In the SQL questions, you are welcome to use views. Comments are not required, although they may help us mark your answers.

There is a blank page at the end for rough work.

A mark of at least 40 out of 100 on this exam is required in order to pass the course.

It’s been a real pleasure teaching you this term. Good luck!
Question 1. [13 marks]
Consider the following schema for a hair salon.

Relations

- Clients(CID, name, phone).
  *CID is the ID of a client, name and phone are their name and phone number.

- Staff(SID, name).
  *SID is the ID of a staff member and name is their name.

- Appointments(CID, date, time, service, SID)
  *CID is the ID of the client whose appointment it is, date and time indicate when the appointment happens, service is the name of the service they have at this appointment, and SID is the ID of the staff member providing the service for this appointment.

Integrity Constraints

- Appointments[CID] ⊆ Clients[CID].

- Appointments[SID] ⊆ Staff[SID].

Recall that no nulls are permitted in the relational model.

Part (a) [5 marks]
Create an ER diagram that represents this data well.
Part (b) [8 marks]
Which of the following statements are enforced by the relational schema we provided? Circle one answer for each. If the statement is enforced, say what part of the schema enforces it. If it is not enforced, write an integrity constraint in relational algebra that would enforce it. Marks will only be given if your explanation is correct.

1. A client can’t have a haircut and a manicure at the same date and time.
   
   Enforced This part of the schema enforces it:
   
   Not enforced This new integrity constraint would enforce it:
   
2. The same phone number can’t be associated with two different clients.
   
   Enforced This part of the schema enforces it:
   
   Not enforced This new integrity constraint would enforce it:
   
3. A client can’t have two phone numbers.
   
   Enforced This part of the schema enforces it:
   
   Not enforced This new integrity constraint would enforce it:
   
4. Every client has at least one appointment.
   
   Enforced This part of the schema enforces it:
   
   Not enforced This new integrity constraint would enforce it:
Question 2.  [12 marks]

Below is a slightly simplified version of a schema we used in lecture. A course’s cName attribute gives its full name, such as “Introduction to Databases”, while dept has a value such as “CSC” and cNum has a value such as 343.

- Student(sID, surName, firstName)
- Course(dept, cNum, cName)
- Offering(oID, dept, cNum, term, instructor)
- Took(sID, oID, grade)

Offering[dept, cNum] ⊆ Course[dept, cNum]
Took[sID] ⊆ Student[sID]
Took[oID] ⊆ Offering[oID]

Write a query in relational algebra that finds the SID and surname of each student who has a passing grade in every first-year course ever offered by department ‘BIO’, but never took a course offered by department ‘CHE’. A first-year course is one whose cNum is between 100 and 199 inclusive. A passing grade is a grade over 50. Use only the basic operators Π, σ, ▷◁, ×, ∩, ∪, −, ρ, :=. Continue your answer on the next page if needed.

Hint: As a first step, write the left-hand sides of a series of assignment statements; use good names for the intermediate relations. This will help you solve the question, and also to earn part marks if you make a mistake.
Continue your answer here if needed.
Question 3. [13 marks]

Part (a) [4 marks]

Here is the schema for table Took expressed in SQL:

```sql
create table Took(
    sID integer,
    oID integer,
    grade integer,
    primary key (sID, oID),
    foreign key (sID) references Student,
    foreign key (oID) references Offering);
```

For each pair of SQL queries below, circle one answer to indicate whether the queries return the same result on all database instances (“equivalent”), return same result on this one instance (from Part (b) on next page), but not on all instances (“same on this instance”), or neither of the above (“neither”).

1. A) select sid from took;

   B) select t1.sid as sid from took t1, took t2;

   Circle one: equivalent same on this instance neither

2. A) select avg(grade) from took;

   B) select avg(t1.grade) from took t1, took t2;

   Circle one: equivalent same on this instance neither

3. A) select distinct sid from took;

   B) (select sid from took) union (select sid from took where grade = 90);

   Circle one: equivalent same on this instance neither

4. A) select distinct sid from took;

   B) select distinct oid as sid from took where grade > 86;

   Circle one: equivalent same on this instance neither
Part (b)  [9 marks]
Consider the following valid instance of Took. (Assume the foreign key relationships are satisfied by appropriate rows in the other tables.)

