National University of Computer and Emerging Sciences, Lahore
Spring Semester 2014

Course: CS502- ADVANCED DATABASES Time Allowed: *90 min.*
Max Points: *30*

Midterm 2

Section: _____ Name: _____ Roll No: _____

Q 1 *(6 points)***:**

a) What is the difference between a file organization and an access method?

b) List the techniques for allowing a hash file to expand and shrink dynamically.

c) What are the primary file organizations?

d) What is the difference between the directories of extendible and dynamic hashing?

Q 2 *(8 points)***:**
Consider a file system on a disk with block size *B=1000 bytes*. A block pointer is *P=6* bytes long, and a record pointer is P_R *=7* bytes long. A file has *r=10,000,000* STUDENT records of fixed-length (un-spanned). Each record has the following fields: *ROLLNO (10 bytes), NAME (25 bytes), DEPTNO (10 bytes), ADDRESS (30 bytes), PHONE (9 bytes), BIRTHDATE (8 bytes), SEX (1 byte), PROGRAMCODE (2 bytes), CGPA (4 bytes, real number)*. An additional byte is used as a deletion marker. No of departments = 200 and students per department = 50,000. RollNo is a primary key column.

Estimate the number of block fetches needed to compute the following queries:

a) SELECT * FROM student WHERE rollno=1234567; (Assume file is not ordered.)
b) SELECT * FROM student WHERE deptno=100; (Assume file is not ordered.)
c) SELECT * FROM student WHERE rollno=1234567; (Assume file is ordered on RollNo.)
d) SELECT * FROM student WHERE deptno=100; (Assume file is ordered on DeptNo.)

Q 3 *(10 points)***:**

Consider a file system on a disk with block size *B=1000 bytes*. A block pointer is *P=6* bytes long, and a record pointer is P_R *=7* bytes long. A file has *r=10,000,000* STUDENT records of fixed-length (un-spanned). Each record has the following fields: *ROLLNO (10 bytes), NAME (25 bytes), DEPTCODE (10 bytes), ADDRESS (30 bytes), PHONE (9 bytes), BIRTHDATE (8 bytes), SEX (1 byte), PROGRAMCODE (2 bytes), CGPA (4 bytes, real number)*. An additional byte is used as a deletion marker.

Suppose the file is not ordered by the key field RollNo and we want to construct a B^+-tree access structure (index) on RollNo. Calculate (rounded up for convenience using FLOOR function in part i, ii, iii):
(i) the orders *p* and p_{leaf} of the B^+-tree; (ii) the number of leaf-level blocks needed if blocks are approximately 69% full; (iii) the number of levels needed if internal nodes are also 69% full; (iv) the total number of blocks required by the B^+-tree; and (v) the number of block accesses needed to search for and retrieve a record from the file--given its RollNo value--using the B^+-tree.

Q 4 *(6 points)***:**

Consider a file system on a disk with block size *B=256 bytes*. A block pointer is *P=10* bytes long, and a record pointer is P_R *=14* bytes long. A record (R) of this file consists of 120 bytes where 50 bytes are for the search field. Show file layout for the index file and data file both when 15 data records having the following search values are inserted:
20, 85, 40, 85, 10, 15, 20, 85, 15, 20, 30, 40, 85, 15, 85 in each of the following cases:

a) Clustering Index (Single-level)
b) Secondary Index (Multi-Level)

You can assume that clustering index uses block anchors (every new value of search field starts at the beginning of a new block) and the secondary index uses neither the variable length record format nor the level of indirection.