SIT103 – Database and Information Retrieval

Sample paper

SPECIAL INSTRUCTIONS

1. Answer **ALL** the questions.
2. There are **100 total** marks for this exam. Please read the section instructions carefully before answering the questions.
3. The exam consists of three sections

- **Section A (15 Marks)**
  There are 15 multiple-choice questions, each worth 1 mark. Select the correct answer for all the questions.

- **Section B (45 Marks)**
  Use the given relational schema and write SQL queries answering all the questions.

- **Section C (40 marks)**
  This section has questions related to data modelling and normalization. Answer all the questions.

Section A (Multiple Choice Questions)

1. The _____ set operator returns only tuples that are in both relations.
   A. intersect
   B. union
   C. project
   D. divide
   E. combine

2. A(n) _____ is where tuples from one relation that do not have matching values in the common attributes of the second relation are also included in the result of performing the join and any missing values in the second relation are set to null.
   A. Natural join
   B. Theta join
   C. Natural join
   D. Join
   E. Equi join
3. A relational schema that has no partial functional dependencies is in _____ form
   A. 1st Normal Form
   B. 2rd Normal Form
   C. 3rd Normal Form
   D. BCNF
   E. Not normalised

4. What is NOT a property of a transaction?
    A. Atomicity
    B. Serialisibility Serializability
    C. Consistency
    D. Isolation
    E. Durability
5. If there is more than one key for relation schema in DBMS then each key in relation schema is classified as
    A. prime key.
    B. super key.
    C. candidate key.
    D. primary key.
    E. Composite Key
6. Form of dependency in which set of attributes that are neither a subset of any of keys nor candidate key is classified as
    A. transitive dependency.
    B. full functional dependency.
    C. partial dependency.
    D. prime functional dependency.
    E. Non dependency
7. _____ is collection of interrelated data and set of program to access them.
    A. Programming Language
    B. Data Structure
    C. Database Management System
    D. Database
    E. Data Definition language
8. _____ensures that once transaction changes are done, they cannot be undone or lost, even in the event of a system failure
    A. Atomicity
    B. Serialisibility
    C. Consistency
    D. Isolation
    E. Durability
9. Deadlocks are possible only when one of the transactions wants to obtain a(n) ____ lock on a data item.
    A. Binary
    B. Exclusive
    C. Shared
    D. Complete
    E. Optimistic
10. A DBMS uses a transaction ____ to keep track of all transactions that update the database
    A. Log
    B. Table
    C. Block
    D. Statements
    E. Evidences
11. Although the ____ locking approach improves the availability of data, its management requires high overhead
    A. Database Level
    B. Table Level
    C. Page Level

D. Block level
E. Record Level

12. This is NOT an advantage of a distributed database
    A. Easy for often distributed business applications
    B. Improved performance
    C. Easy data processing and retrieval
    D. Expansion
    E. Scalability
13. A decision support systems are / does
    A. Short term and day today operations
    B. Transaction processing
    C. Functional system
    D. information extraction, derived from operational data
14. This is Not a main characteristic of a data warehouse
    A. Integrated
    B. Consistent
    C. Subject – oriented
    D. Time variant
    E. Non volatile
15. What is a problem of data replication?
    A. Updating Copies and managing concurrency
    B. Configuring the databases
    C. Writing queries
    D. Scaling the databases
    E. Connecting with the data warehouses

Section B (SQL)

Student(**snum**: integer, sname: string, major: string, level: string, age: integer)

Class (**name: String**, meets_at:time, room : string, fid: integer)

Enrolled(**snum:** integer**, cname**: String)

Faculty(**fid:** integer, fname: string: deptid : integer)

The meaning of these relations is straightforward; for example, Enrolled has one record per student-class pair such that the student is enrolled in the class.

Write following queries in SQL. No duplicates should be printed in any of the answers.

1. Find the names of all juniors (Level = JR) who are enrolled in a class taught by I.Teach (name of the faculty).

```
SELECT DISTINCT s.sname
FROM Student s, Class c, Enrolled e, faculty f
Where s.snum = e.enum AND E.cname= c.name ANDc.fid = f.fid AND f.fname = 'I.Tech'
AND s.level = 'JR'
```

2. Find the age of the oldest student who is ether a history major or is enrolled in a course taught by I.Teach (name of the faculty)

```
SELECT MAX(s.age)
FROM Student s
WHERE (s.major = 'History') OR s.num IN
      (SELECT e.snum
      FROM class c, Enrolled e, Faculty f
      WHERE e.cname =c.name AND c.fid = f.fid AND f.fname ='I.Tech')
```

3. Find the names of all classes that either meet in room R128 or have five or more students enrolled.

```
SELECT c.name
FROM Class c
WHERE c.room = 'R128' OR c.name IN
      (SELECT e.cname
      FROM Enrolled e
      GROUP BY e.cname
      HAVING count(*)>=5)
```