<table>
<thead>
<tr>
<th>sID</th>
<th>oID</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>85</td>
</tr>
</tbody>
</table>

For each SQL query below, circle one answer to indicate whether or not it is legal. If legal, show the output. If not, explain the error. Your output or explanation must be correct to receive any marks.

1. select *
   from took
   group by sid;
   Circle one:   Legal   Illegal
   Output or explanation:

2. select min(grade), count(*), count(distinct oID), sid
   from took
   group by sid;
   Circle one:   Legal   Illegal
   Output or explanation:
3. select *
   from took t1
   where not exists (  
       select * from took t2
       where t2.oid = t1.oid and t1.sid <> t2.sid and t2.grade >= t1.grade);  
Circle one:  Legal    Illegal  
Output or explanation: 

4. (select sid from took) except all (select sid from took where grade < 90);  
Circle one:  Legal    Illegal  
Output or explanation: 

5. select *
   from took t1 left join took t2 on t1.grade> t2.grade and t1.sid <> t2.sid
   order by t1.sid, t1.oid;  
Circle one:  Legal    Illegal  
Output or explanation:
Question 4. [6 marks]
Consider the following schema. Primary keys are underlined.

Sailor(SID, Sname)
Reserve(SID, BID, Date) Reserve[SID] ⊆ Sailor[SID] Reserve[BID] ⊆ Boat[BID]
Boat(BID, Color)

The following syntactically correct queries in SQL and relational algebra attempt to return the set (i.e., with no duplicates) of all SIDs of all Sailors who have never reserved a red boat. For each, circle one answer to indicate whether it is correct or incorrect. Note that the symbol <> means “not equal to” or ≠.

1. select S.SID
   from Sailor S
   where S.SID not in (select R.SID
                          from Reserve R, Boat B
                          where B.Color = 'red')

   Circle one: correct incorrect

2. select distinct S.SID
   from Sailor S, Reserve R, Boat B
   where S.SID = R.SID and R.BID = B.BID and B.Color <> 'red'

   Circle one: correct incorrect

3. select S.SID
   from Sailor S
   where not exists (select R.SID
                    from Reserve R, Boat B
                    where S.SID = R.SID and R.BID = B.BID and B.Color = 'red')

   Circle one: correct incorrect

4. select distinct R.SID
   from Reserve R, Boat B
   where R.BID = B.BID and B.color = 'red'
   group by R.SID
   having count(*) < 1

   Circle one: correct incorrect

5. Π_{SID}(Sailor) − Π_{SID}(σ_{Color='red'}(Reserve ⊙ Boat))

   Circle one: correct incorrect

6. Π_{SID,BID}(Sailor ⊙ Reserve) − Π_{SID,BID}(σ_{Color='red'}(Sailor ⊙ Reserve ⊙ Boat))

   Circle one: correct incorrect
Question 5. [8 marks]

Consider the following SQL definitions:

```sql
create table Treatment (  
    PhysicianID int not null,  
    PatientID int not null,  
    Date date not null,  
    Diagnosis varchar(30),  
    Location char(5) references Location  
        on delete cascade on update cascade,  
    primary key(PhysicianID,PatientID,Date),  
    check (PatientID <> PhysicianID) );
```

```sql
create table Location (  
    id char(5) primary key,  
    head int );
```

Which of the following are valid (legal) instances of table Treatment that satisfy all constraints. If illegal, circle the problem data and write a short explanation of the problem.

Assume that table Location contains three tuples \{ (Yonge, 1), (Bloor, 2), (Maine, 1) \}.

1. Circle one: Legal Illegal

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1/1/02</td>
<td>Chicken Pox</td>
<td>Bloor</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1/1/02</td>
<td>Measles</td>
<td>Bloor</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1/1/02</td>
<td>Chicken Pox</td>
<td>Bloor</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1/1/02</td>
<td>Null</td>
<td>Bloor</td>
</tr>
</tbody>
</table>

2. Circle one: Legal Illegal

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>1/1/02</td>
<td>Null</td>
<td>Bloor</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1/1/02</td>
<td>Measles</td>
<td>Yonge</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1/1/02</td>
<td>Chicken Pox</td>
<td>Null</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1/1/02</td>
<td>Null</td>
<td>Null</td>
</tr>
</tbody>
</table>

