BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2005-2006

Comprehensive Examination

Course No: CS UC352 **Course Title: Database Systems**
Date: 31st May 2006 **Duration: 3 Hours** **Total Marks: 40**

1. A company database needs to store information about employees (identified by *ssn*, with *salary* and *phone* as attributes); departments (identified by *dno*, with *dname* and *budget* as attributes); and children of employees (with *name* and *age* as attributes). Employees *work* in departments; each department is *managed* by an employee; a child must be identified uniquely by *name* when the parent (who is an employee; assume that only one parent works for the company) is known.

   Draw an ER diagram that captures this information. (State explicitly any assumptions that you make about the model.) [3 marks]

2. (a) Define the terms: key, primary key, super key. Give one example of each.
   (b) You are given a relation R with attributes, ABCD. For the set of FDs given below do the following:
   (i) Identify the key(s) of R.
   (ii) Identify the best normal form that R satisfies (3NF or BCNF).
   (iii) If R is not in BCNF, decompose it into a set of BCNF relations.
   $A \rightarrow B,\ BC \rightarrow D,\ A \rightarrow C$ [1½+3 marks]

3. Given two relations R1 and R2, where R1 contains N1 tuples, R2 contains N2 tuples, and N2>N1>0, give the minimum and maximum possible sizes (in tuples) for the result relation produced by each of the following relational algebra expressions. In each case, state any assumptions for R1 and R2 that are needed to make the result meaningful.
   (a) R1 ∪ R2 (b) R1 X R2 (c) $\sigma_{a=5}$ (R1) (d) π_a (R1) [2 marks]

4. Consider the following relations:

   Student (*snum*: **integer**, *sname*: **string**, major: **string**, *year*: **integer**, *age*: **integer**)
   Course (*cname*: **string**, *meets_at*: **time**, *room*: **string**, *fid*: **integer**)
   Enrolled (*snum*: **integer**, *cname*: **string**)
   Faculty (*fid*: **integer**, *fname*: **string**, *deptid*: **integer**)

   Write the following queries in SQL:
   (a) Find the names of all third year students who are enrolled in a course taught by *I.Teach*.
   (b) Find the age of the oldest student who is a CS major.
   (c) Find the names of all courses that either meet in room 413 or have five or more students.

[PTO]

(d) Find the names of all students who are enrolled in two courses that meet at the same time.
(e) Print the year and the average age of students of that year, for each year.
(f) Find the names of students who are not enrolled in any courses. [3 marks]

5. Mention three reasons that cause record lengths to vary. For each of these, describe a suitable storage scheme for storing such records. [3marks]

6. Explain the difference between each of the following:
   (a) Primary versus secondary indices.
   (b) Dense versus sparse indices. [1+1 marks]

7. Suppose we store a relation R(x, y, z) in a partitioned hash table with 1024 buckets (i.e., 10-bit bucket address). Queries about R each specify exactly one of the attributes, and each of the attributes is equally likely to be specified. If the hash function produces 5 bits based only on x, 3 bits based only on y, and 2 bits based only on z, what is the average number of buckets that need to be searched to answer a query? [1½ marks]

8. State whether the following results are true for **bags** or not. If true give a proof, otherwise give a counter example.
   (a) $R \cup R = R$
   (b) $R \cap R = R$
   (c) $R \cap (S \cup T) = (R \cap S) \cup (R \cap T)$ [3 marks]

9. The consistency constraint on a database is $0 \leq A \leq B$. A transaction T in the database is given below.

   B := A+B; A := A+B;

   (a) Check whether T preserves the consistency of the database.
   (b) Show the undo-log and redo-log records assuming that initially A=5, B=10. [1+2 marks]

10. Below are two transactions in a database.
    T1: READ(A, t); t := t+2; WRITE(A, t); READ(B, t); t := t*3; WRITE(B, t);
    T2: READ(B, s); s := s*2; WRITE(B, s); READ(A, s); s := s+3; WRITE(A, s);
    (a) Show that both serial orders (T1, T2) and (T2, T1) have the same effect on the database independent of the initial values of A and B.
    (b) Give examples of a serializable schedule and a nonserializable schedule of the 12 actions of T1 and T2.
    (c) How many serial schedules of the 12 actions are there? [2+2+1 marks]

11. Explain one deadlock detection and recovery and one deadlock prevention methods. Discuss the advantages and disadvantages of these two approaches to deadlock handling. [2+2+1 marks]

12. (a) What is the purpose of semi-structured data in databases? Illustrate how this is achieved with the help of an example.
    (b) Write a one-page note on the latest developments related to XML based on the literature survey you did using the EBSCO online bibliographic database. [2+3 marks]

♣♣♣♣♣

BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2005-2006

TEST II Question Paper

Course No: CS UC352 **Course Title: Database Systems**
Date: 23rd Apr 2006 **Duration: 50 minutes** **Total Marks: 20**

Open book
(Permitted to refer Textbook and class notes.)

