CS 351 DATA ORGANIZATION AND MANAGEMENT

HW2
Date Given : October 7, 2009
Date Due : October 16, 2009

Important Notes: 1. Please submit the Homework to Room EA 231 on the due date by 5:00 pm (no late submission will be accepted). 2. Answer the following 5 question in the order they are given using a standard size paper. 3. Handwritten submissions are accepted, a word document is preferred. 4. Staple all papers and write your name on them.

1. Suppose you have two track-sized buffers for processing a file of b blocks where the effective block transfer time is ebt. (For the concept of track sized buckets please refere to Salzberg's book, p. 37.)
   - **a.** If it takes 0.5 times the read time to process one block of data, what is the total processing time for the whole file? Justify your answer.
   - **b.** If it takes 1.5 times the read time to process one block of data, what is the total processing time for the whole file? Justify your answer.

2. Start with 100,000 records of 400 bytes each in a heap (pile) file on disc. Delete 1 record for every 3 records added until the total number of records is 150,000. How long does it take to reorganize this file? How long does it take to find a record r (T_F) right before reorganization and how long right after?

3. Summary Report Production

Suppose you have a pile file of 50,000 records with the record format:

CITY	COUNTY	TOWN	VILLAGE	POPULATION

A *city* has several counties, a *county* has several towns, a *town* has several villages and each *village* has a population.

You are asked to print a report to list:

The total population of a *town* within a *county* (sum of Village populations)
The total population of a *county* within a *city* and
The total population of a *city* as a listing:

City	**County**	**Town**	**Total Population**
City 1	County 1	Town 1	P_{111}
...			
City 1	County 1	Town k	P_{11k}
		County 1 Population	$\mathbf{P_{co1}}$
City 1	County 2	Town 1	P_{111}
...			
City 1	County 2	Town 1	P_{121}
		City 1 Population	$\mathbf{P_{city}}$

...

City n	County m	Town p	P_{nmp}
		City n Population	$\mathbf{P_{city}}$
TOTAL POPULATION			$\mathbf{P_{total}}$

Describe an efficient method to pocess the given pile file to produce the above report.

4. Given the sequence of keys:

320, 315, 330, 275, 310, 305, 250, 325, 290, 100, 180, 150, 170, 140, 210, 175, 205

a. Form a minimum heap (the root must contain the minimum element) out of the above key sequence by inserting the keys in the order given from left to right

b. Using Replacement Selection with the following additional sequence of unordered keys from left to right generate the first two unit strings (initial strings, segments).

311, 276, 110, 176, 261, 321, 212, ...

5. Suppose that we have two files of magazin subscriptions. One is for M1 magazin and has 10 million records of 100 bytes each. The other is for M2 that has 2 million records of 100 bytes each. 1 million of the records are in both files. Both files are pile files. Assume the parameters of IBM 3380 disk drive and 10MB of memory space available.

a. Devise an efficient method (your own method) to create a file that contains the common records.

b. How long does your method in (a) take?

c. Devise an efficient method to create a combined mailing list with no duplicates.

d How long does your method in (c) take?