Fundamentals of Data Engineering
(COE848)

Mid Term Exam

March 2016

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

It is a 120 minute exam, 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 1:

Consider the following table definitions:

```
CREATE TABLE RelA (Aid integer, a1 integer, a2 integer, PRIMARY KEY (Aid))
CREATE TABLE RelB (Aid integer, Cid integer, b1 integer,
                   PRIMARY KEY (Aid, Cid, b1),
                   FOREIGN KEY (Aid) REFERENCES RelA,
                   FOREIGN KEY (Cid) REFERENCES RelC)
CREATE TABLE RelC (Cid integer, c1 integer, c2 integer, c3 integer, PRIMARY KEY (Cid))
```

Write a SQL statement that would be equivalent to the following:

π_{c1} ((Temp1 $\cap$ Temp2) $\bowtie$ RelC)

where

Temp1 = π_{Cid} (($\sigma_{a2=q}$ RelA) $\bowtie$ RelB) and
Temp2 = π_{Cid} (($\sigma_{a2=r}$ RelA) $\bowtie$ RelB)

Answers: either answer is fine:

```
SELECT DISTINCT C.c1
FROM RelC C, RelB B1, RelA A1, RelB B2, RelA A2
WHERE C.Cid = B1.Cid AND B1.Aid = A1.Aid AND
      C.Cid = B2.Cid AND B2.Aid = A2.Aid AND
      A1.a2 = q AND A2.a2 = r
```

```
SELECT DISTINCT C.c1
FROM RelA A, RelB B, RelC C
WHERE C.Cid = B.Cid AND B.Aid = A.Aid AND A.a2 = q AND
      C.Cid IN (SELECT C2.Cid
                FROM RelC C2, RelA A2, RelB B2
                WHERE C2.Cid = B2.Cid AND
                      B2.Aid = A2.Aid AND A2.a2 = r )
```

```
SELECT DISTINCT C.c1
FROM RelC C
WHERE C.Cid IN (( SELECT B.Cid
                  FROM RelAA, RelB B
                  WHERE B.Aid = A.Aid AND A.a2 = q)
                INTERSECT
                (( SELECT B2.Cid
                  FROM RelAA2, RelB B2
                  WHERE B2.Aid = A2.Aid AND A2.a2 = r))
```

Question 2:

You have two relations R = [C, F, G] and P = [B, C, D, F, H]. Using only relational algebra select (σ) and cross product (X) operators, express the natural join of R and P.

Answers

$$\sigma_{R.C=P.C \text{ AND } R.F=P.F} (R \times P)$$

OR

$$\sigma_{R.F=P.F} \sigma_{R.C=P.C} (R \times P)$$

Question 3:

Assume the following relational schemas (underlined variables in each each schema make up its primary key).

Customers (*SSN*: integer, *name*: string, *address*: string, *city*: string)

Accounts (*SSN*: integer, *AccntNo*: integer)

Transactions (*AccntNo*: integer, *ProductId*: integer, *date*: string, *quantity*: integer) // *quantity* is the number of ProductId purchased on given transaction

Products (*ProductId*: integer, *ProductName*: string, *cost*: real)

Write SQL queries for the following. You may NOT use nested queries and you may NOT use any JOIN keyword constructs.

a)

Write an SQL query that lists the names of all customers living in Nashville (i.e., city = 'Nashville').

```
SELECT C.Name FROM Customers C WHERE C.city = 'Nashville'
```

b)

Write an SQL query that lists the names and account numbers of all customers (regardless of city).

```
SELECT C.Name, A.AcctNo
FROM Customers C, Accounts A
WHERE C.SSN = A.SSN
```

c)

Compute and list the accounts (as *AccntNo*) and purchases (as *ProductId, ProductName, quantity*) on a given date (*date = X*) of all customers living in Nashville. If a customer bought the same product in more than one transaction on the given day, the query would list these in separate rows of the result.

```
SELECT A.AcctNo, P.ProductId, P.ProductName, T.quantity
FROM Customers C, Accounts A, Transactions T, Products P
WHERE C.city = 'Nashville' AND C.SSN = A.SSN AND A.AcctNo = T.AcctNo AND T.date = X
        AND T.ProductId = P.ProductId
```

d)

