**CS 351 DATA ORGANIZATION AND MANAGEMENT**

**HW3**

**Date: December 7, 2010**

**Date Due: December 15, 2010**

**Important Notes: 1. Please submit the Homework to Room EA 126 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.** Draw the B tree of degree (order) 1 by entering 16, 6, 4, 11, 10, 12, 14, 7, 9. Use the following property of B trees while growing the tree: the root has at least two children if it is not a leaf node[1].

**2.** In this question first insert the records with the keys 6, 8, 30, 40, 50, 61, and 70 into a B tree of order (degree) one. Now, insert the above keys in a different order to obtain the minimum number of nodes. If your first attempt gives the minimum number of nodes, please explain why.

**3.** Draw the B+ tree of order 1 (both for index set and sequence set nodes) which results from entering the successive input 6, 12, 4, 11, 10, 16, 14, 7, 9, 5, 15, 3, 1.

**4.** What is the minimum depth of a $B^+$ tree containing 300 records with an index of degree 5 and data nodes holding up to 10 records? What is the maximum depth?
Hint. For the minimum height all nodes must be 100% full, for the maximum height all nodes must be only 50% full.

**5.** Consider a range query based on the key field in a B+ tree environment. The degree of the data nodes is 50 and the degree of index nodes is 200. The number of records to be accessed is 102. What is the minimum and maximum number of data nodes to be accessed? Explain your answer. (Assume that data nodes are connected to each other and data nodes are 100% full.)

**6.** Questions on file structures.
- **a.** What are the advantages of B+ tree structure over ISAM?
- **b.** Can we increase the index area in ISAM files? Does its indexing area policy cause any problems?
- **c.** What are the advantages of B+ tree structure over Linear Hashing?
- **d.** What are the advantages of sequential files over B+ tree structure?
- **e.** Compare range query processing in the B and B+ tree environments in terms of efficiency.

**7.** Suppose the leaf nodes of a B+ tree are 2400 bytes long, record size is 200 bytes; the internal nodes have an average fanout of 50. What size files (in terms of number of records) will have a two-disk-access retrieval with only 5 Megabyte of memory? In your calculation use the approximation mem= $bk/fo^2$ (see 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.

**8.** Consider a sorted sequential file with 1,000,000 records. Assume that records are stored in buckets and the bucket size is 100 records, and assume that bucket I/O time (including s, r and dtt) is 10 ms.
Hint. We did not discuss sorted sequential files, but this question can be answered without reading that section: just consider how the size of the search space decreases as we perform the binary search. And also remember that each block contains 100 records and I/O is done in terms of physical reads.

**a.** How many disk accesses are needed to find a record in this file? Calculate the expected value for $T_F$. Explain briefly. (Note that any desired bucket can be accessed by calculating its displacement with respect to the beginning of the file.)

**b.** Now assume that we have an index structure for our sequential file which is 1% of the original file and also assume that we can keep this index in the main memory. 1% implies that for each (1,000,000 / 100=) 10,000 records we have an entry in this index (the first index entry indicates the value stored in the $10{,}000^{th}$ record, the $2^{nd}$ entry indicates the value stored in the $20{,}000^{th}$ record etc., it is similar to ISAM track index). How many disk accesses are needed to find a record in this case? Calculate the expected value for $T_F$. Explain briefly.

---

[1] We discussed a similar case in one of the sections and overlooked this property.