4. Find the students those who have not enrolled in any class

```
SELECT DISTINCT s.sname
FROM Student s
Where s.snum NOT IN (SELECT e.enum
                     FROM Enrolled e)
```

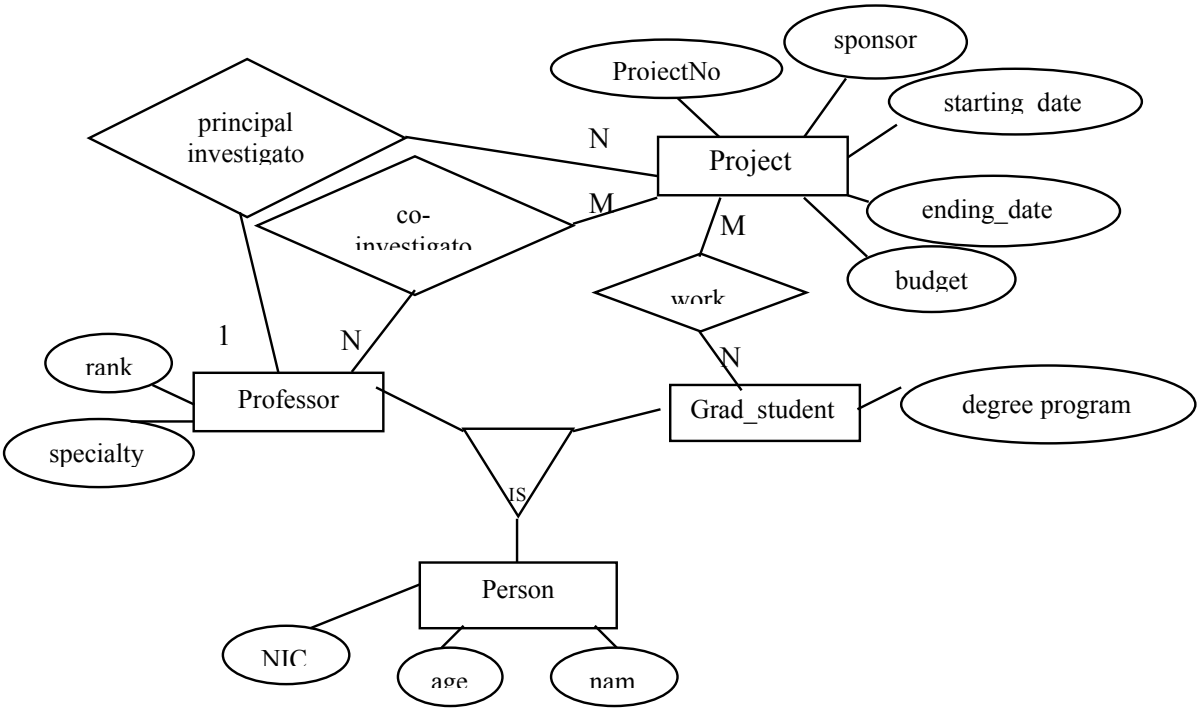
5. Print the Level and the average age of student for that Level, for each level.

```
SELECT s.level, AVG(s.age)
FROM Student s
GROUP BY s.level
```

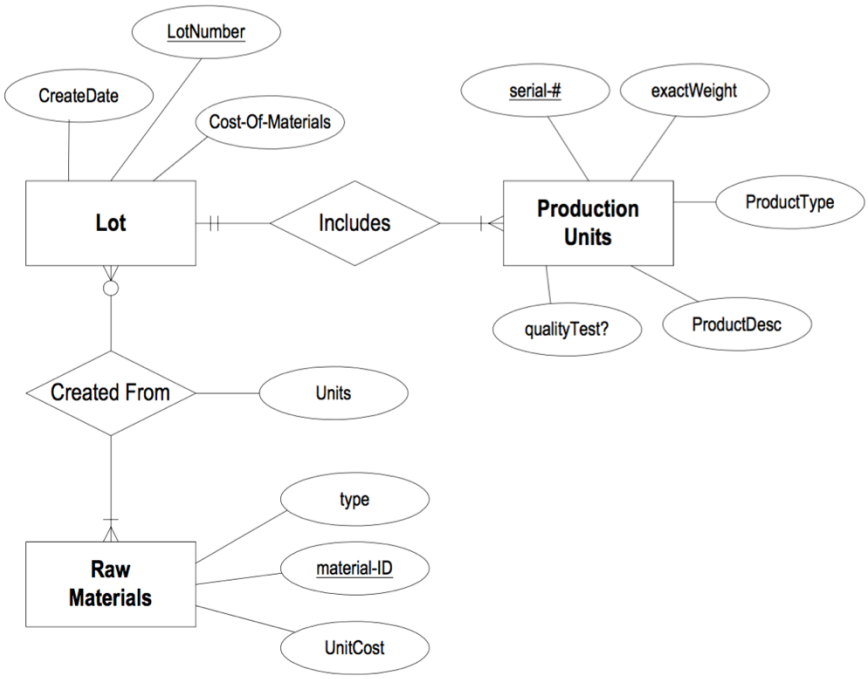
Section C (Designing, Normalisation, Transaction Management and Data Warehouses)

