Solutions

Fundamentals of Data Engineering
(COE848)

Final Exam

April 2018

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 7 questions with a total score of 100.

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 | 5 | |
| 3 | 16 | |
| 4 | 14 | |
| 5 | 15 | |
| 6 | 15 | |
| 7 | 20 | |
| Total | 100 | |

Question 1 (15 Marks): Explain the following in 2-3 sentences:

a) Transitive Dependency

$F \rightarrow B$, $B \rightarrow A$ } transitive dependency $\Rightarrow F \rightarrow A$

b) Surrogate Key

a synthetic, often automatically generated, key that uniquely represents values in each row.

c) Advantages of PreparedStatements (mention two)

+ reusability

+ execution Speed

d) Difference between Char and Varchar

char is fixed length character datatype while var char is variable length.

e) Physical data independence

to be possible to change the physical storage of data without affecting the conceptual schema design of the data.

Question 2 (5 Marks): Within the context of database normalization, there is a rule called *pseudo-transitivity*, which is expressed as follows: A→B, BC→D then AC→D.

Based on the functional dependency rules that you already learnt in class this term, show how *pseudo-transitivity* can be derived.

- A → B given;
- BC → D given
- AC → BC Augmentation of 1 with C
- AC → D Transitivity of 3 and 2

Question 3 (20 Marks): Consider the relation schema S=(A,B,C,D) and the following functional dependencies on S:

A→BCD, B→C, CD→A

For each of the following short questions, be sure to briefly explain your answer.

a) S is not in BCNF, but is it in 3NF? Explain your answer.

Yes, because CD and A are candidate keys. B → C is okay because C is part of a candidate key (CD).

b) Consider the decomposition of S into S1 = ABC and S2 = BCD. Is this a valid decomposition into BCNF? Explain.

No, because B → C and B is not a candidate key (for either S1 and S2)

c) Consider the decomposition of S into S3 = ABD and S4 = BC. Is this a valid decomposition into BCNF? Explain.

Yes, A is a key for ABD and B is a key for BC (also all 2 attr relns are in BCNF)

d) Assume the decomposition in part c is implemented and the user is allowed to enter data into tables S3 and S4. Is it guaranteed that all the initial functional dependencies on S (A→BCD, B→C, CD→A) will be respected in S3 and S4?

No, because CD→A (or A→C) are lost ;

Question 4 (20 Marks): Assume you are in charge of managing the program committee for an important conference. The following database stores information about papers submitted to the conference (table Paper), reviewers on the program committee (table Reviewer), and the assignment of reviewers to papers (table Reviews). Each reviewer on the program committee will have to review a set of papers. Each paper will be reviewed by some subset of reviewers.

Paper(pid, title)
Reviewer(rid, name)
Reviews(rid, pid)

a) Write a SQL query that finds all papers with fewer than three reviewers assigned to them. The output of the query should be a list of paper titles. The result should also include papers without any reviewers assigned to them.

```
SELECT P.title
FROM Paper P LEFT OUTER JOIN Reviews X ON P.pid = X.pid
GROUP BY P.pid, P.title
HAVING count(*) < 3
```

b) Write a SQL query that finds the reviewers with the most papers assigned to them. There can be more than one such reviewer. The output of the query should be a list of reviewer names. A reviewer should be listed if no other reviewer has strictly more papers to review.

```
SELECT R.name
FROM Reviewer R, Reviews X1
WHERE R.rid = X1.rid
GROUP BY R.rid, R.name
HAVING count(*) >= ALL ( SELECT count(*)
                         FROM Reviews X2
                         GROUP BY X2.rid )
```

Question 5 (Mark 15): Refer to the following schema design and answer the queries using relational algebra:

Country (CName, Continent, GDP, Population, LitRate, LifeExp)

Politician (Pname, Gender, Age, Office, Country, Region)
Head_Of_Government (Country, Office)
Alliance_Membership (Country, Alliance)

a) List the country in the database where all of the senators (office=senator) are women.

result $\leftarrow \Pi_{CName}$ (Country) $- \Pi_{CName}(\sigma_{Gender = "Male" \wedge Office = "Senator"}$ Country $\bowtie$ Politician)

b) List the country in the database where all of the senators are women but the head of government is not a woman.

all_female_senators $\leftarrow \Pi_{CName}$ (Country) $- \Pi_{CName}(\sigma_{Gender = "Male" \wedge Office = "Senator"}$ Country $\bowtie$ Politician)

not_female_head $\leftarrow \Pi_{CName}(\sigma_{Gender = "Male"}$ Country $\bowtie$ Politician $\bowtie$ Head_Of_Government)

result $\leftarrow$ all_female_senators $\wedge$ not_female_head

c) List the countries in the database that are in NATO (Alliance=Nato) but do not have a female head of government.

