Carnegie Mellon University
Department of Computer Science
15-415/615- Database Applications
C. Faloutsos & A. Pavlo, Fall 2015
Prepared by Yujing Zhang
DUE DATE: Wednesday, 11/04/2015, 3:00pm

Homework 6

IMPORTANT

• Deposit hard copy of your answers in class at 3:00pm on Wednesday, 11/04/2015.
• Separate answers, as usually, i.e., please solve each of the 5 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 your answers.
• Late homeworks: Follow usual policy: 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; 5 questions total
• Rough time estimate: ≈2-5 hours (0.5-1 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>20</td>
<td></td>
</tr>
<tr>
<td>Functional Dependencies I</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Functional Dependencies II</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 . . . . . . . . . . . . . . . . . . . . . [20 points]
Submit on separate page

Course: 15-415/615; HW: ; Q:  
Name: _________________; andrew-id: _______________________; late days:

For this problem we consider the yelp reviews database with following three tables (slightly simplified from Homework 2, for your convenience):

1. Business(bid, name, city, state), where the primary key is bid;
2. yelp_user(uid, name), where the primary key is uid;
3. Review(bid, uid, stars, date), where the primary key is (bid, uid), bid is foreign key referencing Business, and uid is foreign key referencing yelp_user.

For these tables we are given the following statistics (also, rounded-off with respect to Homework 2, for your convenience):

- Business consists of \( N_1 = 60,000 \) tuples, and there are:
  - \( V(Business, name) = 45,000 \) distinct business names.
  - \( V(Business, city) = 400 \) distinct cities.
  - \( V(Business, state) = 30 \) distinct states.
- yelp_user consists of \( N_2 = 360,000 \) tuples, and there are:
  - \( V(yelp_user, name) = 40,000 \) distinct user names.
- Review consists of \( N_3 = 1,600,000 \) tuples, and there are:
  - \( V(Review, uid) = 360,000 \) distinct uid’s.
  - \( V(Review, bid) = 60,000 \) distinct bid’s.
  - \( V(Review, stars) = 5 \) distinct star ratings (i.e. 1, 2, 3, 4, and 5), without nulls.

For the queries below, assume that there are no correlations between the columns of a table nor any prior knowledge about the data (i.e., assume uniform distribution). Estimate the number of resulting tuples for the query, and give the answer with *** fourth *** significant digit accuracy. We will accept either rounding half up or down, but no partial credit will be given.

(a) [2 points] SELECT * FROM Business WHERE city = ‘Pittsburgh’;

(a) __________

(b) [3 points] SELECT * FROM Business WHERE state = ‘PA’ AND name = ‘McDonald’;

(b) __________

(c) [3 points] SELECT * FROM Review WHERE stars > 3;

(c) __________

(d) [3 points] SELECT city, count(*) FROM Business GROUP BY city;

(d) __________

Question 1 continues...
(e) [4 points] SELECT * FROM Business JOIN Review ON Business.bid = Review.bid WHERE state = 'PA';

(f) [5 points] SELECT * FROM Review as R1 JOIN Review as R2 ON R1.bid = R2.bid;

The query returns all pairs of users (and more info), that have reviewed the same business. (For your ease of computation, the query reports mirror- and self- pairs.)
Question 2: Functional Dependencies I ............... [10 points]
Submit on separate page
Course: 15-415/615; HW: ; Q:
Name: __________________; andrew-id: __________________; late days:

Consider the following legal instance of a relational schema $S$ with attributes $ABC$:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

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

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

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

ii. [2 points] ☐ Yes  ☐ No : $B \rightarrow A$ is violated.

iii. [2 points] ☐ Yes  ☐ No : $C \rightarrow A$ is violated.

iv. [2 points] ☐ Yes  ☐ No : $BC \rightarrow A$ is violated.

v. [2 points] ☐ Yes  ☐ No : $AC \rightarrow B$ 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

Homework 6 continues...
Question 3: Functional Dependencies II .............. [20 points]

Submit on separate page

Course: 15-415/615; HW: ; Q: 
Name: __________________; andrew-id: __________________; late days:

For the next set of questions consider the relational schema \( R = \{A, B, C, D, E, F, G, H\} \) and the set of functional dependencies \( \text{FD:} \)

\[
\begin{align*}
A & \rightarrow B \\
B & \rightarrow C \\
AD & \rightarrow CEF \\
BE & \rightarrow FG \\
CF & \rightarrow GH \\
G & \rightarrow H
\end{align*}
\]

