

CS 351 DATA ORGANIZATION AND MANAGEMENT

Homework 1

Date Given : October 19, 2011

Date Due : October 31, 2011

Important Notes: 1. Please submit the Homework to Room EA 511 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. 5. I plan to assign the next homework before the due date of this assignment.

Q1. Secondary storage medium related: **a.** Find the technical specification (seek time, rotational latency time, disk capacity, track capacity, disk capacity) of a current hard disk drive designed for desktops. **b.** Explain if it supports the concept of cylinder. **c.** Compare it with a typical solid state disk in terms of speed and storage capacity. **d.** If you compare the characteristics of the hard disk you present in item **a** with IBM 3380: what do you see especially interesting in terms of speed and in terms of other aspects?

Please specify your resources (web pages etc.).

Q2. In the IBM 3380 environment consider sequential processing of bk number buckets vs. random processing of the same number of buckets. Each bucket contains 10 blocks.

Compute the times T_s (=total time for sequential processing) and T_r (=total time for random processing) for $bk=50$ and for $bk=1,000,000$ buckets.

Compute the T_r/T_s for very large bk (i.e. when bk goes to infinity).

In the following questions please state any assumption you make.

Q3. Consider two unsorted sequential files F_1 and F_2 in the IBM 3380 disk drive environment. These files contain 250,000 records and R (record size is 200 bytes). In both files 20% of the records are marked as deleted. Fifty percent of the active records of F_1 and F_2 are common. We want to create an intersection file. For this purpose we have 10 MB (10,000,000 bytes) of main memory allocated to file intersection. The file intersection program works as follows: we read 10MB of records of F_1 to main memory and search the active records in F_2 and if the record is also active in F_2 it is marked as common and such records are written to the intersection file F_{12} all together (i.e., after processing all records stored in that 10 MB area). This process is repeated until we are done. **a.** How many (s+r) operation is needed for F_1 , F_2 , and F_{12} ? **b.** How much time is needed to read the records of F_1 ? **c.** How much time is needed to read the records of F_2 ? **d.** How much time is needed to write the active records to F_{12} ? **e.** What happens if we use two disk drives. Please show your work.

Q4. Again consider the files specified in Q3. **a.** How much time is needed to reorganize F_1 and F_2 . **b.** Answer Q3 this time by assuming that we are using the reorganized versions of F_1 and F_2 . Please show your work.

Q5. Consider two files F_1 and F_2 . They both contain n number of records and B_{fr} is the blocking factor for both of them. M is the amount of bytes available in the memory (like 10 MB above). The p percent of the records are common in F_1 and F_2 . Develop an algorithm for creating a union file called F_{12} . **a.** Present your algorithm in pseudo code. **b.** Develop a model (equation) to predict the time needed for union. **c.** Can we increase the speed of file processing if we use two disk drives? Please explain.