3. Circle one: Legal Illegal

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3/3/02</td>
<td>Chicken Pox</td>
<td>Yonge</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1/1/02</td>
<td>Flu</td>
<td>Null</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2/2/02</td>
<td>Measles</td>
<td>Yonge</td>
</tr>
</tbody>
</table>

4. Circle one: Legal Illegal

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>3/3/02</td>
<td>Flu</td>
<td>Yonge</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3/3/02</td>
<td>Measles</td>
<td>Yonge</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3/3/02</td>
<td>Chicken Pox</td>
<td>Yonge</td>
</tr>
</tbody>
</table>
5. Consider the following instance of Treatment and again assume that Location contains three tuples 
{ (Yonge, 1), (Bloor, 2), (Maine, 1) }.

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1/1/02</td>
<td>Glaucoma</td>
<td>Bloor</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1/1/02</td>
<td>Flu</td>
<td>Maine</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2/2/02</td>
<td>Chicken Pox</td>
<td>Yonge</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3/3/02</td>
<td>Flu</td>
<td>Bloor</td>
</tr>
</tbody>
</table>

Show the contents of tables Treatment and Location after the following update.

```sql
delete from Treatment
where Location = 'Bloor';
```

6. Consider the following instance of Treatment and again assume that Location contains three tuples 
{ (Yonge, 1), (Bloor, 2), (Maine, 1) }.

<table>
<thead>
<tr>
<th>PhysicianID</th>
<th>PatientID</th>
<th>Date</th>
<th>Diagnosis</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1/1/02</td>
<td>Glaucoma</td>
<td>Bloor</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1/1/02</td>
<td>Flu</td>
<td>Maine</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2/2/02</td>
<td>Chicken Pox</td>
<td>Yonge</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3/3/02</td>
<td>Flu</td>
<td>Bloor</td>
</tr>
</tbody>
</table>

Show the contents of tables Treatment and Location after the following update.

```sql
update location
set id = 'Queen'
where head = (select max(head) from location);
```
Question 6. [4 marks]
Consider the relation $R$ on attributes $ABCDEFG$, with the following functional dependencies:

$$A \rightarrow BEF, \quad G \rightarrow BC, \quad CD \rightarrow G.$$ 

Throughout this question, we use the term “key” in the sense of “minimal key”.

Part (a) [1 mark]
Is $AD$ part of a key for $R$?  Yes  No

Explain:

Part (b) [1 mark]
Is $AB$ part of a key for $R$?  Yes  No

Explain:

Part (c) [1 mark]
Is $ADCE$ a key (not part of a key, $is$ a key) for $R$?  Yes  No

Explain:

Part (d) [1 mark]
Is $ADG$ a key (not part of a key, $is$ a key) for $R$?  Yes  No

Explain:
**Question 7.** [6 marks]

Consider the relation $R$ on attributes $LMNOPQ$, with the following functional dependencies:

$$
P \rightarrow OM, \quad MP \rightarrow N, \quad L \rightarrow NO, \quad ON \rightarrow LM, \quad OP \rightarrow L.
$$

**Part (a) [2 marks]**

Create a valid instance of $R$ that satisfies the functional dependencies, has two tuples, and contains redundancy.

State exactly what data in this instance is redundant and explain what makes it redundant.

**Part (b) [2 marks]**

What important property does BCNF decomposition guarantee that 3NF synthesis does not guarantee?

What important property does 3NF synthesis guarantee that BCNF decomposition does not guarantee?

**Part (c) [2 marks]**

Consider this set of relations and their functional dependencies:

$\begin{align*}
R1(A, B, C) & \quad R2(C, D, E) & \quad R3(E, F, G) \\
A \rightarrow B & \quad \text{none} & \quad EF \rightarrow G \\
C \rightarrow AB &
\end{align*}$

Could these relations be the result of the BCNF decomposition algorithm? **Yes**  **No**

Explain:
Question 8. [6 marks]
Consider relation $R(A, B, C, D, E, F)$ with functional dependencies $S$.

$$S = \{AD \rightarrow BC, \ ABD \rightarrow E, \ BE \rightarrow CDF, \ CD \rightarrow A, \ BF \rightarrow C\}$$

