Midterm Exam
CS 437 / 537 – Database System
Oct 18, 2016

Start: 9:00am, Stop: 10:15am (1 hour and 15 minutes)

Name:

NetId:

<table>
<thead>
<tr>
<th>Section</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short answer</td>
<td>/ 15</td>
</tr>
<tr>
<td>Multiple choices</td>
<td>/ 30</td>
</tr>
<tr>
<td>SQL programming &amp; E-R design</td>
<td>/ 30</td>
</tr>
<tr>
<td>DB knowledge</td>
<td>/ 25</td>
</tr>
<tr>
<td>Total Score</td>
<td>/ 100</td>
</tr>
</tbody>
</table>
Part 1: Short answer (15 points)

Please briefly answer the following questions

1.1. What is the difference between a database system and a file system? (3 points)

Any reasonable answer is acceptable.

A file system specifically focuses on file or object storage; on the contrary, a database system can offer better data integrity protection, consistency, fault tolerance and data isolation.

1.2. In relational algebra, define set intersection using other relational algebra operators. (3 points)

Assume two relations r and s, and set intersection should be: r – (r – s) . You can also use other operators, e.g., union operators. However, you need to make sure the solution is a general solution assuming you do not know the number of attributes in each table.

1.3. What is the difference between BCNF and 3NF? (3 points)

Any reasonable answer is acceptable.

A relation R is in BCNF if it is in 3NF and for each functional dependency X -> A in R, X should a key or superkey in R. In other words, BCNF acts differently from 3NF only when there exist multiple overlapping candidate keys.

1.4. What is the difference between JOIN and INNER JOIN? (3 points)

JOIN is the default representation of INNER JOIN, so they are the same. Many students answered this incorrectly. Note that you cannot assume JOIN as different ways, because in SQL (we are now talking about exactly SQL language), JOIN is the shorter representation of INNER JOIN.

1.5. Why do we need embedded SQL? (3 points)

Embedded SQL enables programmers to conveniently write program and SQL. Any reasonable answer is acceptable.

Part 2: Multiple choice (15 x 2 points each = 30 points)

2.1. Which of the following statements is NOT correct?

a. Logical schema is the overall logical structure of the database
b. Physical schema is the overall physical structure of the database
c. Database Instance is the actual content of the database at a particular point in time
d. Physical data independence is the ability to modify the logical schema without changing the physical schema
2.2. Which of the following statements is NOT correct?
   a. \( K \) is a superkey of a given relation schema \( R \) if values for \( K \) are sufficient to identify a unique tuple of each possible relation \( r(R) \)
   b. Superkey \( K \) is a candidate key if \( K \) is minimal
   c. In some case, primary key may not be included in candidate keys
   d. Foreign key has a constraint -- value in one relation must appear in another

2.3. Which of the following statements about DB design phases is NOT correct?
   a. Logical design is to decide on the database schema
   b. Physical design is to decide on the physical layout of the database
   c. Developers make the decision on what attributes should be recorded in the DB
   d. Computer scientists make the decision on what relation schemas should the database have and how should the attributes be distributed among the various relation schemas

2.4. Which of the following statements about entity sets is NOT correct?
   a. A specific person cannot be an entity
   b. A set of companies is an entity set
   c. \( \text{course} = (\text{ID}, \text{name}, \text{credits}) \) could be an entity
   d. A subset of the attributes form a primary key of the entity set

2.5. Which of the following statements about a relationship set is NOT correct?
   a. Binary relationship includes two entity sets
   b. Most relationship sets in a DB system are non-binary
   c. A binary relationship’s degree is two
   d. Entity sets of a relationship set need not be distinct

2.6. Which of the following statements about weak entity sets is NOT correct?
   a. An entity set that does not have a primary key is called as a weak entity set
   b. The existence of a weak entity set depends on the existence of an identifying entity set
   c. The discriminator of a weak entity set is the set of attributes that distinguishes among all the entities of a weak entity set
   d. In E-R diagrams, a weak entity set is depicted via a rectangle

2.7. Which of the following statements about outer joins is NOT correct?
   a. Outer join is an extension of the join operation that avoids loss of information
   b. It specifies left or right or full
   c. It could be used with the clauses: natural, on, and using
   d. If none of the clauses is used it is natural join

2.8. Which of the following statements about views is NOT correct?
   a. A view provides a mechanism to hide certain data from the view of certain users
   b. View definition is the same as creating a new relation by evaluating the query expression
   c. A view can be defined using the create view statement
d. Once a view is defined, the view name can be used to refer to the virtual relation that the view generates

2.9. Which of the following statements about built-in data types in SQL is NOT correct?
   a. `timestamp`
   b. `date`
   c. `timestamp`
   d. `interval`

2.10. Which of the following statements about embedded SQL is NOT correct?
   a. Before executing any SQL statements, the program in some cases does not need to connect to the DB
   b. Variables of the host language can be used within embedded SQL statements
   c. EXEC SQL statement is used to identify embedded SQL request to the preprocessor
   d. Embedded SQL is one of tools on accessing SQL from a programming language