(a) [6 points] Which of the following is a minimum cover of the FD? If none, mark accordingly, and give your own answer.

i. The given FDs (Eq 1-6), is a minimum cover already.

ii. \( \{A \rightarrow B, B \rightarrow C, AD \rightarrow E, AD \rightarrow F, BE \rightarrow F, BE \rightarrow G, CF \rightarrow G, G \rightarrow H\} \)

iii. \( \{A \rightarrow B, B \rightarrow C, AD \rightarrow E, BE \rightarrow F, CF \rightarrow G, G \rightarrow H\} \)

iv. \( \{A \rightarrow B, B \rightarrow C, AD \rightarrow E, AD \rightarrow F, BE \rightarrow F, CF \rightarrow G, G \rightarrow H\} \)

v. 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)-(6))? 

i. [3 points] □ Yes □ No : \( A \rightarrow C \)

ii. [3 points] □ Yes □ No : \( AE \rightarrow F \)

iii. [3 points] □ Yes □ No : \( CE \rightarrow F \)

iv. [3 points] □ Yes □ No : \( BDE \rightarrow AG \)

(c) [1 point] True or False: The attribute closure \( \{B\}^+ \) is \( \{B, C, F\} \).

□ True □ False

(d) [1 point] True or False: The attribute closure \( \{AD\}^+ \) is \( \{A, B, C, D, E, F, G\} \).

□ True □ False

Homework 6 continues...
Question 4: 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 relation with attributes, $X = \{P, Q, R, S, T\}$, Let the following functional dependencies $FD$ be defined over the relation $X$:

$$P \rightarrow Q$$
$$Q \rightarrow R$$
$$S \rightarrow T$$

(a) [2 points] Provide the attribute closure of $\{PS\}$.

(b) Consider the decomposition $PQR, ST$. Mark 'True' or 'False':
   
   i. [1 point] □ True □ False : It is lossless
   ii. [1 point] □ True □ False : It is dependency-preserving
   iii. [2 points] □ True □ False : All tables of the decomposition, are in 3NF or higher
   iv. [2 points] □ True □ False : All tables of the decomposition, are in BCNF

(c) Consider the decomposition $PQ, QR, ST$. Mark 'True' or 'False':
   
   i. [1 point] □ True □ False : It is lossless
   ii. [1 point] □ True □ False : It is dependency-preserving
   iii. [2 points] □ True □ False : All tables of the decomposition, are in 3NF or higher
   iv. [2 points] □ True □ False : All tables of the decomposition, are in BCNF

(d) Consider the decomposition $PQR, PS, ST$. Mark 'True' or 'False':
   
   i. [1 point] □ True □ False : It is lossless
   ii. [1 point] □ True □ False : It is dependency-preserving
   iii. [2 points] □ True □ False : All tables of the decomposition, are in 3NF or higher
   iv. [2 points] □ True □ False : All tables of the decomposition, are in BCNF
Consider the relation with attributes, $E = \{U, V, W, X, Y, Z\}$. Suppose that the following functional dependencies hold:

\[
\begin{align*}
    U & \rightarrow VW \\
    WX & \rightarrow Z \\
    V & \rightarrow X \\
    V & \rightarrow Y \\
    Z & \rightarrow U
\end{align*}
\]

(a) [6 points] List all the candidate key(s) for $E$. A, possibly correct, answer may be: “{$UV$} and {$UW$}”.

(b) [2 points] Is the relation $E$ in BCNF? □ Yes □ No

(c) [3 points] Justify: Explain why $E$ is (or is not) in BCNF. Your answer should follow the style, e.g.: “all FDs follow the rules of BCNF” or “FD (11) violates the rules: ‘Z’ is a determinant, but not a candidate key”

(d) [2 points] Is the relation $E$ in 3NF? □ Yes □ No

(e) [3 points] Justify: Explain why $E$ is (or is not) in 3NF. Follow the style that we mentioned above.

(f) [6 points] Give a 3NF decomposition of $E$ that is lossless, dependency preserving, and has as few tables as possible.

(g) [8 points] Give a BCNF decomposition of $E$ that is lossless, and has as few tables as possible.