CS 351 DATA ORGANIZATION AND MANAGEMENT

Homework 1 (New version, Oct. 10, 2010)
Date Given : October 7, 2010
Date Due : October 18, 2010

Important Notes: 1. Please submit the Homework to Room EA 126 on the due date by 5:00 pm –on the same day you can also give it to me in the classroom- (no late submission will be accepted). 2. Answer the question in the order they are given using a standard size paper. 3. Handwritten submissions are accepted; however a word document is preferred and appreciated. 4. Staple all papers and write your name on them.

Q1. Prepare a table giving comparative information as to Capacities, Speeds of Access and Costs ($/Mbyte) for the storage devices in the hierarchy. Include Semiconductor RAM, Hard Discs, Floppy or Flexible Disks, CD-ROMs, DVDs, BluRay Optical Media and Magnetic Tapes/Cassettes.
Indicate the unit of measurement information for the characteristics under consideration.

We have the same question asked last year (the answers are available on our course web site, some numbers there may need some corrections, e.g., magnetic tape speed?), indicate the one with the most significant change in terms of cost and storage capacity. Please specify your resources (web pages, magazines, etc.).

Q2. Find two modern-day hard disk drives available in the market (indicate your resources) and for them a) compute the time it takes for one full revolution of a disk, b) compute the average rotational delay for these disks, c) specify seek time and d) data transfer speeds.

Q3. Consider sequential processing of physically adjacent n blocks on successive tracks of the same cylinder vs. random processing of the same number of blocks.(you can ignore the effect of first bbt and use ebt instead for sequential read and write operations.This is valid for Q4, Q5 and Q6 also)
Assume the disk parameters are: s= 16 msec, r= 8.3 msec, btt= 0.8 msec and ebt= 0.84 msec.
Compute the times Ts (=total time for sequential processing) and Tr (=total time for random processing) for b=100 and for b=1,000,000 blocks.
Compute the Tr/Ts for very large b (i.e. when b goes to infinity).

Q4. Consider two unsorted sequential files (piles) F1 and F2 with 100,000 records and record size of 800 bytes. Please also assume the IBM 3380 environment (block size = 2400, ebt= 0.84 msec, r= 8.3 msec, s= 16 msec, etc.). The records are the bibliographic information about the books of Bodrum and Marmaris public libraries. Assume that files do not contain duplicate records (i.e., a record can only appear only once in a file) and 70% of the records of these files are common.

a) Develop an algorithm to create a file which is the difference of these two file, i.e.,
FD (Difference)= F1 – F2 (FD contains all records of F1 that do not appear in F2).

A simple step by step English explanation is enough, i.e., you do not need to write it in pseudo code. However, if you prefer pseudo code it is also acceptable. In your algorithm assume that we do not have any additional memory other than the two buffers provided by the operating system. That is your algorithm will be similar to the least efficient algorithm that we analyzed for file intersection.

b) Develop a formula for the file processing time, do not include the (s+r) in the formula and in your calculations. Calculate the file processing time using this formula and the IBM 3380 environment.

c) Now find the number of (s+r) operations that you need to perform and calculate the total time requirements of seek and rotational latency times by using the IBM 3380 environment.

Q5. a) Answer the same questions as Q4 (sections a, b, c) but this time assume that 10 MB of main memory is available. In this 10 MB of memory you can only keep the records of F1 to search them in F2.

b) Consider the same question but this time assume that we have 20 MB of memory to keep the records of F1. Will there be a significant change in file processing time? Explain your answer.

Q6. Answer the same questions as Q5.a (it has three sections) but this time assume that you can use 10 MB of memory to keep the records of F1 and F2 half and half (i.e., allocate 5 MB to keep the records of F1 and allocate 5 MB to keep the records of F2).