First Name: ____________________________

Last Name: ____________________________

MSU Net ID: ____________________________

Sample Exam (2019) for CSE 480: Database Systems

Instructions:

• DO NOT START THE EXAM UNTIL TOLD TO DO SO

• You need to answer 6 of the 7 questions.

• On one of the questions make a large slash through it, which indicates that it should not be graded.

• Answer the questions in the spaces provided on the page. If you run out of room for an answer, continue on the back of that page.

• Show us the crossed out question and present your MSU ID (or other form of ID) when you hand in your exam.

Figure 1: http://xkcd.com/1129/
Question 1: SQL statement ........................................... 50 points

I have 2 tables containing information about athletes and teams. The table `athletes` has the following columns: name (the athlete’s name, unique), year (the graduating year of the athlete), team (the id of the team the athlete is part of), and grade (the athlete’s GPA). The table `teams` has the following columns: name (the team’s name), id (the team’s identification number, unique), and school (the name of the team’s school).

I want to know (for each team) what the was the average grade of the students who played for that team. However, I am only concerned with teams with more than one player. Order the result from the highest average grade to the least.

Here is some example data:

<table>
<thead>
<tr>
<th>Table: athletes</th>
<th>name</th>
<th>year</th>
<th>team</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Josh</td>
<td>2017</td>
<td>7</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Emily</td>
<td>2016</td>
<td>7</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Tyler</td>
<td>2015</td>
<td>8</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Grant</td>
<td>2015</td>
<td>8</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Charles</td>
<td>2016</td>
<td>9</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Laura</td>
<td>2016</td>
<td>10</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table: teams</th>
<th>name</th>
<th>id</th>
<th>school</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparties</td>
<td>7</td>
<td>MSU</td>
</tr>
<tr>
<td></td>
<td>Greens</td>
<td>8</td>
<td>MSU</td>
</tr>
<tr>
<td></td>
<td>Blues</td>
<td>9</td>
<td>UM</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>10</td>
<td>MSU</td>
</tr>
</tbody>
</table>

And the desired results for this example data:

<table>
<thead>
<tr>
<th>team</th>
<th>average_grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparties</td>
<td>3.5</td>
</tr>
<tr>
<td>Greens</td>
<td>2.75</td>
</tr>
</tbody>
</table>

Fill in the blanks when the clause should be used in the SELECT statement (you must always use qualified names):

```
SELECT __________________________
FROM ____________________________
     ____________________________ JOIN ____________________________
ON _____________________________
WHERE ___________________________
GROUP BY ________________________
HAVING __________________________
ORDER BY _________________________;
```
Question 2: Corrupted Database.................................50 points

Unfortunately, Pikachu got a bit upset too close to our database, and it erased some of
the values. The data isn’t essential, but we would like to recover what we can. Here’s
what we know:

• The relation has 6 attributes (A-F) as shown below.
• The relation has one multivalued dependency: \( AB \rightarrow\rightarrow F \)
• The relation has two functional dependencies: \( C \rightarrow D \)
• and \( AC \rightarrow E \)

Use the data given in the relation below to fill in the missing values. If the value is
unknown, put a question mark (?) instead of a number.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>9</td>
<td>2</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Question 3: Schedules ................................................................. 50 points

(a) (10 points) What is the difference between a serial and a serializable schedule?

(b) (10 points) Which of the four ACID principles are violated by a non-serializable schedule?
   ○ Atomicity  ○ Consistency  ○ Isolation  ○ Durability

(c) (10 points) When does a conflict occur between two transactions?

(d) (10 points) Define what conflict-serializable means.

(e) (10 points) Are all serializable schedules conflict-serializable?
Question 4: Precedence ................................................................. 50 points

(a) (10 points) What are the 3 conditions that determine if an action \( A_1 \) from one transaction takes precedence over an action \( A_2 \) in a different transaction?

