

## Midterm 2002 - spring

### CSC228H University of Toronto

**Duration – 50 minutes**

**Aids Allowed: none. No calculators.**

Student Number: \_\_\_\_\_

Last Name: \_\_\_\_\_

First Name: \_\_\_\_\_

Instructor: \_\_\_\_\_

TA: \_\_\_\_\_

---

***Do not turn this page until you have received the signal to start***

---

#1: \_\_\_\_\_ / 12

#2: \_\_\_\_\_ / 12

#3: \_\_\_\_\_ / 16

#4: \_\_\_\_\_ / 8

#5: \_\_\_\_\_ / 10

Total: \_\_\_\_\_ / 58

Good Luck!

PLEASE HAND IN

**Question 1. [12 marks]**

Given a Disk with the following characteristics:

64 sectors / track

512 bytes / block

Rotational delay: 10 ms

Transfer rate: 2048 bytes / ms

Seek time: 20 ms

**Part (a)** [6 marks]: How many tracks would be needed to store a file with 4096000 bytes?

$$4096000 \text{ [bytes]} / 512 \text{ [bytes / block]} = 8000 \text{ blocks}$$

$$8000 \text{ [blocks]} / 64 \text{ [blocks / track]} = \mathbf{125 \text{ tracks}}$$

**Part (b)** [6 marks]: How long would we take to read the whole file in sequence, assuming no interruptions while reading a track?

$$125 \times (20 + 10) = 3750 \text{ ms (due to seek time and rotational delay)}$$

$$4096000 / 2048 = 2000 \text{ ms (due to transfer rate)}$$

$$t = 3750 + 2000 = \mathbf{5750 \text{ ms}}$$

**Question 2. [12 marks]**

Assuming we have a data file with 1000 records. Each record contains 4 fields with one unsigned integer (4-bytes) each.

What would the **minimum** file size (no compression) be if, on average:

**Part (a)** [ 3 marks ] : The values of the fields are 2 (decimal) digits long and you write the file in text mode, with a 1-byte field delimiter and a 1-byte record delimiter.

**On average**, 2 decimal digits  $\rightarrow$  2 bytes

$$( 4 \text{ [fields / record]} \times ( 2 \text{ [bytes / field]} + 1 \text{ [byte / field delimiter]} ) + 1 \text{ [byte / record delimiter]} ) \times 1000 \text{ [records]} =$$

**13000 bytes**

**Part (b)** [ 3 marks ] : The values of the fields are 2 (decimal) digits long and you write the file in binary mode, with fixed length fields.

Since the fields are fixed length, we have to attend all possibilities, not only the average case, therefore, **4 bytes**. We don't have delimiters in this case.

$$4 \text{ [fields / record]} \times 4 \text{ [bytes / field]} \times 1000 \text{ [records]} = \mathbf{16000 \text{ bytes}}$$

**Part (c)** [ 3 marks ] : The values of the fields are 2 (decimal) digits long and you write the file in text mode, with fixed length fields.

If you want to represent the 4 bytes in the field (  $FF\ FF\ FF\ FF = 2^{32} - 1 = 4,294,967,295$  ) we have to allocate space for 10 digits  $\rightarrow$  10bytes in text mode. We have again not to consider the average case, since the fields are fixed length.

$$4 \text{ [fields / record]} \times 10 \text{ [bytes / field]} \times 1000 \text{ [records]} = \mathbf{40000 \text{ bytes}}$$

**Part (d)** [ 3 marks ] : Why would it be difficult to deal with a binary file and field delimiters, in this example?

Because in a binary file, all characters (00 – FF) may occur in the data, becoming difficult to choose a character as a field delimiter.

**Question 3. [16 marks]**

Given a B-tree order 7, what is:

**Part (a)** [2 marks]: The minimum number of children that the root node can have?

2

**Part (b)** [2 marks]: The maximum number of children that the root node can have?

7

**Part (c)** [2 marks]: The minimum number of data values that the root node can have?

1

**Part (d)** [2 marks]: The maximum number of data values that the root node can have?

6

**Part (e)** [2 marks]: The minimum number of children that a node can have?

4

**Part (f)** [2 marks]: The maximum number of children that a node can have?

7

**Part (g)** [2 marks]: The minimum number of data values that a node can have?

3

**Part (h)** [2 marks]: The maximum number of data values a node can have?

6

**Question 4. [8 marks]**

When would you use the following file compression techniques (give ONE typical case for each):

Run-Length code:

- Whenever there is a significant amount of equal characters in sequence.

Variable-Length code:

- Whenever certain characters are more frequent than others.

**Question 5. [10 marks]**

**Part (a)** [5 marks]: Explain briefly the co-sequential processing.

- 1 sorted Master File and 1 Transaction File
- or two or more sorted input files producing one or more output files
- The Master File and the Transaction File are read in sequence AND in parallel
- The Transaction File specifies operations in the Master File like matching, merging and updating

**Part (b)** [5 marks]: What are the bigohs involved when the transaction file is sorted and non-sorted? (State the meaning of the variable(s) you use)

**Answer:**

Considering:  $M \rightarrow$  Number of entries in the Master File

$N \rightarrow$  Number of entries in the Transaction File

**Sorted:**  $O(M + N)$  [ Since I read in sequence in each file, never going back ]

**Non-sorted:**  $O(M \times N)$  [ Since for each element in the TF I have to seek for the correspondent element in the MF, always starting over from the beginning]

Obs.: Explanation not required, but the elements have to be identified.