Solutions

Fundamentals of Data Engineering (COE848)

Mid Term Exam

March 2017

This is a CLOSED BOOK exam. Textbooks, notes, laptops, calculators, personal digital assistants, cell phones, and Internet access are NOT allowed.

This is a 120-minute exam.

There are 6 questions with a total of 100 marks.

Please read each question carefully, and write your answers legibly in the space provided. You may do the questions in any order you wish, but please
USE YOUR TIME WISELY.

When you are finished, please hand in your exam paper and make sure you are **signed out**.
Good luck!

Student Name: _____
Student ID: _____
Score: _____ %

Question	Maximum Mark	Received
1	15	
2	20	
3	15	
4	15	
5	15	
6	20	
Total	100	

Question 1: Explain the following terms in 2-3 sentences:

a) Schema

b) DDL

c) Atomicity

d) JDBC

e) Query Optimization

Question 2: Consider the following table definitions:

R(a,b)
T(b,c)

Express R ⋉ T using only the primary relational algebra operators. (Hint: ⋉ is right outer join).

$R \ltimes T = (R \bowtie T) \cup (\{(NULL, NULL)\} \times (T - \pi_{b,c}(R \bowtie T)))$

Question 3:

Given the following table definitions:

Flights(flno, from, to, distance, departs)
Aircraft(aid, aname, range)
Certified(eid, aid)
Employees(eid, ename, salary)

Provide appropriate relational algebra expressions for each of the following queries:

a. Find the name of pilots who are certified on some Boeing aircraft.

$\Pi_{ename}\left(\sigma_{aname='Boeing\%'}\left(Aircraft \bowtie Certified \bowtie Employees\right)\right)$

b. Find the name of the aircraft(s) that can fly non-stop from LA to NY.

$\Pi_{aname}\left(\sigma_{range \geq distance}\left(Aircraft \bowtie Flights\right)\right)$

from = 'LA'

And

to = 'NY'

And

flno = aid

c. Find the flight number of the flights that can be piloted by every pilot whose salary is over $100,000 (consider that distance needs to be less than or equal to range)

$A = \Pi_{flno, eid}(\delta_{range \geq distance}(Aircraft \bowtie Flights) \bowtie Certified$

$B = \Pi_{eid}(\delta_{salary > 100,000}(Employees \bowtie Certified))$

result = A/B

d. Find the name of employee(s) with the highest salary.

$A = \Pi_{eid, ename}(Employees)$

$B = \Pi_{e2.eid, e2.ename}(\delta_{e1.salary > e2.salary}(\rho_{e1}(employees) \bowtie \rho_{e2}(Employees)))$

result = A - B

e. Find the name of employees certified on exactly two aircrafts.

$A = \Pi_{eid, ename}(\delta_{r1.aid \neq r2.aid \text{ and } r1.eid = r2.eid}(\rho_{r1}(certified) \bowtie \rho_{r2}(certified)))$

$B = \Pi_{eid, ename}(\delta_{r1.aid \neq r2.aid \text{ and } r1.aid \neq r3.aid \text{ and } r2.aid \neq r3.aid \text{ and } r1.eid = r2.eid = r3.eid}(\rho_{r1}(Certified) \bowtie \rho_{r2}(Certified) \bowtie \rho_{r3}(Certified)))$

result = A - B

Question 4:

Consider the following relational algebra statements or written queries; write equivalent SQL statements:

serves(bar, drink)
likes(drinker, drink)
frequents(drinker, bar)

a.

$$\pi_{bar}(serves \bowtie \sigma_{drinker=Joe}(likes)).$$

```
Select  Serves.bar
from    Serves, likes
where   serves.drink = likes.drink
  and   likes.drinker = 'Joe';
```

b.

$$\pi_{drinker}(frequents \bowtie serves \bowtie likes)$$

```
Select  frequents.drinker
from    frequents, Serves, likes
where   frequents.bar = Serves.bar
  and   Serves.drink = likes.drink
  and   likes.drinker = frequents.drinker
```

c. Find those drinkers that frequent bars that serve at least one of the drinks they like.

```
Select f.drinker
from   fraquets f, serves s,
       S.bar = f.bar
and
   S.drink   IN ( Select drink
                  from  likes l
                  where
                    f.drinker = l.drinker )
```

d. Find the bars that are frequented only by those drinkers who like all the drinks served by those bars.

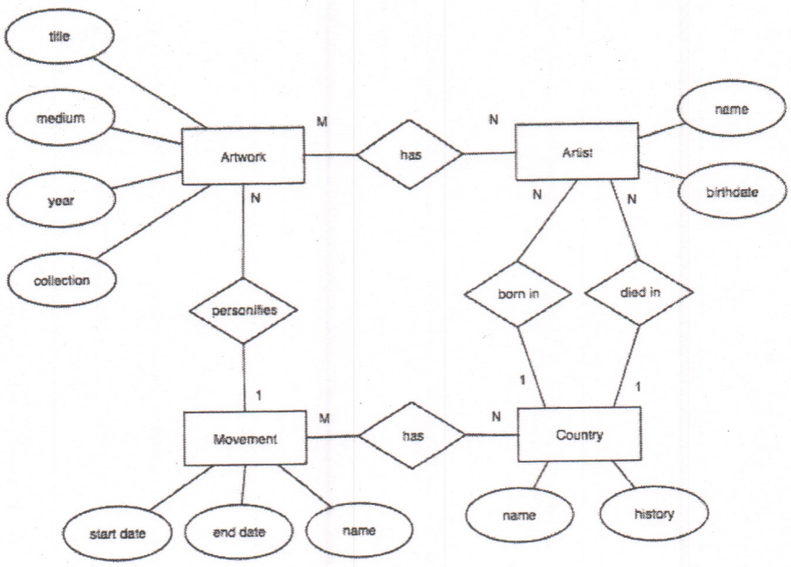
```
Select s3.bar
from serves s3
where s.bar NOT IN (  Select s2.bar
                      from frequts f, likes l, serves s2
                      where l.drinker = f.drinker
                      and s2.bar = f.bar
                      and l.drink <> some( select drink
                                           from serves s
                                           where s.bar = f.bar ))
```

Question 5:

Convert the ER diagram below to a database schema. Indicate the keys for each table in your answer. You do not need to write SQL DDL commands.

Note: For a table T with attributes k and p where k is the primary key, you can use the following notation in your answer:

T (k,p)



has (id1 , id2)

Artwork (id, title, medium, year, collection, personifies)

Artist (name, birthdate, died_in, born_in)

Country (name, history)

Movement (name, end_date, start_date)

Question 6:

Draw a simple ER diagram to capture the following requirements. If you ER model contains any redundancy discuss why.

a. One superhero protects each city
b. Every supervillain operates out of a secret location (not always a city)
c. Some cities have more than one supervillain while some have none
d. Every super-hero has a supervillain nemesis
e. A super-hero and supervillain always fight in some city
f. Teams of super-heroes can fight teams of supervillains