# APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

## FOURTH SEMESTER B.TECH DEGREE EXAMINATION, MAY 2017

### Department: Computer Science and Engineering

*Subject: - CS208: Principles of Database Design*

Time: 3 hours                                                                 Max. Marks: 100

---

## PART A
### Answer *all* questions

1. Highlight the major differences between structured, unstructured, and semi-structured data. (3)

2. What facts about the relationships between entities TEACHER and COURSE are conveyed by the following ER diagram? (3)

| EMPLOYEE | (0, 2) | WORKSIN | (3, 5) | PROJECT |

3. Illustrate how referential integrity constraint is represented using a foreign key. (3)

4. Distinguish between super key, candidate key and primary key using a real convincing example. (3)

**Total: (12)**

## PART B
### Answer any *two* full questions

5. Design an ER diagram for the following scenario:. (9)
   There is a set of teams, each team has an ID (unique identifier), name, main stadium, and to which city this team belongs. Each team has many players, and each player belongs to one team. Each player has a number (unique identifier), name, DoB, start year, and shirt number that he uses. Teams play matches, in each match there is a host team and a guest team. The match takes place in the stadium of the host team. For each match we need to keep track of the following: The date on which the game is played The final result of the match. The players participated in the match. For each player, how many goals he scored, whether or not he took yellow card, and whether or not he took red card. During the match, one player may substitute another player. We want to capture this substitution and the time at which it took place. Each match has exactly three referees. For each referee we have an ID (unique identifier), name, DoB, years of experience. One referee is the main referee and the other two are assistant referee.

6. 
```
EMPLOYEE(ENO, NAME, ADDRESS, DOB, AGE, SALARY, DNUM, SUPERENO)
DEPARTMENT(DNO, DNAME, DLOCATION, DPHONE, MGRENO)
PROJECT(PNO, PNAME, PLOCATION, PCOST, CDNO)
```
(9)

In the above relations DNUM is a foreign key that identifies the department to which an employeebelongs. MGRENO is a foreign key identifying the employee who manages thedepartment. CDNO is a foreign key identifying the department that controls theproject. SUPERENO is a foreign key identifying the supervisor of each employee.

Write relational algebra expressions for the following queries:-
   (a) Names of female employees whose salary is more than 20000.
   (b) Salaries of employee from 'Accounts' department
   (c) Names of employees along with his/her superviser's name
   (d) For each employee return name of the employee alog with his department name and the names of projects in which he/she works.

7.  a.  With the helpof an example, establish that a relationship of degree 3 need not be equivalent to three relationships with degree 2. (4)

    b.  Show how the division operation can be implemented using the basic relation algebra operations. Illustrate its application through a realistic example. (5)

**Total: (18)**

## PART C
### Answer *all* questions

8.  What is meant by correlated nested query? Illustrate with a real example. (3)

9.  List any two uses of triggers. Write a trigger which raises an error if more than three students are deleted simultaneously from the STUDENT table. (3)

10. List out and illustrate the *two* anomalies in relational databases design. (3)

11. Given the FDs P→Q, P→R, QR→S, Q→T, QR→U, PR→U, write the sequence of Armstrong's Axioms needed to arrive at a. P → T    b. PR → S  c. QR → SU (3)

**Total: (12)**

## PART D
### Answer any *two* full questions

12. For the relation schema below, give an expression in SQL for each of the queries that follows: (9)

   **employee(employee-name, street, city)**
   **works(employee-name, company-name, salary)**
   **company(company-name, city)**
   **manages(employee-name, manager-name)**

   a)  Find the names, street address, and cities of residence for all employees who work for the Company 'RIL Inc.' and earn more than $10,000.
   b)  Find the names of all employees who live in the same cities as thecompanies for which they work.
   c)  Find the names of all employees who do not work for 'KYS Inc.'. Assume that all people work for exactly one company.
   d)  Find the names of all employees who earn more than every employeeof 'SB Corporation'. Assume that all people work for at most one company.

13. Consider the un-normalized relation R(A, B, C, D, E, F, G) with the FDs A→B , AC→G, (9)

AD→EF, EF→G, CDE→AB.  Trace the normalization process to reach 3NF relations.

14. Let R = ABCDE, R1 = AD, R2 = AB, R3 = BE, R4 = CDE, and R5 = AE. Let the FDs be: A→C, B→C, C→D, DE→C, CE→A. Show how to test if the decomposition of R into {R1,R2,R3,R4,R5} is a lossless join decomposition. (9)

**Total: (18)**

## PART E
### Answer any four full questions

15. a. What is the difference between *materialization* and *pipelining* in query execution? Illustrate. (2)
    b. Illustrate the dirty read problem with a real example. (3)
    c. Determine if the following schedule is recoverable. Is the schedule cascade-less? Justify your answers. (5)

    r1(X), r2(Z), r1(Z), r3(X), r3(Y), w1(X), c1, w3(Y), c3, r2(Y), w2(Z), w2(Y), c2

    (Note: $r_i(X)/w_i(X)$ means transaction $T_i$ issues read/write on item X; $c_i$ means transaction $T_i$ commits.)

16. a. Discuss the significance of RDF with the help of a small example. (4)
    b. For a relation R(A,B,C,D) with A as primary key assume that, (6)
    number of records in the relation = 500.000 , size of a record (fixed length records) = 200 byte, size of attribute A = 16 bytes, size of a pointer = 4 bytes, block size = 2400 bytes
    Sequential organization on key A is used.
       i.   Compute the number of blocks of the file.
       ii.  Compute the number total number of blocks of a static multilevel index   A.
       iii. Compute the maximum number of block accesses for a query like *select \* from R where A = "….."*

17. a. Write a note on Biological databases (3)
    b. Consider the following tables representing enrollment of students to courses: STUDENT (<u>ROLLNO</u>, NAME, AGE, ADDRESS, EMAIL), COURSE(<u>CNO</u>, CNAME, AREA)  and ENROLLEMENT(<u>ROLLNO, CNO</u>,GRADE), where ROLLNO and CNO in ENROLLMENT are foreign keys referring to the primary keys with the same names. (7)
    Show an initial query tree for the following query and optimize it using the rules of heuristics:

    ```
    SELECT CNAME, NAME, EMAIL, GRADE
    FROM STUDENT, COURSE, ENROLLMENT
    WHERE COURSE.CNO=ENROLLMENT.CNO
    and STUDENT.ROLLNO = ENROLLMENT.ROLLNO
    and COURSE.CNAME = 'dbs'
    and STUDENT.NAME = 'Raju'
    ```

    Assume that CNAME is a candidate key of COURSE.

18. a. "In the recovery of *concurrent* transactions under the *immediate database modification* scheme, undo operations must be done *before* redo operations." Justify this statement with the support of a small sample log. (3)
    b. Argue that two-phase locking always ensure serializability. (4)
    c. List out any three features of GIS. (3)

**Total: (40)**