Princeton University

COS425 - Database Systems
William O'Connell - Spring 1996

Homework #4 - Indexing and Hashing

Graded out of 100 points; Due: in class, Tuesday, April 16th 1996.

1. [15 pts] Critique each of the following access methods {**ISAM, B+-tree, Static Hashing, Linear Hashing,** and **Extendible Hashing**}. The topics that should be included in the critique are:

- Point (exact match) queries
- Range Queries
- Flexibility in dynamic dataset growth.
- Effective use of Storage Space

2. [20 pts] Construct a B+-tree for the following set of key values

{2, 3, 5, 11, 17, 19, 23, 29, 31}

under the assumption that the number of search-key values that fit in one node is:

a. 3
b. 5
c. 7

3. [15 pts] For each B+-tree of problem 2, show the form of the tree after each of the following series of operations:

- insert 9
- insert 10
- insert 8
- delete 23
- delete 19

4. [15 pts] Suppose that we are using extendible hashing on a file containing records with the following search-key values:

2, 3, 5, 7, 11, 17, 19, 23, 29, 31

Show the extendible hash structure for this file if the hash function is $h(x) = x \bmod 8$ and buckets can hold three records.

5. [15 pts] Show how the extendible hash structure of problem 4 changes as the result of each of the following steps:

- delete 11
- delete 31
- insert 1
- insert 15

6. [20 pts] (a) Does data clustering improve performance, describe your answer. (b) How is data clustering best dealt with small objects? With large objects? (c) How does buffering assist clustering and why do buffer management in the DBMS versus OS?