

## CS 351 DATA ORGANIZATION AND MANAGEMENT

### HW4

Date: December 24, 2011

Date Due: January 4, 2012

**Important Notes:** 1. Please submit the Homework to Room EA 511 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 and section on them. 5. When needed assume/use the parameters/values of IBM 3380.

1. In an extendible file environment pseudo keys are obtained by  $\text{mod}(\text{key}, 5693)$ ,  $R$  (record size) = 400 bytes. In the implementation environment assume that disk block address representation (pointers) requires 4 bytes. a) What is the maximum memory requirement of the file directory? b) Is it possible to have overflow blocks if so how? c) What is the maximum file size in terms of number of records?
2. In an extendible hashing file with  $d$  (directory level) equal to 4 what is the minimum and maximum number of pages with  $p$  (page hashing level) equal to 1, 2, 3, 4, and 5. Explain briefly.
3. Draw the B tree of degree (order) 1 by entering 12, 14, 7, 9, 16, 6, 4, 11, 10.
4. Draw the B+ tree of order 1 (both for index set and sequence set nodes) which results from entering the successive input 14, 7, 9, 5, 15, 3, 1, 6, 12, 4, 11, 10, 16.
5. Consider a file with 200,000 records of bank accounts.  $R$  (record size) is 400 bytes. Different from our usual assumption for B+ trees the data nodes are not connected to each other. Other than that assume that it is a typical B+ tree environment. a) Calculate the time needed to find the average account balance, and also calculate  $T_F$ , and  $T_N$ . b) Now assume that we have a sequential file for the same set of records. Find the time needed to find the average bank account, and also calculate  $T_F$ , and  $T_N$ .
6. Suppose the data nodes of a B+ tree are 2400 bytes long, record size is 400 bytes; assume that the degree of index nodes is 80. What size files (in terms of number of records) will have a two-disk-access retrieval with 10 Megabyte of memory? In your calculation use the bottom-up approximation introduced by Salzberg. In this approximation how many layers are assumed for the index file (including the final layer above the data nodes)? Draw a simple figure to explain.