

BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2006-2007

Comprehensive Examination

Course No: CS UC352
Date: 28th May 2007
Duration: 3 Hours

Course Title: Database Systems
Weightage: 40%
Max. Marks. 40

Write answers to all questions, including answers to question no.1, in the main answer book only

1.

- (a) We use the term ----- to denote a candidate key that is chosen by the database designer as the principal means of identifying tuples within a relation.
- (b) Give a relational algebraic expression for the natural join of two relations r and s having schema R and S with $R \cap S = \{A_1, A_2, \dots, A_n\}$ using selection, projection, cross product, and set operations.
- (c) In SQL the ----- clause provides a way of defining a temporary view whose definition is available only to the query in which this clause occurs.
- (d) In an E-R diagram what type of an attribute is represented by a double ellipse?
- (e) Define Boyce-Codd normal form (BCNF).
- (f) State Armstrong's axioms for functional dependencies.
- (g) What is the expansion of XML?
- (h) What is the name of the file organisation in which records of several different relations are stored in the same file?
- (i) Indices whose search key specifies an order different from the sequential order of the file are called ----- indices.
- (j) What is the indexing scheme based on tree structure in which pointers to records in the file are kept only at the leaf nodes?
- (k) What is the worst-case time complexity, in terms of disk I/O, of the binary search if the file is ordered on an attribute and the selection condition is an equality comparison on that attribute?
- (l) What is the name of the strategy used for reducing the number of temporary files that are produced by passing the results of one operation to the next one?
- (m) Find an expression equivalent to $\sigma_{\theta_1}(E_1) \bowtie \sigma_{\theta_2}(E_2)$ that does not contain the σ operator.
- (n) Give an estimate of the size of the join of two relations $r(R)$ and $s(S)$ having A as the only common attribute.
- (o) What are the ACID properties of transactions in a database?
- (p) State a condition for a concurrent schedule to be conflict serializable.
- (q) Which are the two phases of a two-phase locking protocol?

- (r) In the timestamp-ordering protocol what is the condition that will permit a transaction T to read an item Q?
- (s) What are the two basic methods of storing relations in a distributed database?
- (t) What is the purpose of the bully algorithm in distributed databases?

(20 x 0.5 = 10 marks)

2. A database consists of the following relations, where the primary keys are underlined.

employee(person_name, street, city)
 works(person_name, company_name, salary)
 company(company_name, city)
 manages(person_name, manager_name)

(a) Give **relational algebraic expressions** to express the following queries. (2.5 marks)

- (i) Find the names of all employees who work for First Bank Corporation.
- (ii) Find the names, street address, and cities of residence of all employees who work for First Bank Corporation.
- (iii) Find the names of all employees in this database who live in the same city as the company for which they work.
- (iv) Assume the companies may be located in several cities. Find all companies located in every city in which Small Bank Corporation is located.
- (v) Find all employees in the database who live in the same cities and on the same streets as do their managers.

(b) Give an expression in **SQL** for each of the queries of part (a). (2.5 marks)

3.(a) Construct an E-R diagram for car insurance company whose customers own one or more cars each. Each car has associated with it zero to any number of recorded accidents. Also convert the E-R diagram to a set of relations. (2+1 marks)

(b) A schema $R = (A, B, C, D, E)$ has a set of functional dependencies $F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$. Is the decomposition of R into (A, B, C) and (C, D, E) lossless? Justify your answer. (2 marks)

4. (a) List two advantages and two disadvantages of each of the following strategies for storing a relational database:

- (i) Store each relation in one file.
- (ii) Store multiple relations in one file. (2 marks)

(b) Explain briefly static and dynamic hashing. What are the advantages and disadvantages of these methods. (3 marks)

5.(a) Give the pseudo code of any method of your choice of an algorithm for join operation. What is the cost of the algorithm in terms of number of seeks and number of block transfers? (3 marks)

(b) Consider the relations $r_1(A, B, C)$, $r_2(C, D, E)$, and $r_3(E, F)$ with primary keys A , C , and E respectively. Assume that r_1 has 1000 tuples, r_2 has 1500 tuples, and r_3 has 750 tuples. Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$ and give an efficient strategy for computing the join. (2 marks)

6.(a) Consider the following two transactions:

T1: read(A);
read(B);
if A=0 then B = B+1;
write(B);

T2: read(B);
read(A);
if B=0 then A = A+1;
write(A);

- (i) Show that every serial execution involving these two transactions preserves the consistency of the database.
- (ii) Show a concurrent execution of T1 and T2 that produces a nonserializable schedule.
- (iii) Is there a concurrent execution of T1 and T2 that produces a serializable schedule?

