National Exams December 2015

98-Comp-B3, Data Bases and 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 assumptions made.

2. This is a CLOSED BOOK EXAM.
   No calculators are needed or permitted.

3. Answer FIVE (5) questions as follows:
   a) One question from question 1 and 2 (Only one question will be marked).
   b) One question from question 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) 10 marks; (b) 10 marks;
   Question 6: (a) 10 marks; (b) 10 marks;
   Question 7: (a) 6 marks; (b) 7 marks; (c) 7marks
   Question 8: (a) 5 marks; (b) 5 marks; (c) 5 marks; (d) 5 marks

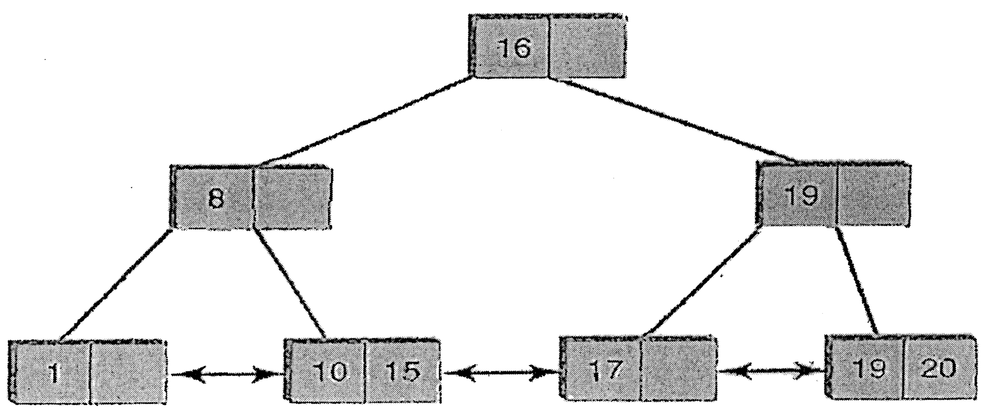
5. All answers should be clear, legible and brief.

Question 1

a. Explain the difference between an equality search and a range search. **(3 marks)**
b. Does the final structure of a B+ tree depend on the order in which the items are added to it? Explain your answer. **(3 marks)**
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: Edith, Carol, Betty, Debbie, Alice, Zelda, Wilma. **(14 marks)**
Note that question l(c) requires that you show a B+ tree for each insertion.

Question 2

a. Explain why a file can have only one clustered index? **(3 marks)**
b. Explain why a secondary, unclustered index must be dense? **(3 marks)**
c. Draw a B+ tree that results from inserting 3, 5, 11, 16, 18, 21 into the index of the B+ tree below **(14 marks)**



Note: For question 2(c) 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 (20 marks)

OllyWood Shows has decided to store information about actors who perform in television shows in a database. The company has wisely chosen to hire you as a database designer. The following information describes the situation that the OllyWood Shows database must model.

- Television shows are bundled and released as a set of episodes for a given season. An episode appears in only one season. A television show may have one or more seasons. A season has a season ID and title.
- Each episode recorded at OllyWood Films has a unique identification number, a title, a copyright date, a season ID, and a show ID.
- Each television show has a show ID, and a format (e.g., Blue Ray, DVD, Video).
- Each actor that acts at OllyWood Shows has an actor ID, SIN, a name, an address, and a phone number. Poorly paid actors often share the same address, and no address has more than one phone.
- Each role portrayed in an episode at OllyWood Films has a role name (e.g., John Wayne, Moses, Cleopatra), a gender (male, female, other) and a type (e.g., Lead Character, Supporting Character, Minor Character, Extra), an episode identification number, and an actor ID.
- A role can be portrayed in many different episodes in many different seasons. The same actor can portray a given role in many different episodes, or even all episodes in all seasons for a show. It is theoretically possible for a role to be played by a different actor in each episode. But the role for a single episode is played by exactly one actor.

Design a conceptual schema for OllyWood Shows 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 (20 Marks)

Consider the following description of the cBay Online Auction database system in which members (buyers and sellers) participate in the sale of items. The data requirements for this system are summarized as follows:

- The online site has members, each of whom is identified by a unique member number and is described by an email address, name, password, home address, and phone number.
- A member may be a buyer or a seller, or both. A buyer has a shipping address recorded in the database. A seller has a bank account number and routing number recorded in the database.
- Items for sale are placed by a seller and are identified by a unique item number assigned by the system. Items are also described by an item title, a description, starting bid price, bidding increment, the start date of the auction, and the end date of the auction.
- Items are categorized based on a fixed classification hierarchy. For example, a modem may be classified as Class: Computer, Subclass: Hardware and Description: Modem.
- Buyers make bids for items they are interested in. The bid price and the time of the bid are recorded.
- At the end of the auction, the bidder with the highest bid price is declared the winner. The transaction between the winner and the seller may then proceed.
- The buyer and the seller may record feedback regarding their completed transaction. The feedback contains a rating of the other party participating in the transaction (from 1 to 10) and a comment.

