M2 Solutions

Ques 1. **[2 Marks]** Circle the correct answers

1. Normalization is not for
   a) reducing the number of tables in the schema.
   b) eliminating uncontrolled redundancy of data.
   c) eliminating anomalies that could otherwise occur with insert, delete or update.
   d) ensuring that functional dependencies are enforced.

2.If a non-key attribute of a relation is null, then that table automatically violates which normal form
   a) NONE
   b) 1NF
   c) 2NF
   d) 3NF
   e) BCNF

3. A table is in 2NF if the table is in 1NF and what other condition is met?
   a) There are no functional dependencies.
   b) There are no null values in primary key fields.
   c) There are no repeating tuples.
   d) There are no attributes that are not fully functionally dependent on the relation's primary key.

4. Which is true?
   a) A superkey is a key.
   b) Two tuples can have the same key value.
   c) A relation can have more than one candidate key.
   d) None of the above

Normalization, 3NF, BCNF

Ques 2. [4 Marks] Consider the relation shown

A	B	C
A1	B1	C1
A1	B1	C2
A2	B1	C1
A2	B1	C3

a) List all the functional dependencies that the given relation instance satisfies.	C→B A→B AC→B
b) Assume that the value of attribute C of the last record in the relation is changed from C3 to C2. Now list all the functional dependencies that this relation instance satisfies.	Same as part (a). Functional dependency set is unchanged.

Ques 3. [2 Marks] Consider a relation R = (A, B, C, D, E).
R is decomposed it into R1 = (A, B, C), R2 = (A, D, E). The set of functional dependencies is:
A → BC
CD → E
B → D
E → A.
Show that this decomposition is a lossless-join decomposition.
R1 ∩ R2 = A;
-A is the common attribute between the two relations, thus we want to show that A is a superkey of either R1 or R2.
(A → BC)
(A → ABC)
(R1 ∩ R2 → R1)
this is a lossless-join decomposition

Ques 4. **[6 Marks]** Consider the below instance of the CLASS relation. Determine whether each of the following statements is true or false; briefly justify your answer.

CID	CName	Credit	SID	SName	Grade
C1	Math	4	S1	Jimmy	Pass
C1	Math	4	S2	Sean	Pass
C2	AI	4	S3	Mary	Pass
C3	DB	3	S4	Peter	Fail
C3	DB	3	S5	Lila	Pass
C3	DB	3	S6	Tony	Pass
C4	SE	3	S1	Jimmy	Pass
C4	SE	3	S2	Sean	Fail
C5	DM	3	S7	Boi	Pass

Statements	T/F?	Brief Justification
a) If class (*C5, DM)* is deleted then the information about student *S7* would also be removed because of deletion anomaly.	T	There is only one instance of C5,DM and for student S7. If this row is removed we also lose the information for student S7 since this is the only instance.
b) Inserting CLASS (C6, Calculus, 3) is an example of insertion anomaly.	T	SID is part of primary key. Hence it must be inserted and cannot be null

Ques 5. **[6 Marks]** Consider the following two schemas

Schema 1: R (A,B,C,D)

Schema 2: R1 (A,B,C) and R2 (B,D)

Determine whether each of the following statement is true or false; briefly justify your answer.

Statements	T/F?	Brief Justification
a) Consider Schema 1 and suppose that the only functional dependencies that hold on the relations in this schema are A→B and C→D. Is Schema 1 in Boyce-Codd Normal Form (BCNF)?	F	Neither A nor C are keys, so A → B and C → D are BCNF violations
b) Consider Schema 2 and suppose that the only functional dependencies that hold on the relations in this schema are A→B, A→ C, B→ A, A→D and all possible implicit FDs. Is Schema 2 in BCNF?	T	A+ = {A,B,C,D} B+ = {A,B,C,D} Both A and B are superkeys for R1 and R2
c) If A→D is deleted from part (b). Is Schema 2 in BCNF?	T	Removing A->D, A still remains the Superkey for R1. R2 by itself is in Primary key because it has 2 entities and trivial FD.

Ques 6. **[15 Marks]** Consider the following relation R(A, B, C, D) and the functional dependencies that hold over this relation.

