National Exams May 2014
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 calculators.
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) 4 marks; (b) 4 marks; (c) 6 marks; (d) 6 marks

    Question 2: (a) 4 marks; (b) 4 marks; (c) 8 + 4 = 12 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) 7 marks; (b) 5 marks; (c) 4 marks; (d) 4 marks

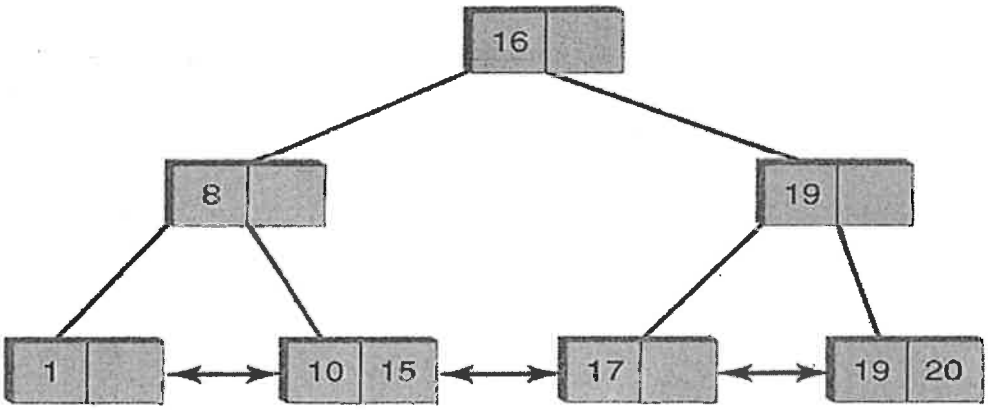
    Question 8: (a) 5 marks; (b) 5 marks; (c) 4 marks (c) 6 marks
5. All answers should be clear, legible and brief.

Question 1

a. Explain the differences between RAID level 0 and level 1.
b. RAID levels 3 and 5 involve striping and storing parity information, but how is RAID level 5 different from level 3?
c. Explain the rationale for using fillfactors less than 1 in
   i. Sorted indices,
   ii. B+ tree indices,
   iii. Hash indices.
d. Explain the difference between an equality search and a range search.

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. Consider the B+ tree in the figure below. Suppose that it was obtained by inserting a key into a leaf node of some other tree, causing a node split. What were the original tree and the inserted key? Is the solution unique? Explain your answer.



Question 3

Create an Entity-Relationship (ER) model for the bus company database system below. Identify the entities and their attributes, as well as the relationships between entities. For every relationship, identify its cardinality, possible participation constraints (or structural constraints) and, if appropriate, role names, weak entity types and identifying relationships.

The database should cover the following information:

- The company owns a fleet of buses. The following information is known for each bus: its license plate number, its type and the date the next maintenance is due. Each bus type has a name and a number of seats. (All busses of the same type have the same capacity).
- The company operates a set of bus lines. Each line has a line_name, line_source, and line_destination.
- The company schedules its trips on a weekly basis (that is, the schedules of two weeks are identical). A set of trips is scheduled for each line every week. A trip goes from line_source to line_destination. Different scheduled trips of one line may differ in their days of the week, their times, their stops, and the types of buses they can use. Each scheduled trip will make several stops at known locations (or bus stops) at known times.

Thus, the following information is relevant for each scheduled trip:

  - The day of the week and time of departure from line_source, as well as the day of the week and time of arrival to line_destination. No trip can last more than 24 hours, so the arrival day is redundant.
  - The bus types that can be used for the trip.
  - The locations and times of the stops. (Note that the list of locations for all the stops served by the bus company exists independently of particular trips and, thus, should be regarded as a separate abstract category).
- The company employs drivers and other personnel. For each employee, the name, social insurance number, address, and hourly wage and are relevant.
- The company keeps track of the actual trips performed. An actual trip is a realization of a scheduled trip. Since the trips are scheduled on a weekly basis, there are 52 actual trips per year for each scheduled trip. For every actual trip, the driver and bus used are known.

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 Entity Relationship (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 doctor-patient database schema (the primary keys are underlined):

Doctor(*licno*: **integer**, *drname*: **string**, *specialty*: **string**)
Patient(*patid*: **integer**, *patname*: **string**, *address*: **string**, *phone*: **string**, *DOB*: **date**)
Visit(*licno*: **integer**, *patid*: **integer**, *type*: **string**, *date*: **date**, *diagnosis*: **string**, *charge*: **string**)

This database is kept by a medical group, which consists of several physicians. Each physician has only one specialty. Patient records are shared among the doctors. For each patient visit, information is stored about the type (regular office, hospital, house call, etc), a single diagnosis for that visit and the charge.

A visit is identified by *licno*, *patid*, *date* because a patient may visit more than one doctor on a particular day. We assume, however, that a patient never visits the same doctor more than once in one day.

Write SQL for the following queries:

a. Get a list of all the doctors and their specialties.
b. Find *licno* of all doctors who saw patient Mary Adams.
c. Find the names of all doctors who have made house calls.
d. Find the names and addresses of all patients who have been diagnosed as having peptic ulcers.

Question 6

Consider the following relations containing student and course information:

Student (*Id, Name, Country*)
Course (*CrsCode, CrsName, Type, Instructor*)
Results(*Id, CrsCode, Grade)*

All the key fields are underlined. The *Type* field specifies the course type, e.g. MATH, STAT, SYSC, TTMG, ELEC, etc. The Results relation lists the grade that students (in Student relation) obtain for courses (in Course relation). Write SQL queries for a, b, and c:

a. Find the *Id* of students who take TTMG or SYSC course.
b. Find the *Id* of students who take every course.
c. Find the *Id* of students who take every SYSC course or take every TTMG course.
d. Consider the same schema (i.e. student and course relations), state in English what the following relational algebra query computes.

$$\pi_{Name}(\pi_{CrsCode}(\sigma_{Type = \text{'SYSC'}}Course) \bowtie (\sigma_{Grade=\text{'D'}}Result) \bowtie Student)$$

Question 7

Consider a relation schema with attributes ABCGWXYZ and the set of functional dependencies:

$$\begin{aligned} \mathbf{F} = \{ & XZ \rightarrow ZYB, \\ & YA \rightarrow CG, \\ & C \rightarrow W, \\ & B \rightarrow G, \\ & XZ \rightarrow G\}. \end{aligned}$$

Solve the following problems using the appropriate algorithms.

a. What is the minimal cover?
b. Is the dependency XZA → YB implied by **F** and **why**?
c. Is the decomposition into XZYAB and YABCGW lossless and **why**?
d. Is the decomposition in (c) above dependency preserving and **why**?

Question 8

a. The features that distinguish transactions from ordinary programs are abbreviated by the acronym ACID otherwise known as "ACID properties." What does this acronym (i.e. ACID) denote? Specify and explain each letter of the acronym.

b. Other than the letter "C" in the acronym "ACID", describe how the other three properties (A, I, and D) can cause system performance to suffer.

c. Given below is a schedule (for two transactions T_1 and T_2) that is produced by a non-strict two-phase locking concurrency control:

$$\mathbf{w_1(x)\ r_2(x)\ w_2(x)\ commit_2\ commit_1}$$

(i) Is the schedule serializable and **why**?
(ii) Is the schedule in commit order and **why**?

d. What happens to the schedule below at a REPEATABLE READ isolation level? Explain your answer.

$$\mathbf{r_1(x)\ r_1(y)\ w_1(x)\ r_2(y)\ r_2(x)\ w_1(y)\ commit_2\ commit_1}$$