(b) (20 points) What are the conflicts and implied transaction precedence for the following schedule:
\[
S: r_1(A); r_2(B); w_1(A); w_2(A); w_3(A); w_3(B); r_1(B);
\]

(c) (20 points) Draw the Precedence Graph for the schedule and indicate if it is conflict-serializable.
Question 5: Simple Locks ................................................................. 50 points

(a) (10 points) According to two-phase locking, when can locks not be acquired?
   - Before all read and write actions
   - Before the last read or write action
   - After the first unlock action
   - Only upon commit or rollback

(b) (10 points) According to strict two-phase locking, when can unlocks be performed?
   - Before all read and write actions
   - Before the last read or write action
   - After the first unlock action
   - Only upon commit or rollback

(c) (30 points) For this question there is only one type of lock (an exclusive lock).
   For the following schedule (S), output all of the read and write actions with the
   needed lock and unlock actions (i.e. \( l_1(A) \) and \( u_1(A) \)). Only lock/unlock when
   such actions are required. You can assume a commit occurs when a transaction has
   completed all of its actions.
   S: \( r_1(A); w_1(B); w_2(C); r_1(A); w_2(A); r_3(B); \)
Question 6: Multiple Types of Locks ........................................... 50 points

Below are three transactions (and six total actions). Unlocks must happen after all the
actions in a transaction have taken place (strict two-phase locking). For this problem,
there are shared locks \((sl_1(A))\) and exclusive locks \((xl_1(A))\) and either/both are unlocked
with \((u_1(A))\). Below, the order of each of the actions are interleaved. For each part,
output the necessary locks and unlocks that the transaction should perform to make the
action take place (don’t forget to include the action itself).

<table>
<thead>
<tr>
<th>(T_1)</th>
<th>(T_2)</th>
<th>(T_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_1(X))</td>
<td>(w_2(Y))</td>
<td>(w_3(Z))</td>
</tr>
<tr>
<td>(w_1(Y))</td>
<td>(w_3(Y))</td>
<td></td>
</tr>
<tr>
<td>(w_1(X))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) \(T_1\): \(r_1(X)\)

(b) \(T_2\): \(w_2(Y)\)

(c) \(T_1\): \(w_1(Y)\)

(d) \(T_3\): \(w_3(Z)\)

(e) \(T_1\): \(w_1(X)\)

(f) \(T_3\): \(w_3(Y)\)
Question 7: Transaction Modes .................................................. 50 points

After the statement is completed by the associated connection, write which connections are holding each type of lock in the table. If no connection holds a type of lock, leave it blank. A connection can only hold one lock at a time. If an action can’t be granted the necessary locks, cross out that action and proceed as if that action wasn’t in the schedule (don’t perform a rollback).

<table>
<thead>
<tr>
<th>ID</th>
<th>SQL statement</th>
<th>Shared</th>
<th>Reserved</th>
<th>Exclusive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CREATE TABLE ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>BEGIN TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>INSERT INTO ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SELECT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BEGIN IMMEDIATE TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DELETE FROM ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ROLLBACK TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BEGIN TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BEGIN TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>COMMIT TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SELECT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SELECT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>INSERT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>INSERT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>COMMIT TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>COMMIT TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>BEGIN EXCLUSIVE TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SELECT ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>UPDATE ...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>COMMIT TRANSACTION;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Question 8: Deadlocks ........................................... 50 points

Our database is running into problems, as many transactions are waiting for locks held by other transactions. Here is what each transaction is waiting for:

- $T_1$ is waiting on $T_4$
- $T_2$ is waiting on $T_7$
- $T_3$ is waiting on $T_2$
- $T_4$ is waiting on $T_1$
- $T_5$ is waiting on $T_8$
- $T_6$ is waiting on $T_2$
- $T_7$ is waiting on $T_6$
- $T_8$ is not waiting

(a) (20 points) Draw the Wait-For graph for transactions $T_{1-8}$.

(b) (10 points) Which transactions are deadlocked?

- $T_1$
- $T_2$
- $T_3$
- $T_4$
- $T_5$
- $T_6$
- $T_7$
- $T_8$

(c) (20 points) What needs to be done to resolve the deadlock?

.................................................................
Question 9: Deadlock Resolution ........................................................... 50 points

Below is a timeline as to when transactions were started, received locks, and requested a lock. For this problem you can assume all locks are exclusive. You can also assume transaction restarts are handled after the events described.