F = A → B, D
C, D → B
C → D
B → D

a) Determine all candidate keys of R.

The two attributes that don't appear on the RHS of any FD are A and C
A^+ → ABD - Not a key
C^+ → CD - Not a key
AC^+ → ACBD - This closure gives us all the attributes and is a key.
If we add any other attributes to AC, it will be a superkey and AC is therefore the only key of R.

b) Compute the closure of {C, B}

CB^+ → CB - Trivial
CB^+ → CBD - From the third or fourth FD

Answer:
BCD

c) Compute the minimal cover

Step 1 - Split the FDs to only have 1 attribute on the RHS:
A → B
A → D
CD → B
C → D
B → D

Step 2 - Try to reduce the LHS of every FD:
We can only do this for CD → B
CD^+ → CDB
D^+ → D
C^+ → CDB

We can remove D from CD → B making the new set:
A → B
A → D
C → B
C → D
B → D

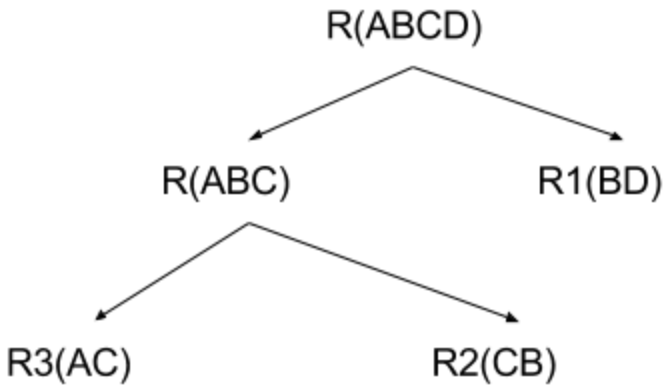
Step 3 - Remove redundant FDs

FD	Closure with FD	Closure w/o FD	Result
$A \rightarrow B$	$A^+ \rightarrow ABD$	$A^+ \rightarrow AD$	Different: Keep FD
$A \rightarrow D$	$A^+ \rightarrow ADB$	$A^+ \rightarrow ABD$	Same: Remove FD
$C \rightarrow B$	$C^+ \rightarrow CBD$	$C^+ \rightarrow CD$	Different: Keep FD
$C \rightarrow D$	$C^+ \rightarrow CDB$	$C^+ \rightarrow CBD$	Same: Remove FD
$B \rightarrow D$	$B^+ \rightarrow BD$	$B^+ \rightarrow B$	Different: Keep FD

Minimal Cover

$A \rightarrow B$
$C \rightarrow B$
$B \rightarrow D$

d) Is R in 3NF? If not, then decompose it. Show all the steps. If you are decomposing R then determine the candidate keys of each decomposed relations.



We add the missed FD A → B as R4(AB) and get the final scheme:
R1(BD), R2(CB), R3(AC), R4(AB)
Candidate keys are R1 : {B}, R2 : {C}, R3 : {AC}, R4{A}

SQL

Ques 7. **[6 Marks]** Consider the below relations where primary keys are underline and foreign key is bold.
employee (ssn,name, address, phone)
*project(pno, pname,**essn**)*
Determine whether each of the following statement is true or false; provide a brief justification of your choice. State any assumptions that you make.

Statements	T/F?	Brief Justification
a) In SQL, the query *"SELECT name FROM employee where ssn>10"* will not return duplicates.	F	

b) Consider below queries: *Query 1: SELECT COUNT (*) FROM employee* *Query 2: SELECT COUNT(name) FROM* employee Query1 and Query2 always return the same result.	F	Explanation: They will return a different result if employee contains tuples with null values for name, since the first query will count those tuples, while the second will not.
c) SELECT e.name FROM employee e, project p WHERE e.ssn = p.essn is equivalent to SELECT name FROM employee	F	

Ques 8. **[5 Marks]** Give the SQL command needed to create the Item relation. You may choose the reasonable domains for your attributes. You must specify the following:

- ItemNo is the primary key
- {IName, Quantity} is a candidate key
- IName and Quantity are not allowed to contain null values
- Price must be greater than 100.

```
create table Item (
ItemNo varchar(16) primary key,
IName varchar(50) not null,
quantity varchar(200) not null,
price float,
unique (name, quantity),
check (price > 100))
```

Ques 9. **[6 Marks]** Consider the following relations (Keys are underlined and foreign keys are bold)
Author (Name, Address, Gender, Research_area)
Journal (JName, Paper_title, Published_year, **AName**)

Write SQL statements that correspond to the following relational algebra expressions.