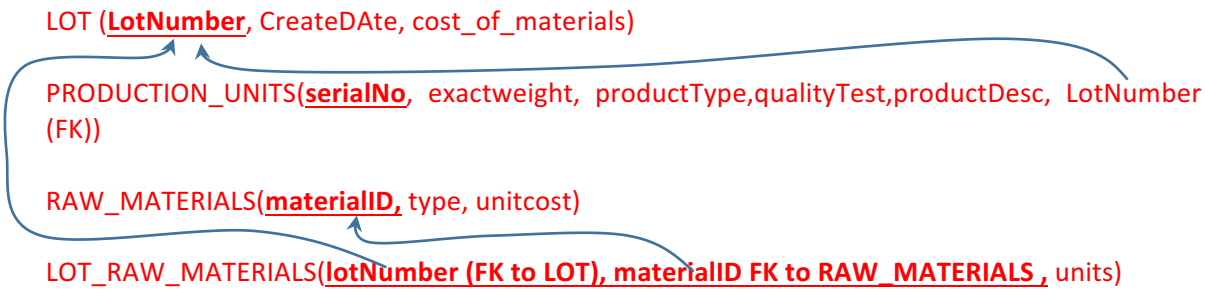
1. Consider the following requirements for a university database:
   - A Person has a NIC (unique), age and a name.
   - Professor is a Person with the following attributes: rank and research specialty.
   - Projects have a project number (unique), a sponsor (e.g. NSF), a starting date, ending date and a budget.
   - Graduate students are also persons. Graduate students need to store information about the degree program (e.g. M.S. or Ph.D.) that they are enrolled.
   - Each project is managed by one professor (known as the project's principal investigator).
   - A project must have a principal investigator.
   - Each project is worked on by one or more professors (known as co-investigators).
   - Professors can manage and/or work on multiple projects.
   - Each project is worked on by one or more graduate students (known as project's research assistants).
   - Graduate students can work on multiple projects.

   Draw the EER diagram for the above requirements.



2. convert the ER diagram into a relational database schema. Be certain to indicate primary keys and referential integrity constraints.





3. Consider following medical records from an animal clinic. Identify dependencies and Normalise them up to BCNF.

HEALTH HISTORY REPORT

PET ID	PET NAME	PET TYPE	PET AGE	OWNER	VISIT DATE	PROCEDURE
246	ROVER	DOG	12	SAM COOK	JAN 13/2002	01 - RABIES VACCINATION
					MAR 27/2002	10 - EXAMINE and TREAT WOUND
					APR 02/2002	05 - HEART WORM TEST
298	SPOT	DOG	2	TERRY KIM	JAN 21/2002	08 - TETANUS VACCINATION
					MAR 10/2002	05 - HEART WORM TEST
341	MORRIS	CAT	4	SAM COOK	JAN 23/2001	01 - RABIES VACCINATION
					JAN 13/2002	01 - RABIES VACCINATION
519	TWEEDY	BIRD	2	TERRY KIM	APR 30/2002	20 - ANNUAL CHECK UP
					APR 30/2002	12 - EYE WASH

Pet (pet_id, pet_name, pet_type, pet_age, owner, { visitdate, procedure_no, procedure_name })
1NF:

Pet (**pet_id**, pet_name, pet_type, pet_age, owner)

Pet_Visit (**pet_id, visitdate**, procedure_no, procedure_name)

note: a procedure may occur on multiple dates, therefore visitdate and procedurte number are included as part of the key

2NF:There arepartial dependancies.

e.g.
Pet_Visit (pet_id => visitdate, procedure_no)
Procedure (procedure_no => procedure_name)

Pet (**pet_id**, pet_name, pet_type, pet_age, owner)

Pet_Visit (**pet_id**, visitdate, procedure_no)

Procedure (**procedure_no**, procedure_name)

3NF: There are no transitive dependancies. Therefore relational schema in 3NF

Pet (**pet_id**, pet_name, pet_type, pet_age, owner)

Pet_Visit (**pet_id**, visitdate, procedure_no)

Procedure (**procedure_no**, procedure_name)

Hence, now the all the relational schemas in 3NF.

4. Name four properties of a transaction and describe each of them with details.
    a. ATOMICITY
    b. CONSISTANCY
    c. ISOLATION
    d. DURABILITY

See the slides for description.

5. Name and explain the replication types of a distributed database.
   Lecture 11 Slide No 7.
6. Write a note on how data and managing data are important for today businesses.
   Today businesses are run by data. Need to explain about automated supply chains, marketing, etc.

   You need to write more lenthy answers for question 4,5 and 6 in part C.