Homework 6

IMPORTANT
• Deposit hard copy of your answers in class at 1:30pm on Tue, 3/25/2014.
• Separate answers, as usually, i.e., please solve each of the 4 questions on a separate page, and type the usual, full information, on each page: your name, Andrew ID, course #, Homework #, and Question #.

Reminders
• Plagiarism: Homework may be discussed with other students, but all homework is to be completed individually.
• Typeset all of your answers whenever possible. Illegible handwriting may get no points, at the discretion of the graders.
• Late homeworks: please email late homeworks
  – to all TAs
  – with the subject line exactly 15-415 Homework Submission (HW 6)
  – and the count of slip-days you are using.

For your information:
• Graded out of 100 points; 4 questions total
• Rough time estimate: ≈6 hours (1-2 hours for each question)

<table>
<thead>
<tr>
<th>Question</th>
<th>Points</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query Optimization</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Functional Dependencies</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Decompositions</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Normal Forms</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Question 1: Query Optimization ........................ [30 points]
Submit on separate page
Course: 15-415/615; HW: ; Q: Name: ____________; andrew-id: ____________; late days: ___________________________

For this problem we consider a database with following three tables:

1. Movies(title, year)
2. Actors(actorID, name)
3. Acted_in(actorID, title, year)

For these tables we know the following statistics:

- **Movies** consists of $N_1 = 50,000$ tuples
  - $V$(title, Movies) = 30,000 distinct movie titles
  - $V$(year, Movies) = 90 distinct years (1925-2015)
- **Actors** consists of $N_2 = 200,000$ tuples
  - $V$(actorID, Actors) = 200,000 distinct actor ID’s
  - $V$(name, Actors) = 160,000 distinct names
- **Acted_in** consists of $N_3 = 1,000,000$ tuples
  - $V$(actorID, Acted_in) = 180,000 distinct actor ID’s
  - $V$(title, Acted_in) = 29,000 distinct movie titles
  - $V$(year, Acted_in) = 90 distinct years (1925-2015)

(a) Yes/No questions:

i. **[3 points]** Ignoring semantics, and given the above statistics, could **title** be a candidate key for **Movies**?
   - □ Yes □ No

ii. **[3 points]** Again, ignoring semantics, could **actorID** be a candidate key for **Actors**?
    - □ Yes □ No

iii. **[3 points]** Could **actorID** be a candidate key for **Acted_in**?
    - □ Yes □ No

(b) Selectivity estimations. Give *fourth* significant digit accuracy. No partial credit will be given.

i. **[3 points]** Estimate the number of resulting tuples for the query:
   
   ```
   SELECT * FROM Movies WHERE year = 1995;
   ```
   - i. __________

ii. **[4 points]** Estimate the number of resulting tuples for the query:

   ```
   SELECT * FROM Movies
   WHERE year = 2000 AND title = "Dude, Where’s my Car?”;
   ```
   - ii. __________

iii. **[4 points]** Estimate the number of resulting tuples for the query:

   Question 1 continues...
iii. ____________

iv. [5 points] Estimate the number of resulting tuples for the query:

```sql
SELECT *
FROM Actors JOIN Acted_in AS Ai
ON Actors.actorID = Ai.actorID;
```

iv. ____________

v. [5 points] Estimate the number of resulting tuples for the query:

```sql
SELECT *
FROM Movies JOIN Acted_in AS Ai
ON Movies.year = Ai.year AND Movies.title = Ai.title;
```

ev. ____________

Homework 6 continues...
Question 2: Functional Dependencies ............ [20 points]

2.1 (This question is a modified version of exercise 19.6 in the textbook.) For the first set of questions consider the following legal instance of a relational schema $S$ with attributes $ABC$:

$$
\begin{array}{c|cc}
S & A & B & C \\
\hline
1 & a & X \\
4 & a & Y \\
5 & b & X \\
\end{array}
$$

Table 1: Legal instance of schema $S$ for question 2.1

(a) Which of the following dependencies are violated by the instance of $S$ in Table 1?

i. [1 point] □ Yes □ No : $A \rightarrow B$ is violated.

ii. [1 point] □ Yes □ No : $B \rightarrow A$ is violated.

iii. [1 point] □ Yes □ No : $BC \rightarrow A$ is violated.

iv. [1 point] □ Yes □ No : $B \rightarrow C$ is violated.

v. [1 point] □ Yes □ No : $C \rightarrow AB$ is violated.

(b) [1 point] By only observing the instance of $S$ in Table 1 can you identify the functional dependencies that hold on schema $S$?

□ Yes □ No

2.2 For the next set of questions consider the relational schema $r = \{P, Q, R, S, T, U, V\}$ and the set of functional dependencies FD:

$P \rightarrow S$ \hspace{1cm} (1)

$PQ \rightarrow ST$ \hspace{1cm} (2)

$S \rightarrow RU$ \hspace{1cm} (3)

$RU \rightarrow S$ \hspace{1cm} (4)

$PT \rightarrow V$ \hspace{1cm} (5)

(a) [3 points] Which of the following is a minimum cover of the FD?

(a) The given FD is a minimum cover.

