The Bare Basics

Storing Data on Disks and Files

Chapter 9

Disks and Files

- DBMS stores information on ("hard") disks.
- This has major implications for DBMS design!
  - READ: transfer data from disk to main memory (RAM).
  - WRITE: transfer data from RAM to disk.
  - Both are high-cost operations, relative to in-memory operations, so must be planned carefully!

Why Not Store Everything in Main Memory?

- *Costs too much.*
  - Same amount of money will buy you say either 128MB of RAM or 20GB of disk.
- *Main memory is volatile.*
  - We want data to be saved between runs. (Obviously!)
- Typical storage hierarchies:
  - Main memory (RAM) for currently used data (primary storage) .
  - Disk for the main database (secondary storage).
  - Tapes for archiving older versions of data (tertiary storage).

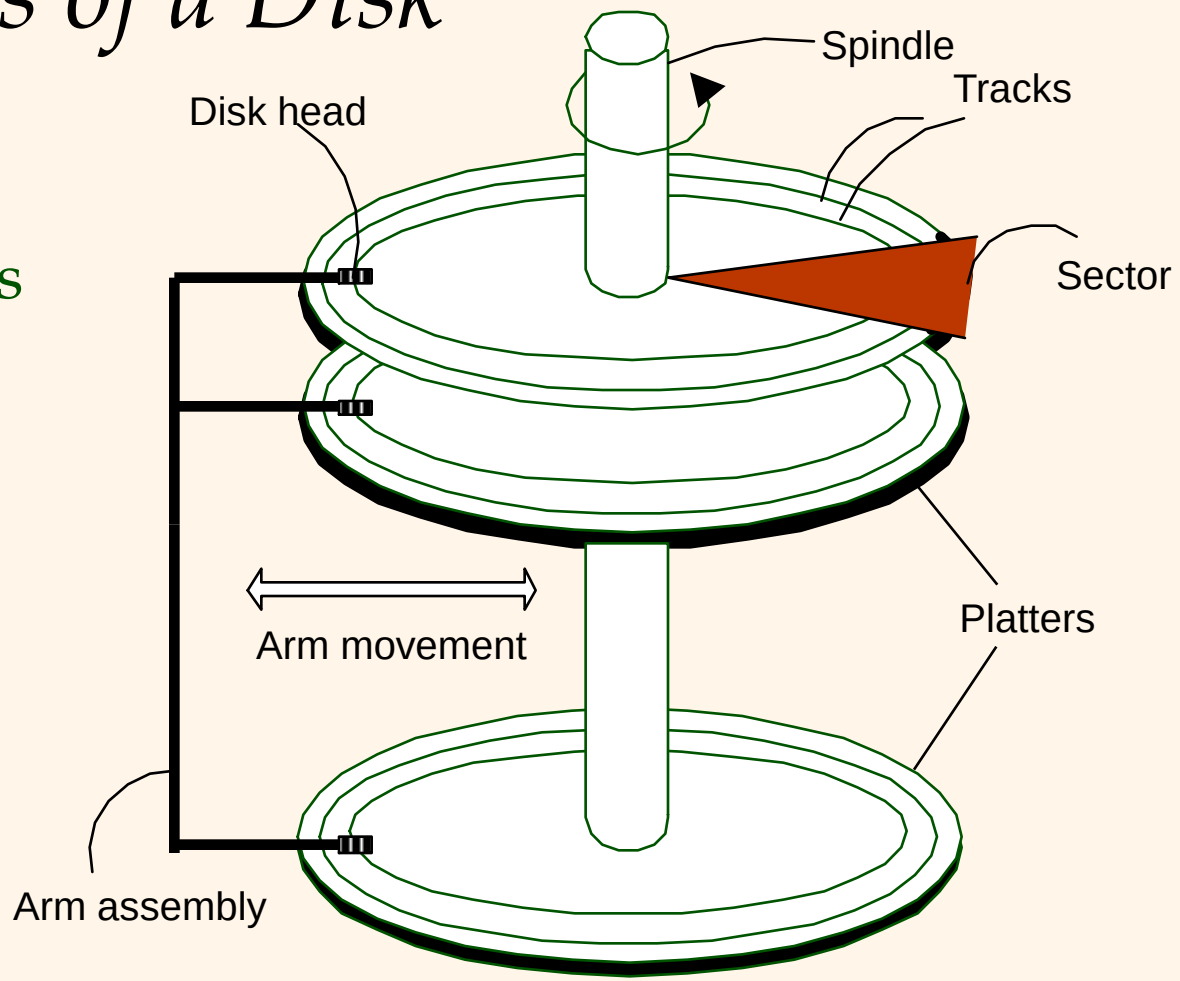
Disks

- Secondary storage device of choice.
- Main advantage over tapes:
  - *random access* vs. *sequential*.
- Data is stored and retrieved in units :
  - called *disk blocks* or *pages*.

- Unlike RAM, time to retrieve a disk page varies depending upon location on disk.
  - Therefore, relative placement of pages on disk has major impact on DBMS performance!

Components of a Disk

- The platters spin (say, 90 rps).
- The arm assembly is moved in or out to position a head on a desired track.
- Tracks under heads make a *cylinder* (imaginary!).
- Only one head reads/writes at any one time.



- *Block size* is a multiple of *sector size* (which is fixed).

Accessing a Disk Page

- Time to access (read/write) a disk block:
  - *seek time* (moving arms to position disk head on track)
  - *rotational delay* (waiting for block to rotate under head)
  - *transfer time* (actually moving data to/from disk surface)
- Seek time and rotational delay dominate.
  - Seek time varies from about 1 to 20msec
  - Rotational delay varies from 0 to 10msec
  - Transfer rate is about 1msec per 4KB page
- Lower I/O cost: reduce seek/rotation delays!

Arranging Pages on Disk

- `*Next*' block concept:
  - blocks on same track, followed by
  - blocks on same cylinder, followed by
  - blocks on adjacent cylinder
- Blocks in a file should be arranged sequentially on disk (by `next'), to minimize seek and rotational delay.
- For a sequential scan, *pre-fetching* several pages at a time is a big win!

Disk Space Management

- Lowest layer of DBMS software manages space on disk.

- Higher levels call upon this layer to:
  - allocate/de-allocate a page
  - read/write a page

- Higher levels don't need to know how this is done, or how free space is managed.