1. $T_W$ starts
2. $T_W$ gets lock on A
3. $T_X$ starts
4. $T_X$ gets lock on B
5. $T_W$ wants lock on B, waits on $T_X$
6. $T_Z$ starts
7. $T_Z$ gets lock on C
8. $T_Z$ wants lock on B, rollback!

(a) (10 points) Which deadlock resolution strategy is being used?
○ Wait-Die ○ Wound-Wait

(b) (20 points) If the same events occurred using the other deadlock resolution strategy, what would the timeline be?

........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................

(c) (10 points) Before the rollback on the original Step 8, draw the Wait-For Graph for the transactions.

(d) (10 points) If you want to have fewer rollbacks (but perhaps undoing more work) which strategy should you use?
○ Wait-Die ○ Wound-Wait
Question 10: Optimistic Scheduling.................................................. 50 points

There are two transactions ($T_i$ and $T_j$). $T_i$ started before $T_j$. Both $T_i$ and $T_j$ performing
reads and/or writes on database element $E$.

(a) (10 points) Give an example of a Read-too-late event and why it is a problem.

(b) (10 points) Give an example of a Write-too-late event.

(c) (10 points) Why are physically unrealizable behaviors a problem?

(d) (10 points) What is the Thomas Write Rule?

(e) (10 points) When does an optimistic scheduler (time stamp/validating) outperform
a pessimistic scheduler (locking)?
Question 11: Legal Optimistic Schedules................................................. 50 points

Below is a time line of when two transactions ($T_i$ and $T_j$) performed actions on the database.

1. $T_i$ begins transaction
2. $r_i(A)$
3. $T_j$ begins transaction
4. $w_i(B)$
5. $w_j(A)$
6. $T_i$ commits transaction
7. $T_j$ commits transaction

For proposed action, indicate if that action were added (in isolation from the other proposed actions) would the resulting schedule result in physically unrealizable behavior.

(a) $w_i(A)$ between steps 3 and 4 ○ Legal ○ Physically Unrealizable
(b) $w_j(B)$ between steps 3 and 4 ○ Legal ○ Physically Unrealizable
(c) $w_j(B)$ between steps 4 and 5 ○ Legal ○ Physically Unrealizable
(d) $r_j(B)$ between steps 4 and 5 ○ Legal ○ Physically Unrealizable
(e) $w_j(C)$ between steps 4 and 5 ○ Legal ○ Physically Unrealizable
(f) $r_i(A)$ between steps 5 and 6 ○ Legal ○ Physically Unrealizable
(g) $w_i(A)$ between steps 5 and 6 ○ Legal ○ Physically Unrealizable
(h) $r_j(A)$ between steps 6 and 7 ○ Legal ○ Physically Unrealizable
Question 12: Entity/Relationship Diagram. .................................................. 50 points
We are creating a database to record which farms raise which types of animals. Below are some facts that need to be represented in the database:

- Each farm has a name and a unique address
- Each farm may raise multiple animals, and each animal can be uniquely identified by the combination of its species and id number.
- Each animal is raised by one farm.
- Each animal also has a nickname, but this may not be unique to each.
- Some animals are breeders, meaning that we know its sex and the season in which it can be breed.
- Some animals are producers, meaning that they produce one or more products (e.g. eggs, milk, wool).
- The products that can be produced have a unique id, as well as, a name and price.

Draw an E/R diagram illustrating the structure of a database capturing the above information.
Question 13: Converting E/R Diagram ........................................ 50 points

Below is an E/R representing information about pets and their owners.

(a) (40 points) Write a relational database schema representing the E/R diagram. Be sure to combine relations when possible. Use the Object-Oriented method to create relations for the subclasses.

(b) (10 points) Into what relation(s) would you put a pet that is purebred and trained?
Question 14: Dependencies ................................................................. 50 points

Below is a relation about pets and owners.

