failure of storage devices gracefully?