Problem 1.

Consider a hard disk that uses Advanced Format with sector size of 4096 bytes (4KB), 1000 tracks per surface, 50 sectors per track, five (5) double-sided platters, and average seek time of 8 msec. Suppose that the disk platters rotate at 7200 rpm (revolutions per minute).

Suppose also that a block (page) of 4096 bytes is chosen. Now, consider a file with 150,000 records of 100 bytes each that is to be stored on such a disk and that no record is allowed to span two blocks. Also, no block can span two tracks.

1. How many records fit onto a block?
2. How many blocks are required to store the entire file? If the file is arranged "sequentially on the disk, how many cylinders are needed?
3. How many records of 100 bytes each can be stored using this disk?
4. What time is required to read a file containing 100,000 records of 100 bytes each sequentially? You can make an assumption about how long it takes moving the heads from one cylinder to the next here.
5. What is the time required to read a file containing 100,000 records of 100 bytes each in a random order? To read a record, the block containing the record has to be fetched from the disk. Assume that each block request incurs the average seek time and rotational delay. You need also to include the transfer time for each block.

Problem 2.

Consider a B+-tree of order two (d=2). Thus, the maximum number of pointers per node is 5 and the maximum number of entries is 4.

1. Show the results of entering one by one the keys that are three letter strings: (era, ban, bat, kin, day, log, rye, max, won, ace, ado, bug, cop, gas, let, fax) (in that order) to an initially empty B+-tree. Assume that you use lexicographic ordering to compare the strings. Show the state of the tree after every 4 insertions.
2. What is the utilization of the tree? The utilization of the tree is defined as the total number of entries in all the nodes of the tree (both leaf and non-leaf nodes) over the maximum number of entries that the same nodes can store.

Exercise 3

Answer the following questions for each of these scenarios, assuming that our most general external sorting algorithm is used:

1. (a) A file with 10,000 pages and three available buffer pages.
2. (b) A file with 20,000 pages and five available buffer pages.
3. (c) A file with 2,000,000 pages and 17 available buffer pages.

1. How many runs will you produce in the first pass?
2. How many passes will it take to sort the file completely?
3. What is the total I/O cost of sorting the file?
4. How many buffer pages do you need to sort the file completely in just two passes?