National Exams May 2011

98-Comp-B3, Data Bases & File Systems

3 hours duration

Notes:

1. If doubt exists as to the interpretation of any question, the candidate is urged to submit with the answer paper a clear statement of any assumption made.

2. This is a Closed Book exam. Candidates may use a Casio or Sharp approved calculator.

3. Answer **five** questions as follow:
   a. **One** question from **questions 1 and 2** (only one question will be marked)

   b. **One** question from **questions 3 and 4** (only one question will be marked)

   c. **Three** questions from **questions 5, 6, 7, and 8** (only three questions will be marked)

4. All questions are of equal value. The marking scheme is as follows:

   Question 1: (a) 3 marks; (b) 3 marks; (c) 14 marks
   Question 2: (a) 3 marks; (b) 3 marks; (c) 14 marks
   Question 3: 20 marks
   Question 4: 20 marks
   Question 5: (a) 3 marks; (b) 6 marks; (c) 5 marks; (d) 6 marks
   Question 6: (a) 3 marks; (b) 5 marks; (c) 7 marks; (d) 5 marks
   Question 7: (a) 3 marks; (b) 3 marks; (c) 14 marks
   Question 8: (a) 5 marks; (b) 3 marks; (c) 2 marks; (d) 10 marks

5. All answers should be clear, legible and brief.

Question 1

a. Explain why a file can have only one clustered index?

b. Explain why a secondary, unclustered index must be dense?

c. Starting with an empty B+ tree with up to two keys per node; show how the tree grows when the following keys are inserted one after another:

18, 10, 7, 14, 8, 9, 21

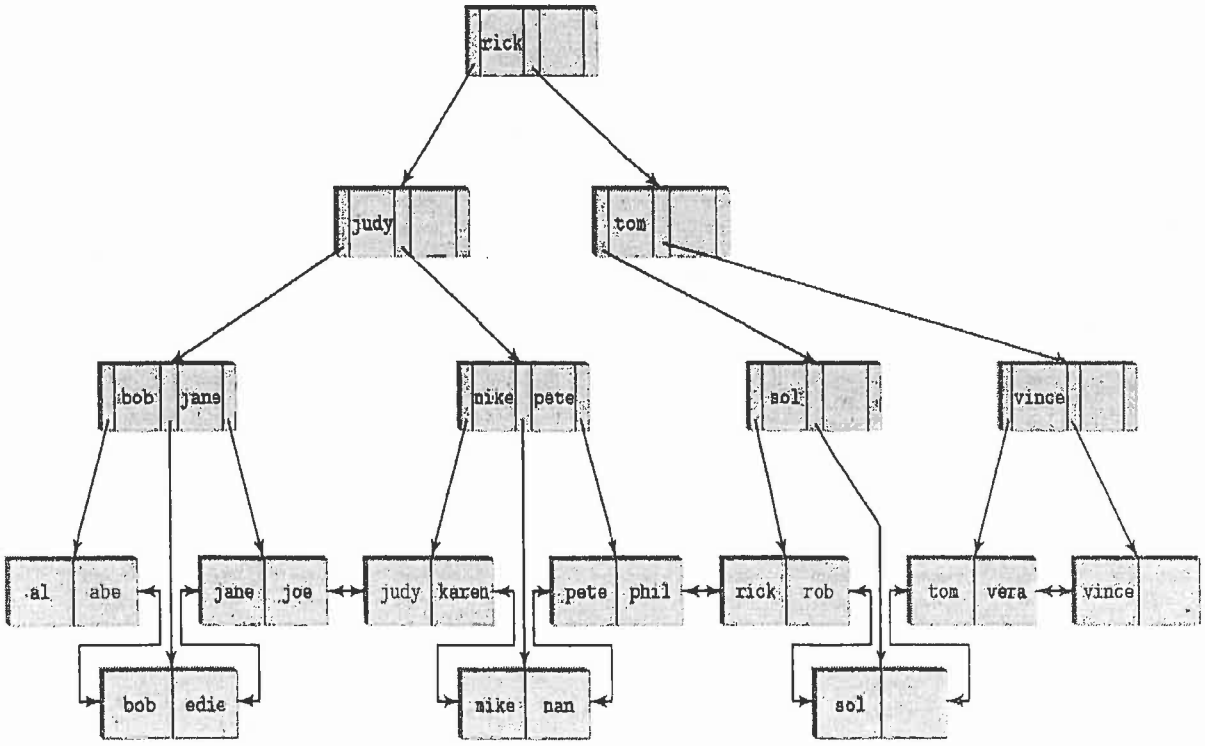
Note that question 1(c) demand that you show a B+ tree for each insertion.

Question 2

a. Explain the difference between an equality search and a range search.

b. Does the final structure of a B+ tree depend on the order in which the items are added to it? Explain your answer.

c. Draw a B+ tree that results from inserting alice, betty, carol, debbie, edith, and zelda into the index of the B+ tree below



Note: For this question you may just show the part of the diagram that changed as a result of the insertions. In addition, only one B+ tree is needed (i.e. the final B+ tree after insertions)

Question 3

Downtown Records has decided to store information about musicians who perform on its albums (as well as other company data) in a database. The company has wisely chosen to hire you as a database designer (at your usual consulting fee of $3000/day). The following information describes the situation that the Downtown database must model.