NatoCountries $\leftarrow \Pi_{CName}(\sigma_{Alliance = "NATO"}$ Country $\bowtie$ Alliance_Membership)

not_female_head $\leftarrow \Pi_{CName}(\sigma_{Gender = "Male"}$ Country $\bowtie$ Politician $\bowtie$ Head_Of_Government)

result $\leftarrow$ NatoCountries $\wedge$ not_female_head

Question 6 (15 Marks): Consider the following DTD Document.

```
<!DOCTYPE Hogwarts [

<!ELEMENT Hogwarts (Rooms, Teachers, Courses) >

<!ELEMENT Rooms (Room*) >
 <!ELEMENT Room EMPTY >
  <!ATTLIST Room
    name     ID     #REQUIRED
    nrSeats CDATA #IMPLIED >

<!ELEMENT Teachers (Teacher*) >
 <!ELEMENT Teacher (Title*) >
  <!ELEMENT Title (#PCDATA) >
  <!ATTLIST Teacher
    name ID     #REQUIRED
    room IDREF #REQUIRED >

<!ELEMENT Courses (Course*) >
 <!ELEMENT Course (Class*) >
  <!ATTLIST Course
    name        ID     #REQUIRED
    teacher     IDREF #REQUIRED
    nrStudents CDATA #IMPLIED >
  <!ELEMENT Class EMPTY >
   <!ATTLIST Class
     day     CDATA #REQUIRED
     hour    CDATA #REQUIRED >
]>
```

a) Write an XPath expression to find all courses that have at least 20 students:

```
//Course[@nrStudents >= 20]
```

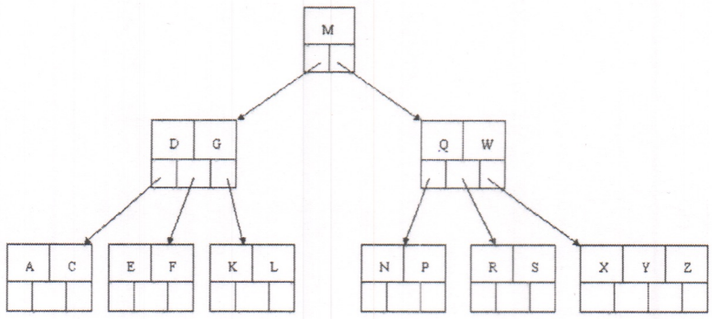
b) Write an XPath expression to list all professors at the school:

```
//Teacher[Title = "Professor"]
```

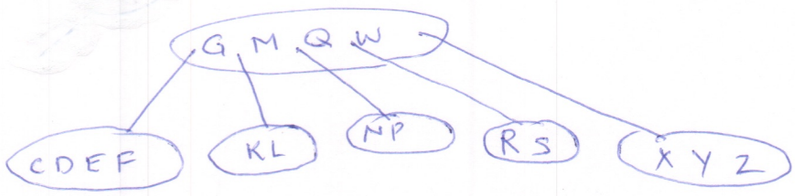
c) Find all rooms that are used on Mondays

```
//Room[@name = //Teacher[@name = //Course[Class/@day = "Monday"]/@teacher]/@room]
```

Question 7 (20 Marks): Consider the following 5-way B-Tree:



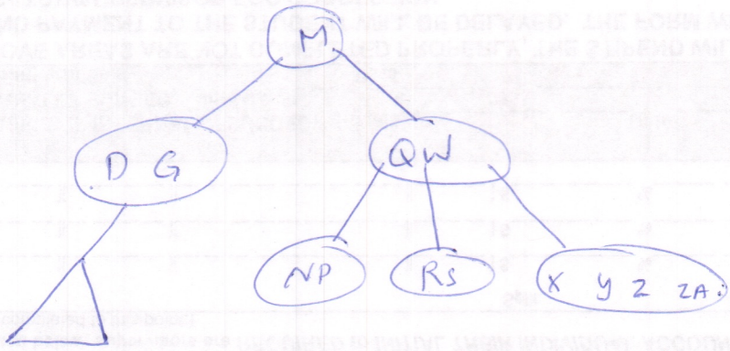

a) Delete 'A' from the tree. Show your steps.




b) Delete 'Q' from the original tree. Show your steps.


c) Insert 'ZA' into the original tree and then 'ZB' into the new tree. Show your steps.

ZA




ZB