List the total number (sum) of all quantities for the product identified by ProductId = X, regardless of AccntNo, that were purchased on date = Y.

```
SELECT SUM(T.quantity)
FROM Transactions T
WHERE T.ProductId = X AND T.date = Y
```

E)

For EACH item, list the item's ProductId, ProductName, and total (sum) of all item quantities purchased by Nashville customers, BUT ONLY for those product's with cost > 50 and having a total purchased quantity of greater than 100.

```
SELECT P.ProductId, P.ProductName, SUM(T.quantity)
FROM Customers C, Accounts A, Transactions T, Products P
WHERE C.city = 'Nashville' AND C.SSN = A.SSN AND A.AcctNo = T.AcctNo AND
        T.ProductId = P.ProductId AND P.cost > 50
GROUP BY P.ProductId, P.ProductName
HAVING SUM(T.quantity) > 100
```

Patient (ssn , name , address , age)

Doctor (ssn , name , specialty , experience)

Primary Physician (patientSSN , doctorSSN , firstAppointment)

Question 4:

Three relational schema are given below. Express each of the provided information needs as a relational algebra query using only projection (π), selection (σ), renaming (ρ), inner joins ($\bowtie$), set union/intersection ($\cup$, $\cap$), and division ($/$).

a)

Find the ssns of doctors who see patients that are at least 25 years old.

$P = \Pi_{ssn}(\sigma_{age \geq 25}(Patient))$

$\Pi_{doctorssn}(P \bowtie_{ssn = PatientssN} PrimaryPhysician)$

b)

Find the ssns of patients that see a doctor that has more than 20 years of experience or less than 2 years of experience.

$InterestingDocs \leftarrow (\sigma_{experience > 20 \text{ or } experience < 2}(Doctor))$