<table>
<thead>
<tr>
<th>Pet</th>
<th>Species</th>
<th>Age</th>
<th>Owner</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoe</td>
<td>Dog</td>
<td>8</td>
<td>Josh</td>
<td>123-4567</td>
</tr>
<tr>
<td>River</td>
<td>Cat</td>
<td>11</td>
<td>Josh</td>
<td>123-4567</td>
</tr>
<tr>
<td>Harry</td>
<td>Rat</td>
<td>4</td>
<td>Emily</td>
<td>246-1357</td>
</tr>
<tr>
<td>Ron</td>
<td>Rat</td>
<td>4</td>
<td>Emily</td>
<td>246-1357</td>
</tr>
<tr>
<td>Snape</td>
<td>Rat</td>
<td>3</td>
<td>Emily</td>
<td>246-1357</td>
</tr>
</tbody>
</table>

(a) (30 points) Which functional dependencies are obeyed?

- Pet → Species
- Species → Owner Phone
- Phone → Owner Phone
- Species → Pet
- Pet → Phone
- Age → Owner

(b) (20 points) The above relation obeys this multivalued dependency (Owner →→ Phone).

What other rows must also be added if the row below is added? Fill in the additional dependent rows as needed (you may not need them all).

<table>
<thead>
<tr>
<th>Pet</th>
<th>Species</th>
<th>Age</th>
<th>Owner</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mal</td>
<td>Bird</td>
<td>12</td>
<td>Josh</td>
<td>555-5555</td>
</tr>
</tbody>
</table>
Question 15: Closures .................................................. 50 points

Here’s a relation (R), its attributes and its functional dependencies (F):

R(A, B, C, D, E)
C D → B
A → D
E → C

(a) (10 points) Which of the following are in the attribute set closure \( \{AB\}^+ \)?

- \( \{A\} \)
- \( \{AB\} \)
- \( \{D\} \)
- \( \{BC\} \)
- \( \{CD\} \)
- \( \{ABD\} \)

(b) (10 points) Which of the following are in the functional dependency closure of F \( (F^+) \)?

- \( C D \rightarrow B \)
- \( A \rightarrow A \)
- \( D \rightarrow B \)
- \( E \rightarrow C D \)
- \( A C \rightarrow B \)
- \( A E \rightarrow B \)

(c) (20 points) Which of the following are superkeys?

- \( \{ABCDE\} \)
- \( \{A\} \)
- \( \{BCE\} \)
- \( \{AE\} \)
- \( \{AB\} \)
- \( \{ABE\} \)

(d) (10 points) Which of the following are keys?

- \( \{ABCDE\} \)
- \( \{A\} \)
- \( \{BCE\} \)
- \( \{AE\} \)
- \( \{AB\} \)
- \( \{ABE\} \)
Question 16: Lossless Joins ................................................................. 50 points

Here’s a relation (R), its attributes and its functional dependencies (F):

\[ R(A, B, C, D, E) \]

\[ C \rightarrow D \]

\[ A \rightarrow D \]

\[ E \rightarrow C \]

(a) (10 points) Which of the following sets of relations maintain the lossless join property?

- \( R_1(ABCDE), R_2(ABCD) \)
- \( R_1(AD), R_2(ABCE) \)
- \( R_1(AB), R_2(BCDE) \)
- \( R_1(CD), R_2(ABCE) \)

(b) (20 points) Which of the following sets of relations are entirely in Boyce-Codd Normal Form?

- \( R_1(ABCDE) \)
- \( R_1(AD), R_2(ABCE) \)
- \( R_1(CDB), R_2(ACDE) \)
- \( R_1(CDB), R_2(AD), R_3(CE), R_4(AE) \)

(c) (20 points) Which of the following sets of relations are dependency preserving?

- \( R_1(ABCDE) \)
- \( R_1(AD), R_2(ABCE) \)
- \( R_1(CDB), R_2(ACDE) \)
- \( R_1(CDB), R_2(AD), R_3(CE), R_4(AE) \)
Question 17: Decomposition ......................................................... 50 points

Here’s a relation (R), its attributes and its functional dependencies (F):

R(A, B, C, D, E)
C D → B
A → D
E → C

(a) (30 points) Decompose the above relation using the Boyce-Codd Normal Form decomposition. Use the order of F when checking for violations.

(b) (20 points) What is the key for the relation R?