Compute a minimal basis for $S$. Explain all steps and show your rough work. There will be no marks for a correct answer without this. Put your final answer where shown on the next page.
Write your minimal basis for S below. Merge the RHSs of the functional dependencies, and present the functional dependencies in alphabetical order as you did on Assignment 3.
Question 9. [4 marks]
Suppose we are employing the 3NF synthesis algorithm on a relation \( R(A, B, C, D, E, F) \), and we already have the following minimal basis:

\[ S = \{ CD \rightarrow A, \ CEF \rightarrow AB, \ D \rightarrow E \} \]

Use the 3NF synthesis algorithm to obtain a set of relations that are in 3NF. Explain all steps in your answer. There will be no marks for a correct answer without a good explanation of the steps.
Question 10. [6 marks]
Consider the relation R(A, B, C, D, E) with the following functional dependencies:

\[ A \rightarrow B, \quad C \rightarrow DE, \quad E \rightarrow AD \]

Part (a) [3 marks]
Suppose we are considering a decomposition of this relation into two relations R1(A, D, E) and R2(B, C, D). Use a technique such as the chase test to check whether this is not a lossless join decomposition. Show your reasoning.

Circle your conclusion: Lossless Lossy

Part (b) [3 marks]
Suppose instead we are considering a decomposition of this relation into the relations R1(A, B), R2(C, D) and R3(A, C, E). Use a technique such as the chase test to check whether this is not a lossless join decomposition. Show your reasoning.

Circle your conclusion: Lossless Lossy


**Question 11.** [6 marks]
Consider the ER diagram below.

![ER Diagram](image)

In this question you will think about how the number of entities in one entity set relates to the number of entities in the other entity sets. When asked for a maximum, if there is no limit write “no limit”.

**Part (a)** [2 marks]
Suppose $|A|$ is 50.

Minimum value of $|B|$: ________________  Maximum value of $|B|$: ________________

Minimum value of $|C|$: ________________  Maximum value of $|C|$: ________________

**Part (b)** [2 marks]
Suppose $|B|$ is 100.

Minimum value of $|A|$: ________________  Maximum value of $|A|$: ________________

Minimum value of $|C|$: ________________  Maximum value of $|C|$: ________________

**Part (c)** [2 marks]
Suppose $|C|$ is 80.

Minimum value of $|B|$: ________________  Maximum value of $|B|$: ________________

Minimum value of $|A|$: ________________  Maximum value of $|A|$: ________________
**Question 12.** [6 marks]

**Part (a) [3 marks]**
Consider the ER diagram below.

Translate this ER diagram into a relational schema. Use underlining to express the key of each relation, and subset notation to express all foreign keys.

**Part (b) [3 marks]**
Here is nearly the same ER diagram. The cardinalities have been changed.

Translate this ER diagram into a relational schema. Use underlining to express the key of each relation, and subset notation to express all foreign keys.
Question 13. [10 marks]
Closets, Closets, Closets would like to start mining information about its online reputation. They have
gathered a set of tweets that mention the hashtag #ClosetsClosetsClosets. For each tweet, they have the
tweeter’s Twitter account (meaning the identifier or handle for the account), the date tweeted, and the
contents of the tweet. For each Twitter account they also track the number of followers. A Twitter account
can produce multiple tweets on the same date so each Tweet has a unique TweetID. Of course, each Tweet
is produced by exactly one Twitter account. They also record the set of all hashtags mentioned in each
tweet (excluding #ClosetsClosetsClosets, which is mentioned in all the tweets they gathered). They have
hired a social media consulting firm that has provided them with further data: the facebook account (at
most one) for many of the Twitter Accounts (but not for all Twitter Accounts, since they don’t all an
identifiable facebook account). A facebook account has a unique id and a count (which may be zero or
more) of the number of times the facebook account has liked a product produced by Closets, Closets,
Closets. Some facebook accounts may be associated with multiple twitter accounts.

Design an ER Diagram to represent this information. Your solution should not have redundancy.
Use this page for rough work. If you want work on this page to be marked, please indicate this clearly at the location of the original question.
Use this page for rough work. If you want work on this page to be marked, please indicate this clearly at the location of the original question.