$\pi_{JName,\ Paper\_title}\ (\sigma_{Published\_year=2016}(Journal))$
SELECT DISTINCT journal name, paper title FROM Paper where year = 2016;

$\pi_{JName}\ (\sigma_{Aname=name\ AND\ published\_year=2016}$ (Author X Journal))
SELECT journal name FROM author, Journal WHERE author name = name AND published year = 2016

Ques 10. **[24 Marks]** Consider the Company database schema and instances given in Appendix A. Use SQL to answer the following questions. Please make sure that your queries do not include any extra information.

a) For each project, retrieve the project number, the project name, and the number of employees from department 5 who work on the project.
select PNum, PName, Count(*) from Project, Works_On, Employee where Pnum=PNo and ssn=essn and DNum=5 group by PNum, PName;

b) Retrieve the name of all employees who have two or more dependents. **(Use nested query**)

```
select Name
from   Employee
where  (select count(*) from Dependent where ssn=essn) >=2;
```

c) Create a view that has employee name, supervisor SSN, and employee salary for each employee who works in the 'Research' department

```
create view RESEARCH_EMP AS
 select E.NAME, E.SuperSSN, E.SALARY
 from EMPLOYEE  E,  DEPARTMENT D
 where D.DName = 'Research' and E.DNum = D.DNum
```

d) Retrieve all employees whose address is in Houston, Texas

```
select *
from   Employee
where  Address like '%Houston, TX%';
```

e) Retrieve the list of all project numbers for projects that involve an employee whose name is 'John Smith', either as a worker or as a manager of the department that controls the project.

```
(SELECT  PNum
         FROM     PROJECT P, DEPARTMENT D, EMPLOYEE E
         WHERE    P.DNum=D.DNum AND D.MGRSSN=E.SSN AND E.NAME = 'John Smith')
UNION
(SELECT   PNum
 FROM     PROJECT P, WORKS_ON W, EMPLOYEE E
 WHERE    P.PNum=W.PNo AND W.ESSN=E.SSN AND NAME = 'John Smith')
```

f) Retrieve the name of all employees who works on all projects controlled by department number 5.

```
SELECT Name
FROM Employee
WHERE NOT EXISTS
 (SELECT PNum
 FROM Project
 WHERE PNum IN (SELECT PNum
 FROM Project
 WHERE DNum = 5)
 AND PNum NOT IN (SELECT DNo
 FROM Works_On
 WHERE essn = ssn));

OR
Select Name
from Employee
where not exists
((select PNum from Project where DNum = 5)
except
(select PNo from Works_On where ssn = essn))
```

Datalog

Ques 11. **[12 Marks]** Consider the following relations (Primary keys are underline and foreign keys are bold):

Paper(title, citation, **AName**,year,**JName**)
Journal(JName, year, PName, **AName**)
Author (AName, RArea)
Where Aname is author name, PName is publisher name and RArea is research area.

Write Datalog to answer the following queries.

a) List the journal name and author name of all journals in the database.
Ans(X,Y) :- Journal(X,_,_,Y)

b) List the name of journals in which papers are published by authors having research area "Databases"
Ans (X) :- Journal(X,_,_,Y), Author(Y,"Databases") OR

c)	List the citation of all paper authored in 2016 or authored by Hazra Imran
	Ans(R) :- Paper (_,C,_,2016) Ans(R) :- Paper (_,C,'Hazra Imran',_,_)

d)	Find the names of authors who have authored paper for all journals.
	Good (A) <- Paper(_,_,A,_,_), ¬Bad(A) **Bad(A) <- Journal(_,_,_,_,J), Paper(_,_,A,_,_), ¬Witness(J,A)** **Witness(J,A) <- Paper(_,_,A,_J)** (Name of the author (A) who authored paper in the Journal (J))