CS 351 DATA ORGANIZATION AND MANAGEMENT

HW3
Date Given : October 29, 2009
Date Due : November 5, 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 questions 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. 5. When needed assume/use the parameters/values of IBM 3380.

1. Perform replacement selection sort on the following successively entered records with keys 80, 90, 39, 49, 68, 65 and then continue with the key values 100, 15, 120, 25, 35, 130. Assume 6-record memory capacity, i.e., the maximum size of the priority queue is six nodes. Show enough detail of your steps so that it will be easy to follow.

2. Consider an unordered file with 10,000,000 records of 400 bytes each with blocking factor of 6. Calculate the time needed to sort this file using a 10 MB of computer memory. Just give the time needed to perform the heapsort part without replacement selection (merging part is not included).

3. For the file of question 2 find the complete sort time (heap sort + merging) for the following cases: a) 2–way merge, b) 4-way merge, c) P-way merge where P is equal to the total number of available sorted segments (additional info another name for a sorted segment is "run").

4. Using a bucket size of one and the hash function HF(key)= mod(key, 3) and bucket chaining enter into an empty static hash table (file) the following values: (30, 10, 32, 27, 5, 18). Show the file contents. a) How many disk accesses are required in order to retrieve the record 18. b) What is the average number of accesses for a successful search? c) What is the number of accesses for an unsuccessful search?

5. Consider a Linear Hashing environment with Bkfr= 3 and Lf (desirable load factor)= 2/3. Update the file structure when the "current load factor of the file becomes 2/3 and after adding 2 records." Start with h= 1, bv= 0, and use ~~mod(key, 4)~~ mod(key, 31) to obtain the pseudo key values. Show how it works with the following key values 10, 20, 3, 7, 13, 14, 17, 21, 25, 16, 22, and 30.

   Comment: h is the level value and as the file space grows higher values of h are used.
   See the textbook for this notation. This notation is equivalent to using k and k+1 bits in addressing as used in the Betty Salzberg book.

6. In a Linear Hashing file environment bv= 5 and h= 5. Give the values for the following items.
   Number of blocks hashed at level h
   Number of blocks hashed at level h+1
   The binary address of the last bucket of the file
   The binary address of the last bucket of the file hashed at level 5

   Answer the same questions for bv= 0 and h= 5.
   Comment: See the comment in Q5.

7. Insert the following key values into an extendible hash file: (27, 18, 29, 28, 42, 13, 16), using a bucket size of 3. Hash function is defined as mod(key, 11).

8. Redo the above question but enter the keys in the following order: (27, 42, 18, 29, 28, 13, 16). Compare the results of question 7 and 8 in terms of directory size and number of data pages.