External Data Structures

External Data Structures

An external data structure is one which is stored on external or secondary memory (i.e. disks) rather than in internal memory (RAM). This means that the data structure can:

- Persist beyond the end of the program execution without the programmer having to explicitly save or load the structure to/from disk

- Grow to sizes larger than can fit in internal memory, often hugely larger, being limited only by the amount of secondary storage available.

- Be accessed and updated by multiple different programs at the same time so long as suitable coordination protocols are observed by the different programs.

Properties of Secondary Storage

- Disks store data in block of sizes configured by the operating system, usually 4KBytes but can be up to 64KBytes
- The disk can only transfer whole blocks at a time: to write a single byte to disk, the operating system would have to
    1. read the block, which would contain the byte, into memory
    2. replace the byte in the memory copy of the disk block
    3. write the memory copy of the block back out to the disk
- In spinning hard disks, the time to read (or write) a block includes
    1. time to move the disk head to the correct track (approx 6ms),
    2. time to wait for the block to spin around to the disk head (approx 4ms) and
    3. time for the disk to spin further until all the block has passed under the disk head.
    4. reading or writing consecutive blocks is much faster that reading a random sequence of disk blocks