Reg. No. : ☐☐☐☐☐☐☐☐☐☐☐☐☐

Question Paper Code : 80443

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2024.

Third/Fourth/Fifth/Eighth Semester

Computer Science and Engineering

CS 8492 — DATABASE MANAGEMENT SYSTEMS

(Common to Computer and Communication Engineering/Mechanical and Automation Engineering/Mechatronics Engineering/Computer Science and Business Systems/Information Technology)

(Regulations 2017)

Time : Three hours   Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What are the main functions of a database administrator?

2. Give few SQL Aggregate functions.

3. What is lossless decomposition? Give example.

4. Give an example for a weak and a strong entity.

5. What is a cascadeless schedule? Why is cascadelessness of schedules desirable?

6. What benefit does strict two-phase locking provide? What disadvantages result?

7. What is scrubbing, in the context of RAID systems, and why is scrubbing important?

8. Why might the leaf nodes of a B+-tree file organization lose sequentially?

9. When is it useful to have replication or fragmentation of data?

10. What is the difference between a false positive and a false drop?

PART B — (5 × 13 = 65 marks)

11. (a) (i) Consider the relational database given in Figure 1. What are the appropriate primary keys? (3)

(ii) Consider the relational database given in Figure 1. Give an expression in the relational algebra to express each of the following queries:

(1) Find the names of all employees who live in city "Miami". (2)

(2) Find the names of all employees whose salary is greater than $100,000. (2)

(3) Find the names of all employees who live in "Miami' and whose salary is greater than $100,000. (1)

(iii) Consider the relational database of Figure 1. Give an expression in the SQL to express each of the following queries:

(1) Find the names of all employees who work for "First Bank Corporation". (2)

(2) Find the names and cities of residence of all employees who work for "First Bank Corporation". (2)

(3) Find the names, street address, and cities of residence of all employees who work for "First Bank Corporation" and earn more than $10,000. (1)

```
employee (person name, street, city)
works (person name, company name, salary)
company (company name, city)
```

Figure 1 Relational Database

Or

(b) (i) Explain the difference between logical and physical data independence with example. (6)

(ii) Why would you choose a database system instead of simply storing data in operating system files? When would it make sense not to use a database system? (7)

12. (a) (i) Explain the distinctions among the terms primary key, candidate key, and super key? (6)

(ii) Explain the steps involved in BCNF with example. (7)

Or

(b) (i) Construct an E-R diagram for a hospital with a set of patients and a set of medical doctors, Associate with each patient a log of the various tests and examinations conducted. (6)

(ii) 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. Specify any necessary overlap and covering constraints as well. (7)

2   80443

13. (a) (i) In multiple-granularity locking, what is the difference between implicit and explicit locking? (6)

   (ii) Since every conflict-serializable schedule is view serializable, why do we emphasize conflict serializability rather than view serializability? (7)

Or

(b) (i) Consider the following locking protocol: All items are numbered, and once an item is unlocked, only higher-numbered items may be locked. Locks may be released at any time. Only X-locks are used. Show by an example that this protocol does not guarantee serializability. (6)

   (ii) Explain the purpose of the checkpoint mechanism. How often should checkpoints be performed? How does the frequency of checkpoints affect: (7)
   • System performance when no failure occurs?
   • The time it takes to recover from a system crash?
   • The time it takes to recover from a media (disk) failure?

14. (a) (i) Explain the distinction between closed and open hashing. Discuss the relative merits of each technique in database applications. (6)

   (ii) List the physical storage media available on the computers you use routinely. Give the speed with which data can be accessed on each medium. (7)

Or

(b) (i) Consider the relations r1 (A, B, C), r2(C, D, E), and r3(E, F), with primary keys A, C, and E, respectively. Assume that r1 has 1000 tuples, r2 has 1500 tuples, and r3 has 750 tuples, Estimate the size of r1 ⋈ r2 ⋈ r3, and give an efficient strategy for computing the join. (6)

   (ii) Suppose you have to create a B+-tree index on a large number of names, where the maximum size of a name may be quite large (say 40 characters) and the average name is itself large (say 10 characters). Explain how prefix compression can be used to maximize the average fan out of nonleaf nodes. (7)

15. (a) (i) Discuss the relative advantages of centralized and distributed databases. (6)

   (ii) Explain the distinction between a type x and a reference type ref(x). Under what circumstances would you choose to use a reference type? (7)

Or

(b) (i) Discuss about various information retrieval models and gives measures to evaluate the effectiveness of information retrieval. (6)

(ii) A car-rental company maintains a database for all vehicles in its current fleet. For all vehicles, it includes the vehicle identification number, license number, manufacturer, model, date of purchase, and color. Special data are included for certain types of vehicles: (7)

- Trucks: cargo capacity.
- Sports cars: horsepower, renter age requirement.
- Vans: number of passengers.
- Off-road vehicles: ground clearance, drivetrain (four- or two-wheel drive).

Construct an SQL schema definition for this database. Use inheritance where appropriate.

PART C — (1 × 15 = 15 marks)

16. (a) Design a database for an automobile company to provide to its dealers to assist them in maintaining customer records and dealer inventory and to assist sales staff in ordering cars.

Each vehicle is identified by a Vehicle Identification Number (VIN). Each individual vehicle is a particular model of a particular brand offered by the company (e.g., the XF is a model of the car brand Jaguar of Tata Motors). Each model can be offered with a variety of options, but an individual car may have only some (or none) of the available options. The database needs to store information about models, brands, and options, as well as information about individual dealers, customers, and cars. Your design should include an E-R diagram, a set of relational schemas, and a list of constraints, including primary-key and foreign-key constraints.

Or

(b) (i) The Google search engine provides a feature whereby Web sites can display advertisements supplied by Google. The advertisements supplied are based on the contents of the page. Suggest how Google might choose which advertisements to supply for a page, given the page contents. (7)

(ii) Consider the following two transactions: (8)
T13: read(A);
read(B);
if A = 0 then B: = B + 1;
write(B).
T14: read(B);
read (A)
if B = 0 then A: = A + 1;
write(A).

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

(1) Show that every serial execution involving these two transactions preserves the consistency of the database.

(2) Show a concurrent execution of T13 and T14 that produces a nonserializable schedule.

(3) Is there a concurrent execution of T13 and T14 that produces a serializable schedule?