B461-Fall 2009 Assignment 5 Indexing and Query Processing

1 Assignment 5-Part 1. Due: November 17

- (a) Specify the main differences between tree and hashing data structures to index a file of records.
 - (b) Specify an application wherein a tree index is superior over a hash index, and vice versa.
- 2. Assume that we have a primary index which is maintained. as a B⁺-tree.
 - (a) Argue why search time to a record is given as $O(log_d(n))$ wherein d is the order of the B⁺-tree and n is the number of records indexed.
 - (b) Argue why the insert time of a records is also $O(log_p(n))$.
- 3. Consider the following parameters: block-size = 4096 bytes, block access time = 20 ms (micro seconds), key-size = 9 bytes, block address-size = 8 bytes, record size = 100 bytes. Assume that there is a B^+ -tree (adhering to the parameters just specified) that holds 10 million records and these records are stored in the leaf nodes of the tree.
 - (a) Specify (in ms) the maximum time to determine whether a record with key k is in the B⁺-tree. **Hint**: begin by determining the order d of the B⁺-tree. (Show all the intermediate computations leading to your answer.)
 - (b) Assuming that 10% of the records in the B⁺-tree are between keys k_1 and k_2 . Specify the maximum time to display the records in the B⁺-tree in the range $[k_1, k_2]$. (Show all the intermediate computations leading to your answer.)
- Consider the following B⁺-tree of order 1 that holds records with keys 1, 5, 8, and 10.



- (a) Show the contents of your B^+ -tree after inserting records with keys 6, 9, 12, and 3, in that order.
- (b) Starting from your answer in question 4a the Show the contents of your B^+ -tree after deleting records with keys 1, 12, 3, 9, 6 in that order.
- 5. Extendible hashing
 - (a) Describe how overflow is handled in extendible hashing.
 - (b) Describe how underflow can be handled in extendible hashing.
 - (c) Consider an extendible hashing data structure wherein i) the initial global depth is set at 1 and ii) all directory pointers point to the same empty block which has local depth 0. So the hashing structure looks like this:

global-dept	h 1	local-depth 0
0	>	
1	>	

Assume that a block can hold at most two records. Show the state of the hash data structure after each of the following insert sequences:

- i. records with keys 1 and 5.
- ii. records with keys 2 and 3.
- iii. records with keys 4 and 7.

2 Assignment 5-Part 2. Due: November 19

- 1. Assume that NY is a file which contains records of persons who lived in New York during the last five years. A record in NY is identified by its social security number field. Let SF be a similar file for the city San Fransisco. A person can move from New York to San Fransisco, or vice versa. However, to simplify matters, we will assume that a person can move a most once.
 - (a) Specify a file organization and an algorithm that would enable finding efficiently the persons that have lived in New York and San Fransisco during the last five years.
 - (b) Specify the time complexity of your algorithm.
- 2. Consider a database wherein we maintain 3 data files:
 - (a) a file STUDENT which keeps student information (student-id, name, rank, major; student-id is a key),
 - (b) a file COURSE which keeps course information (course number, name, department; course-number is a key), and
 - (c) an ENROLLMENT file which keeps records of the form (student-id, course-number).

Specify a file organization for each of these files which would enable efficient processing of the following queries. Argue why your organization provides efficient processing for the following types of queries:

- (a) Retrieve the information associated with an individual student or an individual course.
- (b) Retrieve the courses offered by a department.
- (c) Retrieve the course numbers of courses taken by a given student.
- (d) Given two different students. Retrieve the course numbers of courses taken by both students.
- (e) Given a rank and a major, retrieve the students with that rank and major.
- (f) Retrieve the majors of students taking a given course.