(Answer ALL questions)

1. Suppose a record has the following fields in this order: A record header consisting of two 4-byte pointers and a character, a character string of length 15, an integer of 2 bytes, an SQL date, and an SQL time (no decimal point).

   (a) How many bytes does the record take if:
   (i) Fields can start at any byte.
   (ii) Fields can start at a byte that is a multiple of 4.
   (iii) Fields can start at a byte that is a multiple of 8. (3 marks)

   (b) Suppose we wish to pack as many records as we can into a block of 4096 bytes, using a block header that consist of ten 4-byte integers, how many records can we fit in the block in each of the three situations regarding field alignment (i) through (iii) of part (a)? (1.5 marks)

2. (a) In a secondary index scheme, assume that blocks can hold either three records, ten key-pointer pairs, or fifty pointers. If the average search key appears in 10 records, how many blocks do we need to hold 3000 records and its secondary index structure, if we use buckets? How many blocks do we need if we do not use buckets? (2+1 marks)

   (b) Suppose keys are hashed to four-bit sequences in an extensible hashing scheme. Also suppose that blocks can hold three records. If we start with a hash table with two empty blocks (corresponding to 0 and 1), show the organisation after we insert 16 records with keys 1111, 1110, ..., 0001, 0000. (3 marks)

[PTO]

3. The table for a relation PC is given below.

model	speed	ram	hd
1001	700	64	10
1002	1500	128	60
1003	866	128	20
1004	866	64	10
1005	1000	128	20
1006	1300	256	40
1007	1400	128	80
1008	700	64	30
1009	1200	128	80
1010	750	64	30
1011	1100	128	60
1012	733	256	60

(a) We wish to design an index on speed and hard disk only. Plot the entries in the table using these two parameters for coordinate values. (1 mark)

(b) Choose five grid lines (total for the two dimensions), so that there are no more than two points in any bucket. (2.5 marks)

(c) Choose a partitioned hash function with one bit each of the three attributes speed, ram, and hard-disk that divides the data given in the table almost uniformly. (1.5 marks)

4. (a) A database consists of two relations, whose schemas are:

Product (maker, model, type)

PC (model, speed, ram, hd, rd, price)

Suggest two plans for executing the following query in this database and compare the efficiencies of the two plans.

```
SELECT maker
FROM Product, PC
WHERE (Product.model = PC.model) AND
        (speed>=160);
```

(3 marks)

(b) State whether the following transformations are true or false? Prove your answer.

(i) $\sigma_{p1 \wedge p2}(R) = \sigma_{p1}(R) \cap \sigma_{p2}(R)$

(ii) $\sigma_{p \wedge q}(R \blacktriangleright\blacktriangleleft S) = \sigma_{p}(R) \blacktriangleright\blacktriangleleft \sigma_{q}(S)$, where p consists of attributes from R only and q consists of attributes from S only.

(iii) $\sigma_{p}(R \cup S) = \sigma_{p}(R) \cup \sigma_{p}(S)$ (1.5 marks)