2.11. Which of the following statements about trigger is NOT correct?
   a. A trigger is a statement that is executed automatically by the system as a side effect of a modification to the DB
   b. Triggering event can be insert, delete or update
   c. Triggers can be activated before or after an event
   d. To design a trigger scheme, we must specify the actions but may not need to specify the conditions

Consider a database with the following three tables

```
R
A B
1 2
3 2
5 6
7 8
9 8

S
B C
6 2
2 4
8 1
8 3
2 5

T
A C
7 1
1 2
9 3
5 4
3 5
```

And the following four results:

```
A B C
1 2 4
1 2 5
3 2 4
3 2 5
5 6 2
7 8 1
7 8 3
9 8 1
9 8 3

A B C
1 2 2
3 2 5
5 6 2
7 8 1
9 8 3

A B C
1 6 2
3 2 5
5 2 4
7 8 1
9 8 3

A B C
3 2 5
7 8 1
9 8 3
```

2.12. What is the resulting table of $R \bowtie (S \bowtie T)$
Part 3: SQL Programming & E-R Design (30 points)

3.1. Consider the following tables:

- **author**: Each tuple represents an author, whose id and name are given by aid and aname, respectively. The underlined attribute is the candidate key.
- **book**: Each tuple represents a book. The attribute category describes the type of the book (e.g., novel, science and music).
- **student**: Each tuple represents a student.
- **write**: A tuple means that book bid was authored by author aid.
- **borrow**: A tuple means that student sid checked out book bid at checkout-time, and returned it at return-time.

All the attributes are strings, except checkout-time and return-time, which are integers. A smaller checkout-time (or return-time) represents an earlier timestamp.

3.1.1. Find the titles of all the books that have ever been borrowed by bme students (i.e., dept = 'bme'). (4 points)

```
SELECT title
FROM book, borrow, student
WHERE book.bid = borrow.bid AND borrow.sid = student.sid AND dept = 'bme';
```
3.1.2. Find the number of distinct students that have ever borrowed the book with bid='b100'. (4 points)

SELECT count(distinct sid)
FROM borrow
WHERE bid = 'b100';

3.1.3. Find the titles of all the books that have ever borrowed by students from at least 10 distinct departments. (4 points)

SELECT title
FROM book, borrow, student
WHERE book.bid = borrow.bid AND borrow.sid = student.sid
GROUP BY book.bid, title
HAVING count(distinct dept) >= 10;

3.2. Consider the following tables:

student(ssn, name)
prof(ssn, name)
course(number, instructor-ssn, title, credits, roomnum)
enroll(student-ssn, coursenum)
room(number, capacity)

3.2.1. Write an SQL query that lists, in alphabetical order, the title of all courses either are taught by "Smith" OR are taught in room number 444. Do not list duplicate titles. (5 points)

SELECT distinct title
FROM course
WHERE number in
    (SELECT number FROM course, prof
     WHERE instructor-ssn = ssn AND name = "Smith")
OR roomnum = 444
ORDER BY title;

3.2.2. Write an SQL query that considers all the courses that have ever been taught by "Brown" and are of 3 credits, and groups them according to title. (5 points)

I noticed some students used easier way to solve this question, which is good!

SELECT title
FROM course, room
WHERE roomnum = room.number
AND credits = 3
AND course.number in
    (SELECT number
     FROM course, prof
WHERE instructor-ssn = ssn and name = “Brown”)
GROUP BY title;

3.2.3. Render the above databases in the E/R model. Make sure to indicate that every course has a unique instructor and room. (8 points)

I set a low bar for this question, but you need to explicitly mention every course has a unique instructor and room in your E-R diagram.

Part 4: Database Knowledge (5 x 5 points each = 25 points)

Consider the relation R:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

4.1. In the above relation $R(A, B, C, D)$, whether the functional dependency $AC \rightarrow B$ holds in the given relation? Why? (5 points)

Yes or no. From the given table (i.e., only the given four tuples), $AC \rightarrow B$ holds. Answering no can also get full credit if you argue that FD dependence cannot be inferred from a set of given tuples.

4.2. In the above relation $R(A, B, C, D)$, what are the possible keys of the table? (5 points)

Possible keys are D or AC. Answer one can get full credit. You can also get full credit if you argue we cannot infer keys from the tuples.

4.3. Provide a lossless-join decomposition of the above relation $R(A, B, C, D)$ (into to two relations). If there is no such decomposition, please explain why. (5 points)

(ABC) and (ACD). You can have different answers that depend on what primary key you used.

4.4. For the above relation $R(A, B, C, D)$, write a SQL program to return the lowest value of $D$. (5 points)

```
SELECT MIN(D)
FROM R;
```

4.5. What is the result of the following query, for the above relation $R(A, B, C, D)$? (5 points)
SELECT A, B
FROM R
WHERE C > (SELECT D FROM R WHERE A = 3);

This query will return an error, since the subquery returns a scalar. Only five students answered this correctly.