(3 marks)

(b) Add lock and unlock instructions to transactions T1 and T2 of part (a), so that they observe the two-phase locking protocol. Can the execution of these two transactions result in a deadlock? Justify your answer.

(2 marks)

7.(a) Describe briefly what is replication and fragmentation of data. When is it useful to have these? Explain your answer.

(3 marks)

(b) Mention one type of failure in a distributed system. Explain how the two-phase commit (2PC) protocol ensures transaction atomicity despite the failure.

(2 marks)

BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2006-2007

TEST II - Regular

Course No: CS UC352
Date: 22nd April 2007

Course Title: Database Systems
Duration: 50 minutes

Total Marks: 20

Open Book

Permitted to refer Text Book, PPTs, and Class Notes

(Answer ALL questions)

1. (a) Draw an E/R diagram for a sales database that includes information about sales representatives, sales areas, and products. Each representative is responsible for sales in one or more areas; each area has one or more responsible representatives. Similarly, each representative is responsible for sales of one or more products, and each product has one or more responsible representatives. Every product is sold in every area; however, no two representatives sell the same product in the same area. (2 Marks)
- (b) Convert the E/R diagram you obtained in part (a) to a set of tables. Underline the keys of the tables. Are all the constraints of the model included in the your design? If not, state them in terms of the tables and their attributes that you created. (3 Marks)

2. (a) Consider a relation with schema $R(A, B, C, D)$ and FDs $A \rightarrow B$, $B \rightarrow C$, and $B \rightarrow D$.
 - (i) What are the nontrivial FDs that follow from the given FDs? You should restrict yourself to FDs with single attributes on the right side. (1.0 Mark)
 - (ii) What are all the keys of R ? (0.5 Mark)
 - (iii) What are all the superkeys of R that are not keys? (1.5 Marks)
- (b) For the relation of part (a) do the following. (4 Marks)
 - (i) 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.
 - (ii) Decompose the relations, as necessary, into collections that are in BCNF.
 - (iii) Indicate all the 3NF violations.
 - (iv) Decompose the relations, as necessary, into collections of relations that are in 3NF.

[PTO]

BITS, Pilani – Dubai Campus
Knowledge Village, Dubai

Third Year (Computer Science & Engg.)
Second Semester, 2006-2007

TEST 1

Course No: CS UC352
Date: 11th Mar 2007

Duration: 50 minutes

Course Title: Database Systems
Total Marks: 20

(Answer ALL questions)

All the questions in this paper are based on a database consisting of the following tables.

Product(maker, model, type)

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

Laptop(model, speed, ram, hd, screen, price)

Printer(model, color, type, price)

Write SQL statements for the following:

(10 x 2 = 20 marks)

1. Find the model number, speed, and hard-disk size for all PC's whose price is under \$1600.
2. Find all the tuples in the Printer relation for color printers. Color is a boolean-valued attribute.
3. Find those manufacturers that sell Laptops, but not PC's.
4. Find makers of PC's with a speed of at least 160.
5. Find the printers with the highest price.
6. Find the average speed of PC's.
7. Find for each manufacturer, the average screen size of its Laptops.
8. Using two INSERT statements store in the database the fact that PC model 1100 is made by manufacturer C, has a speed 240, RAM 32, hard disk 2.5, a 12x CD, and sells for \$2499
9. Delete all PC's with less than 2 gigabytes of hard disk.
10. Write a suitable schema for the relation Laptop.

7. Find the project numbers of projects using at least one part available from supplier S1.

8. Insert into P the tuple (P1, Nut, Red, 12.0, Dubai).

9. Delete all shipments of quantity less than 100.

10. Double the status of all suppliers from Dubai. (The status of all other suppliers should remain the same.)