- Each musician that records at Downtown has an SIN, a name, an address, and a phone number. Poorly paid musicians often share the same address, and no address has more than one phone.
- Each instrument used in songs recorded at Downtown has a unique identification number, a name (e.g., guitar, synthesizer, flute) and a musical key (e.g., C, B-flat, E-flat).
- Each album recorded on the Downtown label has a unique identification number, a title, a copyright date, a format (e.g., CD or MC), and an album identifier.
- Each song recorded at Downtown has a title and an author.
- Each musician may play several instruments, and a given instrument may be played by several musicians.
- Each album has a number of songs on it, but no song may appear on more than one album.
- Each song is performed by one or more musicians, and a musician may perform a number of songs.
- Each album has exactly one musician who acts as its producer. A musician may produce several albums, of course.

Design a conceptual schema for Downtown using an ER diagram. Be sure to indicate all key and cardinality constraints and any assumptions you make. Identify any constraints you are unable to capture in the ER diagram and briefly explain why you could not express them.

Question 4

Computer Engineering Department frequent fliers have been complaining to Toronto County Airport officials about the poor organization at the airport. As a result, the officials decided that all information related to the airport should be organized using a DBMS, and you have been hired to design the database. Your first task is to organize the information about all the airplanes stationed and maintained at the airport. The relevant information is as follows:

- Every airplane has a registration number, and each airplane is of a specific model.
- The airport accommodates a number of airplane models, and each model is identified by a model number (e.g., DC-10) and has a capacity and a weight.
- A number of technicians work at the airport. You need to store the name, SIN, address, phone number, and salary of each technician.
- Each technician is an expert on one or more plane model(s), and his or her expertise may overlap with that of other technicians. This information about technicians must also be recorded.
- Traffic controllers must have an annual medical examination. For each traffic controller, you must store the date of the most recent exam.

- All airport employees (including technicians) belong to a union. You must store the union membership number of each employee. You can assume that each employee is uniquely identified by a social security number (i.e. SIN).
- The airport has a number of tests that are used periodically to ensure that airplanes are still airworthy. Each test has a Federal Aviation Administration (FAA) test number, a name, and a maximum possible score.
- The FAA requires the airport to keep track of each time a given airplane is tested by a given technician using a given test. For each testing event, the information needed is the date, the number of hours the technician spent doing the test, and the score the airplane received on the test.

Draw an ER diagram for the airport database. Be sure to indicate the various attributes of each entity and relationship set; also specify the key and participation constraints for each relationship set.

Question 5

Consider the following schema:

Suppliers(*sid:* **integer**, *sname:* **string**, *address:* **string**)
Parts(*pid:* **integer**, *pname:* **string**, *color:* **string**)
Catalog(*sid:* **integer**, *pid:* **integer**, *cost:* **real**)

The key attributes are underlined. The Suppliers relation defines supplier id, name and address. The Parts relation gives part id, name, and color. The Catalog relation lists the prices charges for parts by Suppliers. Now, write the following queries in **SQL**:

a. Find the *pnames* of parts for which there is some supplier.
b. Find names of suppliers who supply every red part.
c. Find the *sids* of suppliers who charge more for some part than the average cost of that part (averaged over all the suppliers who supply that part).
d. For every supplier that supplies a green part and a red part, print the name and prices of the most expensive part that s/he supplies.

Question 6

Consider the following relations containing airline flight information:

Flights(*flno:* **integer**, *from:* **string**, *to:* **string**, *distance:* **integer**, *departs:* **time**, *arrive:* **time**)
Aircraft(*aid:* **integer**, *aname:* **string**, *cruisingrange:* **integer**)
Certified(*eid:* **integer**, *aid:* **integer**)
Employees(*eid:* **integer**, *ename:* **string**, *salary:* **integer**)

The key attributes are underlined. The Flights relation lists flight number, origin, destination, distance, and departure and arrival times. The Aircraft relation shows aircraft id, name and cruising range. The Certified relation gives employee id, and aircraft id. The Employees

relation describes pilots (and other kinds of employees); every pilot is certified for some aircraft (otherwise he or she would not qualify as a pilot). Note that only pilots are certified to fly.

Now, write the following queries in **relational algebra**:

a. Find the *eids* of pilots certified for some Bombardier aircraft.
b. Find the names of pilots certified for some Airbus aircraft.
c. Find the *aids* of all aircraft that can be used on non-stop flights from Ottawa to Boston.
d. Identify the flights that can be piloted by every pilot whose salary is more than $160,000.00 Canadian.

Question 7

a. Define the term functional dependency.
b. Why is it that some functional dependencies are called trivial?
c. Consider a database schema with attributes A, B, C, D, and E and the functional dependencies B → E, E → A, A → D, and D → E.
    i. Prove that the decomposition of this schema into AB, BCD, and ADE is lossless.
    ii. Is the decomposition in (i) above dependency preserving?

Question 8

Consider the following two transactions:

T_1:	T_2:
$read_1$ (A);	**$read_2$ (B);**
$read_1$ (B);	**$read_2$ (A);**
if (A = 0) then B := B + 1;	**if (B = 0) then A := A + 1;**
$write_1$ (B).	**$write_2$ (A).**

Let the consistency requirement be **A=0 or B=0**, with **A = B = 0** the initial values.

a) Show that every serial execution involving these two transactions preserves the consistency of the database.
b) What is a serilizable schedule?
c) What is a conflict-serializable schedule
d) Is there a concurrent execution of T_1 and T_2 that produces a conflict-serializable schedule? Explain and justify your answer.