Solutions

Fundamentals of Data Engineering (COE848)

Final Exam

April 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 8 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	10	
2	10	
3	20	
4	15	
5	10	
6	15	
7	5	
8	15	
Total	100	

Question 1 (10 Marks): Explain the following in 2-3 sentences:

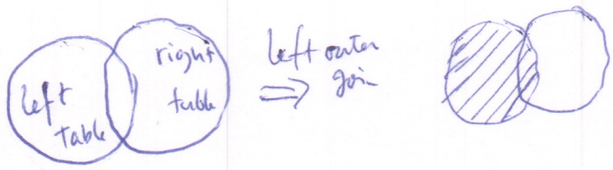
a) Database Normalization

the process to organize tables in order to reduce redundancy & improve data integrity

b) Difference between projection and selection

projection manipulates & determines which columns are returned while selection determines which rows are returned.

c) Left outer join



d) N-way tree

a tree whose nodes have n-1 keys & n pointers.

e) ISAM

indexed sequential access method used for indexing & maintaining indices.

Question 2 (10 Marks): Explain what are the pros and cons of normalizing a database into higher degree normal forms?

+ maintainability
+ better structure

+ redundancy
+ duplicate information

Question 3 (20 Marks): You decide to create a database to track the songs your favorite band plays in its live concerts. To this end, you create a relational schema directly for your database. After much consideration, you believe that a single schema will be sufficient:

Concerts (City, Venue, Year, Month, Date, Song, Album)

However, after using the database for a few months, you realize that your band (and the real world) have some characteristics that you should model in your database. Convert each of the next four sentences about Concerts into a functional or a multi-valued dependency. You can use the first letter of each attribute as an abbreviation for the attribute. Consider each of these four sentences independently. If you cannot write down a functional or a multi-valued dependency, say so, and explain why you cannot, if possible. Do not assume any other constraints, even if they seem reasonable to you.

a) Each song appears in at most one album. In other words, the band does not repeat the same song in different albums.

Song → Album

b) A city does not have two venues with the same name. In other words, City and Venue serve to identify the location of a concert uniquely.

Its not possible to state a functional dependency.

c) In an effort to please its fans, the band plays at most one song from any album in a given concert.

city, venue, year, month, date, album → song

d) The manager books the band in any city at most once every year.

city year → venue, month, date

Question 4 (15 Marks): Express each of the provided information needs as a relational algebra query using only projection (π), selection (σ), renaming (ρ), joins ($\bowtie$), set union/intersection ($\cup$, $\cap$) and division (/).

Student(ssn, name, address, major)
Course(code, title)
Registered(ssn,code)

a) SSNs of students who are registered for both 'Database Systems' and 'Analysis of Algorithms'.

Π_{ssn} (Student $\bowtie$ Registered $\bowtie$ ($\delta_{title='Database\ Systems'}$ Course)) $\cap$

Π_{ssn} (Student $\bowtie$ Registered $\bowtie$ ($\delta_{title='Analysis\ of\ Algorithms'}$ Course))

b) The titles of courses for which no student is registered.

Π_{title} ((Π_{code} (Course) $-$ Π_{code} (Registered)) $\bowtie$ Course)

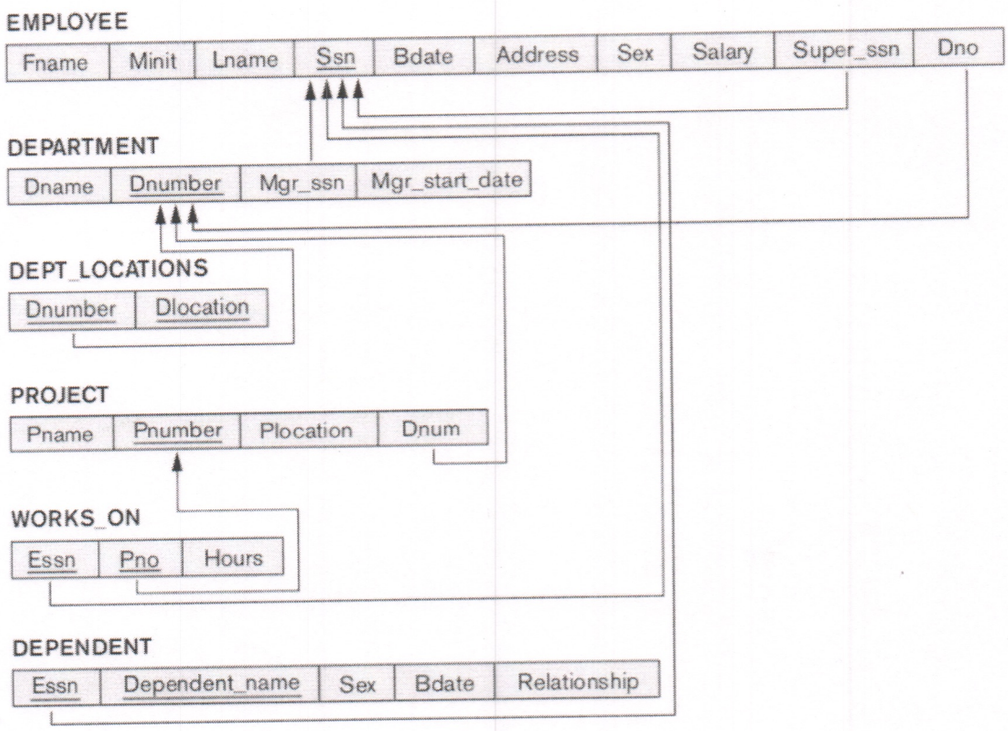
c) Names of students and the titles of courses they registered to.