♣♣♣♣♣

BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2005-2006

TEST I Question Paper

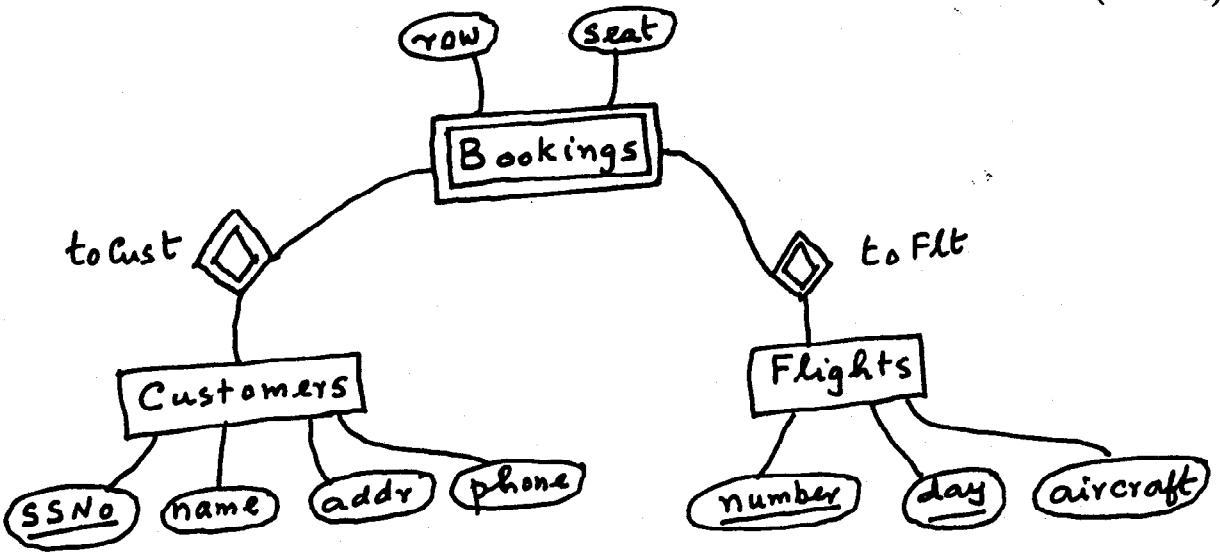
Course No: CS UC352 **Course Title: Database Systems**
Date: 12th Mar 2006 **Duration: 50 minutes** **Total Marks: 20**

Data Provided and complete. Closed book.

(Answer ALL questions)

1. (a) Draw an E/R diagram for bank database that includes information about customers and their accounts. Information about a customer includes their name, address, phone no, and a unique customer_id assigned to each customer by the bank. Accounts have numbers, types (e.g. savings, current) and balances. We also need to record the customer who owns an account. Include arrows where appropriate, to indicate multiplicity of a relationship. Assume that a customer can own many accounts, but an account is owned by only one customer. (3 Marks)
   (b) Define a weak entity set. Give an example. (2 Marks)

2. (a) Convert the following E/R diagram to a set of tables. Underline the keys of the tables. (3 Marks)



[PTO]

(b) Consider a relation with schema R(A, B, C, D) and
FDs $AB \rightarrow C$, $C \rightarrow D$, and $D \rightarrow A$.
(i) What are the nontrivial FD's that follow from the given FDs? You should restrict yourself to FDs with single attributes on the right side. (1.5 Marks)
(ii) What are all the keys of R? (1.0 Mark)
(iii) What are all the superkeys of R that are not keys? (0.5 Mark)

3. For the relation R(A, B, C, D, E) with FDs $AB \rightarrow C$, $C \rightarrow D$, $D \rightarrow B$ and $D \rightarrow E$ do the following. (4 Marks)
(a) Indicate all the BCNF violations. Do not forget to consider the FDs that are not in the given set, but follow from them. It is not necessary to give violations that have more than one attribute on the right side.
(b) Decompose the relations, as necessary, into collections that are in BCNF.
(c) Indicate all the 3NF violations.
(d) Decompose the relations, as necessary, into collections of relations that are in 3NF.

4. A database consists of four relations, whose schemas are:
Product(maker, model, type)
PC(model, speed, ram, hd, rd, price)
Laptop(model, speed, ram, hd, screen, price)
Printer(model, color, type, price)

Note:
(i) Attribute **type** of relation **Product** takes values **pc, laptop,** or **printer**.
(ii) Attribute **type** of relation **Printer** takes values **ink-jet, laser** or **bubble**.
(iii) Attribute **color** of relation **Printer** takes values **true** or **false**.

Write expressions of relational algebra to answer the following queries. (5 Marks)
(a) What PC models have speed of at least 1000?
(b) Which manufacturers make laptops with a hard disk of at least one gigabyte?
(c) Find the model number and price of all products (of any type) made by manufacturer B.
(d) Find the model numbers of all color laser printers.
(e) Find those manufacturers that sell Laptops, but not PCs.

♣♣♣♣♣