Design a conceptual schema for the cBay Online Auction database system 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 5

Consider the following relations containing Art Object information.
Art Objects are categorized based on their Type. For each Type, there is the unique type identifier, together with a description (painting, sculpture, statue or other). For each Art Object, there is the title, country of origin, the type, the purchasing cost as well as the date of the acquisition, the artist, and the collection it belongs to). Art Objects are grouped into Collections, for which the unique name (e.g. Lord Dalhousie Drawing, Dutch Renaissance, etc.), type (e.g. Permanent, Drawing, Renaissance, etc.) and the name of the contact person are kept. The Artist information is also recorded, including his/her name, date of birth, number of awards won and the country of birth.

ArtObject(AOid : integer; Title : string; Origin : string; TypeID : integer; Cost : integer; Acquired : date; Aid : integer; CName : string)
Type(TypeID : integer; description : string)
Artist(Aid : integer; Tname : string;DoB : date;Awards : integer; Country : string)
Collection(Cname : string; CType : string; Contactname : string)

Now, write the following queries in relational algebra:
a. Find the names of the Artists who have created Art Objects that are contained in at least two different Collections. For example, a Self Portrait of Vincent van Gogh may be included in the Permanent Collection, while his painting of Sunflowers may be included in a special exhibition about Artists who have lived in Provence **(10 Marks)**
b. Find the names of the Artists who have won more than three Awards, but who have never created any sculptures. **(10 Marks)**

Question 6

Consider, again, the following database scheme (equivalent to question 5 above).

ArtObject(Aid : integer; Title : string;Origin : string; TypeID : integer; Cost : integer; Acquired : date; Tid : integer;CName : string)
Type(TypeID : integer; description : string)
Artist(Tid : integer; Tname : string;DoB : date;Awards : integer;Country : string)
Collection(Cname : string;CType : string;Contactname : string)

a. What does the following SQL query find? **(5 marks)**

```
SELECT A.Cost, A.Origin
FROM ArtObject A
WHERE A.Acquired > 1999
GROUP BY A.Cost
HAVING 1 < (SELECT count(*)
            FROM ArtObject A2
            WHERE A.Cost = A2.Cost
            AND A2.Acquired > 1999)
```

b. Write the SQL query to find the names of Artists whom are listed in our database, but for whom we have never actually acquired any Art Objects. That is, there are no Art Objects of these Artists contained in our database. **(5 marks)**

Question 7

Consider the following relation regarding the Options installed in Cars sold at a Car Dealership, such as e.g. Cruise Control and Air Conditioning, together with the List and Discounted Prices of these Options.

CarSale(CarID : integer;OptionType : integer;OptionListP rice : currency; SaleDate : date;OptionDiscountedP rice : currency)

Assume that the following functional dependencies hold.

CarID → SaleDate
OptionT ype → OptionListP rice
CarID;OptionT ype → OptionDiscountedP rice

a. Define the term functional dependency. **(6 marks)**
b. Demonstrate that the decomposition of this schema into CarSale(CarID, SaleDate), Options(OptionType, OptionListPrice) and CarOptions(CarID, OptionType, OptionDiscountedPrice) is lossless. **(7 marks)**
c.. Explain why or why not the decomposition in b) above dependency preserving? **(7 marks)**

Question 8

Consider the following two transactions:

```
T1:
        read(x);
        read(y)
        if (x=1) then y:=y+2;
        write(y);
T2:
        read(y);
        read(x);
        if (y=1) then x:-x+2;
        write (x);
```

Let the consistency requirement be x=1 or y=1 with x=y=1 the initial values.
a) Show that every serial execution involving these two transactions preserves the consistency of the database. **(5 marks)**
b) What is a serilizable schedule? **(5 marks)**
c) What is a conflict-serializable schedule? **(5 marks)**
d) Is there a concurrent execution of T1 and T2 that produces a conflict-serializable schedule? Explain and justify your answer. **(5 marks)**