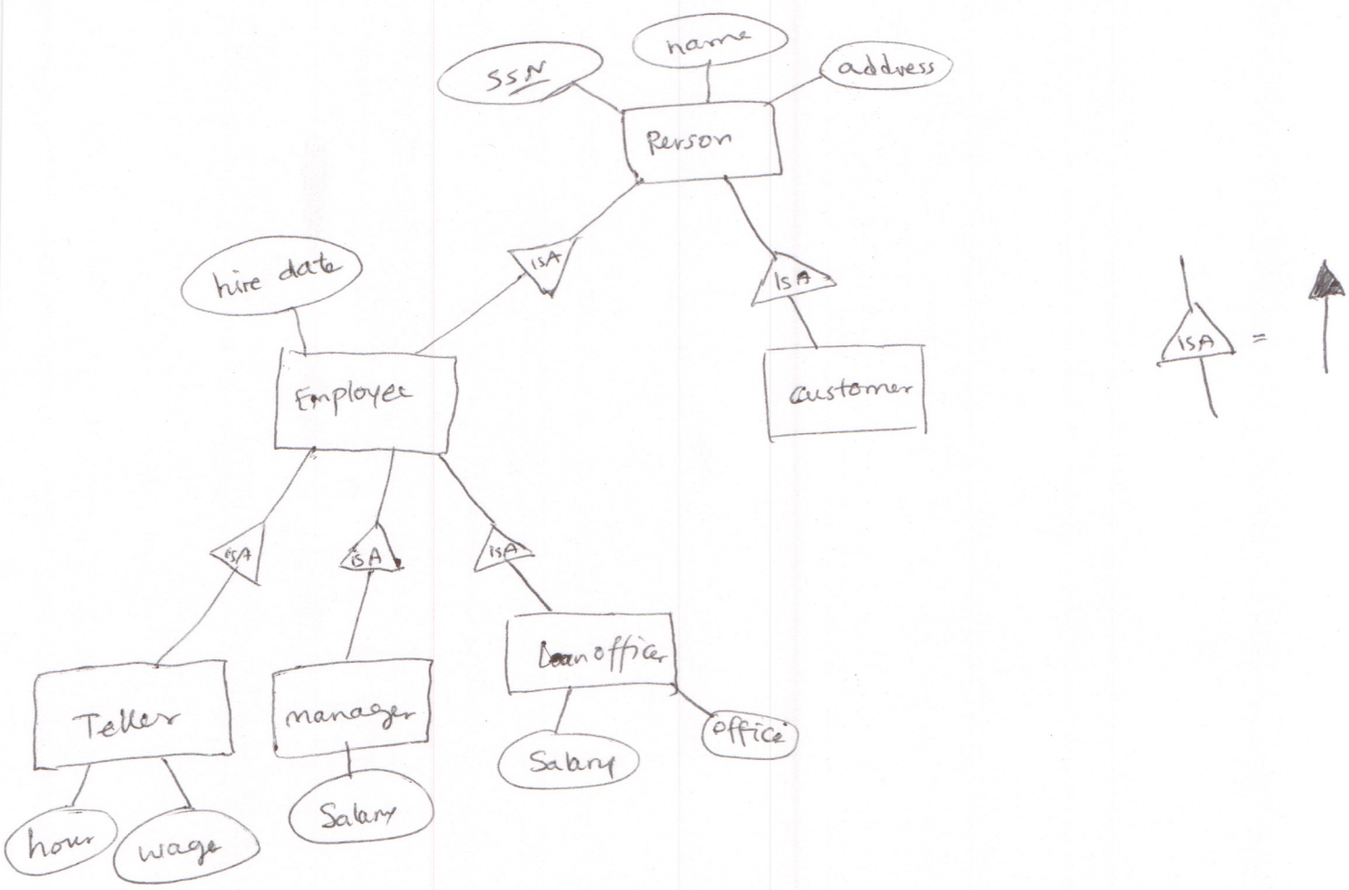
$\Pi_{patientssn}(PrimaryPhysician \bowtie_{doctorSSN = SSN} InterestingDocs)$

Question 5:

You are building a database system for a bank. The bank tells you that it needs to keep track of employees and customers. Every employee and every customer has a (unique) social security number, a name, and an address. Customers also have an account ID, which employees do not have (unless the employee is also a customer). Employees have a hire date, which customers do not have (unless the customer is also an employee). Furthermore, different information is stored about each employee type: tellers work a certain number of hours per week at a certain wage, loan officers work for a salary and have a fixed office location, and managers work for a salary but have no fixed office location.

Draw a simple E-R diagram to capture these requirements. If your E-R model contains any redundancy, describe it.

Any other reasonable diagram would also be fine



Question 6: Explain the following terms in 2-3 sentences:

a) Weak Entity

b) Foreign Key

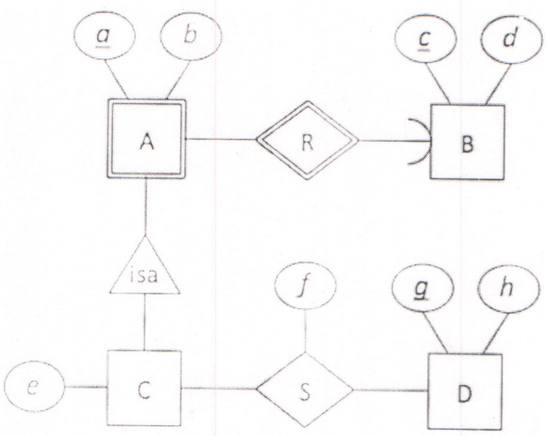
c) Candidate Key

d) ACID

e) Declarative Language

Question 7:

Convert the E/R 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)

Answers:

A (c , a , b)
B (c , d)
C (a , c , e)
D (g , h)
S (a , c , g , f)

Explanation:
Entity A is weak and dependent on B. Hence, the key of A needs to include the key of B.
C is a refinement of A and, hence, needs to include the key of A.