$\Pi_{name,\ title}$ (Student $\bowtie$ registered $\bowtie$ Course)

d) List of courses in which all 'ECMP' major students are registered.

$\Pi_{Code,ssn}(Registered) / \Pi_{SSN}(\sigma_{major='ECMP'}(Student))$

Question 5 (Mark 10): Refer to the Company database shown in the figure to answer the following two questions:



a) Write a SQL statement to remove all locations for the "Special Circumstances" department from the database.

```
delete from Dept_Locations
where Dnumber in ( select Dnumber
                   from Department
                   where Dname = 'Special_Circumstances' )
```

b) Write the SQL statement to update the database so that all projects in all departments managed by "Oliver Perks" are moved to "Arlen".

```
update Project Set Plocation = 'Arlen'
where Dnum in ( Select Dnumber
                from Department d, employee e
                where
                Fname = 'oliver' and Lname = 'Perks'
                and d.mgr_ssn = e.Ssn
```

Question 6 (15 Marks): Consider the following DTD for a TV Schedule.

```
<!DOCTYPE TVSCHEDULE [
<!ELEMENT TVSCHEDULE (CHANNEL+)>
<!ELEMENT CHANNEL (BANNER,DAY+)>
<!ELEMENT BANNER (#PCDATA)>
<!ELEMENT DAY (DATE,(HOLIDAY|PROGRAMSLOT+)+)>
<!ELEMENT HOLIDAY (#PCDATA)>
<!ELEMENT DATE (#PCDATA)>
<!ELEMENT PROGRAMSLOT (TIME,TITLE,DESCRIPTION?)>
<!ELEMENT TIME (#PCDATA)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT DESCRIPTION (#PCDATA)>

<!ATTLIST TVSCHEDULE NAME CDATA #REQUIRED>
<!ATTLIST CHANNEL CHAN CDATA #REQUIRED>
<!ATTLIST PROGRAMSLOT VTR CDATA #IMPLIED>
<!ATTLIST TITLE RATING CDATA #IMPLIED>
<!ATTLIST TITLE LANGUAGE CDATA #IMPLIED>
]>
```

For the following problems, assume you have an XML document that conforms to the above DTD and answer the question:

a) Write an XPath expression to find the time of all program slots that are in English.

//PROGRAMSLOT[/TITLE/@LANGUAGE='ENGLISH']/TIME

b) Write an XPath expression to find the rating of "Shawshank Redemption"

//PROGRAMSLOT[TITLE='Shawshank Redemption']/TITLE/@RATING

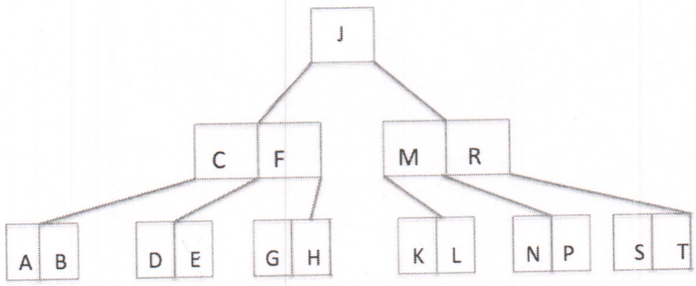
c) Write an XPath expression to find the title of programs.

//TITLE

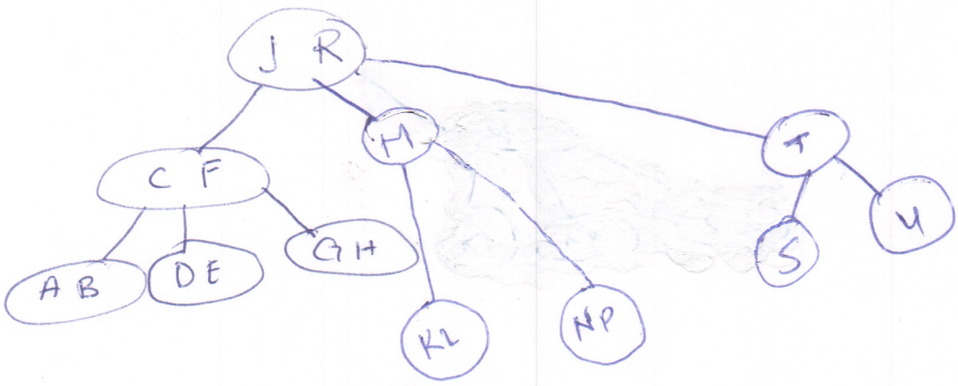
Question 7 (5 Marks): Consider the various cardinality constraints available in ER Diagrams including, many-to-many, many-to-one, one-to-one, and ternary relations, it is impossible to enforce a mandatory participation in which of these cardinality contraints in a relational schema? Why?

many to many.

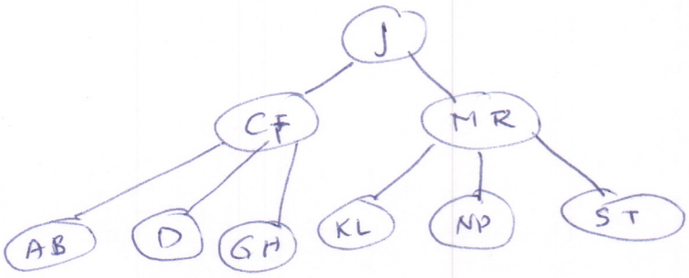
Question 8 (15 Marks): Consider the following 3-way B-Tree:



a) Insert 'U' into the tree. Show your steps.



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



c) Delete 'J' from the original tree. Show your steps.