Department of Computer Science,
New Mexico State University
Qualifying Exam for Databases
Fall, 2013

Open book; Open notes.   **Your code:** _____
Good luck!   **(2hrs, 100 points)**

Answer Questions 1–4 by using the following relational database schema, where the underlined attribute(s) form(s) the primary key of the corresponding schema.

- Student(<u>sid</u>: char(10), name: varchar(64), email: varchar(32), deptname: char(10))

- Take(<u>sid</u>: char(10), <u>cid</u>: char(5), score: float)
  Foreign key: *sid* references *Student (sid)*
  Foreign key: *cid* references *Course (cid)*

- Course(<u>cid</u>: char(5), credit: integer, title: varchar(128), deptname: char(10))

- Prereq (<u>course_id</u>: char(5), <u>prereq_id</u>: char(5))
  Foreign key *course_id* references *Course (cid)*
  Foreign key *prereq_id* references *Course (cid)*

1. (20pts) Write a relational algebra expression (or a set of relational algebra expressions) to answer each of the following queries.

   A. (10pts) Find all the non-CS students who have taken at least one Computer Science course. Show the ids and titles of such 'CS' courses together with the student names and their corresponding department names. (Note that Computer Science courses refer to the courses that are offered by the 'CS' department.)

   B. (10pts) For all the students in the Computer Science department (i.e., deptname='CS'), find the students whose average score is the highest. Show such students' ids, names, and their average scores. (Assume that a student is only allowed to take a course at most once.)

2. (20pts) Write an SQL statement to answer each of the following queries.

    A. (10pts) For each student in the Computer Science department, calculate the total number of credits that they have taken. Show such students' ids, names, and total number of credits taken. (Assume that a student is only allowed to take a course at most once.)

    B. (10pts) For all the students who have taken 'CS482', show all the other Computer Science courses that they have taken. Show the the students' id, names, as well as the ids and titles for the courses that they have taken.

3. (10pts) Suppose that the following function dependencies hold on the *Course* relation.

- $title \rightarrow deptname$
- $cid \rightarrow title$

    A. (5pts) What is/are the super key(s) for this relation? Please justify your answer.

    B. (5pts) Is the relation *Course* in BCNF? Please justify your answer.

4. (30pts) Besides the schema information, you are also given the following information about the buffer, disk, and the relations:

- No record spans multiple disk pages.
- The *integer* type uses 8 bytes, the *float* type uses 16 bytes, and an index entry pointer uses 16 bytes.
- Each buffer page size is 1K (i.e., 1024) bytes.
- Each disk page size is 1K (i.e., 1024) bytes.
- 5 buffer pages are available.
- The *Student* relation contains 10,000 tuples.
- The *Take* relation has a clustered B^+ index on its primary key and the height of this B^+ tree is 3.
- The *Student* relation has a clustered B^+ index on its primary key.

Note: If the provided information is not sufficient to answer the question and you need to make assumptions, please write down these assumptions clearly.

A. (10pts) What is the possible height of the B^+-tree for the *Student* relation? Please justify your answer.

B. (10pts) What is the possible number of tuples in the relation *Take*? Please justify your answer.

C. (10pts) Given a relational algebra expression "*Student* $\bowtie$ *Take*", please estimate (1) the number of results generated, and (2) the I/O cost if you apply index nested loop join algorithm.

5. (20pts) Given the log file as shown in Table 1 and assume that the buffer management system flushed out page P_3 at a time point which is after LSN35 and before LSN40.

LSN	Log
00	begin_checkpoint
05	end_checkpoint
10	update T1 write P3
15	update T2 write P2
20	update T1 write P2
25	T1 abort
30	update T2 write P3
35	CLR: Undo T1 LSN 20
40	T2 commit
45	Update T3 write P2
50	T2 end
55	CRASH, RESTART

Table 1: Log records

Answer the following questions to perform recovery when DBMS restarts:

A. (10pts) During the REDO phase, list the LSNs that need to be redone and explain the reasons.

B. (10pts) After finishing the UNDO phase, show the newly added log records and denote clearly the *undonextLSN* and *ToUndo*.

— End —