

Written Assignment #1
CIS5930: Advanced Topics in Data Management
Fall 2008

Assigned: Tuesday, Sep 09, 2008; Due: In Class on Tuesday, Sep 23, 2008.

Problem 1. [20 points]

Given a B+ tree with fanout f_k , in each node of the tree, we embed a smaller B+ tree with fanout f_e . Notice that we still would like to ensure that each node fits into one page (holding up-to B items). For a data set with N records, answering the followings: what is the constraint that f_k and f_e must satisfy? what is the height of the B+ tree (the outside tree)? What is the point search query cost in terms of I/Os? What is the point search query cost in terms of number of comparison performed?

Problem 2. [30pts]

Given two files of size N_1 and N_2 (say, containing integer numbers), a main memory of size M and a page size of B .

1. (10 pts) Assuming two files are not sorted, what is the cost (in terms of IOs) of finding out the k th element in the union of file one and file two?
2. (10 pts) Assuming two files are not sorted, what is the cost (in terms of IOs) of finding out the top- k th elements in the union of file one and file two?
3. (10 pts) What are the corresponding cost for the above questions when file one and file two are individually sorted at the beginning?

Problem 3. [50pts]

Implement the KD-tree based on the code template provided. You must follow the interface as they are in the template (feel free to add methods if necessary). You should test your implementation with the data set provided.

For range queries, the code should print out both the X and Y coordinates of each point inside the query range, as well as the total number of points enclosed in the query range.

Bonus Points [25pts] Make your KD tree serializable.