(b) $\{P \rightarrow S; PQ \rightarrow T; PQ \rightarrow S; S \rightarrow R; S \rightarrow U; PT \rightarrow V; RU \rightarrow S\}$

(c) $\{P \rightarrow R; P \rightarrow U; PQ \rightarrow T; PT \rightarrow V\}$

(d) $\{P \rightarrow S; PQ \rightarrow T; S \rightarrow R; S \rightarrow U; PT \rightarrow V; RU \rightarrow S\}$

(e) none of the above - the cover is

(b) Yes/No: Which of the following functional dependencies can be deduced, from the above set of functional dependencies (Eq. (1)-(5))? 

Question 2 continues...
i. [1 point]  □ Yes  □ No : $P \to U$

ii. [2 points]  □ Yes  □ No : $PT \to SV$

iii. [1 point]  □ Yes  □ No : $SQ \to V$

iv. [1 point]  □ Yes  □ No : $PS \to RV$

v. [1 point]  □ Yes  □ No : $P \to V$

vi. [1 point]  □ Yes  □ No : $PSRU \to QT$

(c) [2 points] True or False: The attribute closure $\{P\}^+$ is $\{R, S, U\}$.
□ True  □ False

(d) [2 points] True or False: The attribute closure $\{PQ\}^+$ is $\{P, Q, R, S, T, U, V\}$.
□ True  □ False

Homework 6 continues...
Question 3: Decompositions .......................... [20 points]
Submit on separate page
Course: 15-415/615; HW: ; Q:
Name: ___________________________; andrew-id: _______________________; late days:
For this set of questions consider the following relational schema $S = \{A, B, C, D, E, F, G\}$:

\[
\begin{align*}
A & \rightarrow D \\
AB & \rightarrow E \\
D & \rightarrow C \\
D & \rightarrow F \\
AE & \rightarrow G \\
CF & \rightarrow D
\end{align*}
\]

Optional, but strong hint: derive the cover of the above functional dependencies.

(a) [3 points] Is the decomposition $\{ACF, ABEG, AD\}$ lossless?
   □ Yes □ No

(b) [4 points] Is the decomposition $\{DCF, ABEG, AD\}$ lossless?
   □ Yes □ No

(c) [4 points] Is the decomposition $\{ABDE, BEG, ADCF\}$ lossless?
   □ Yes □ No

(d) [3 points] Is the decomposition $\{ACF, ABEG, AD\}$ dependency preserving?
   □ Yes □ No

(e) [3 points] Is the decomposition $\{DCF, ABEG, AD\}$ dependency preserving?
   □ Yes □ No

(f) [3 points] Is the decomposition $\{ABDE, BEG, ADCF\}$ dependency preserving?
   □ Yes □ No
Question 4: Normal Forms .......................... [30 points]
Submit on separate page
Course: 15-415/615; HW: ; Q:
Name: _____________________; andrew-id: _____________________; late days:

Consider the relation schema \( r = \{P, Q, R, S, T, U, V\} \) and the functional dependencies FD:

\[
\begin{align*}
PR & \rightarrow S \\
P & \rightarrow T \\
PT & \rightarrow R \\
S & \rightarrow U \\
ST & \rightarrow V \\
TV & \rightarrow S \\
QT & \rightarrow V \\
V & \rightarrow Q
\end{align*}
\]

Consider the relational schemas:

- \( r_1 = \{P, R, S, T\} \)
- \( r_2 = \{Q, T, V\} \)
- \( r_3 = \{S, T, U, V\} \)

(a) [2 points] What is the projection of the FDs on \( r_1 \)?

(b) [2 points] Indicate all the candidate key(s) for \( r_1 \):

- \( \{P\} \)
- \( \{PR\} \)
- \( \{PRT\} \)
- \( \{PR\} \) and \( \{PT\} \)
- Other: ____________

(c) [3 points] Is \( r_1 \) 3NF? \( \square \) Yes \( \square \) No

(d) [3 points] Is \( r_1 \) BCNF? \( \square \) Yes \( \square \) No

(e) [2 points] What is the projection of the FDs on \( r_2 \)?

(f) [2 points] Indicate all the candidate key(s) for \( r_2 \):

- \( \{Q\} \) and \( \{T\} \)
- \( \{QT\} \)
- \( \{TV\} \)
- \( \{QT\} \) and \( \{TV\} \)
- \( \{QT\} \) and \( \{QV\} \)
- Other: ____________

(g) [3 points] Is \( r_2 \) 3NF? \( \square \) Yes \( \square \) No

---

Question 4 continues...
(h) [3 points] Is $r_2$ BCNF? □ Yes □ No

(i) [2 points] What is the projection of the FDs on $r_3$?

(j) [2 points] Is $r_3$ 3NF? □ Yes □ No

(k) [2 points] Is $r_3$ BCNF? □ Yes □ No

(l) [3 points] Decompose $r_3$ to two relational schemas $r_{3,1}$ and $r_{3,2}$ so that they are in 3NF, and the decomposition is lossless and dependency preserving. Give those relational schemas.

(l) __________

(m) [1 point] Yes/No: is it possible to decompose $r_3$ into two BCNF schemas $r'_{3,1}$ and $r'_{3,2}$, with a lossless and dependency-preserving decomposition?
□ Yes □ No