Carnegie Mellon University
Department of Computer Science
15-415/615- Database Applications
C. Faloutsos & A. Pavlo, Spring 2014
Prepared by Alex Beutel
DUE DATE: Tue, 3/25/2014, 1:30pm

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)

Revision : 2014/04/02 01:00

Question	Points	Score
Query Optimization	30	
Functional Dependencies	20	
Decompositions	20	
Normal Forms	30	
Total:	100	

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. **555.5555**

Solution: (Optional) justification: $N_1/90$

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. **0.018518**

Question 1 continues. . .

Solution: (Optional) justification: $N_1/90/30,000$

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

SELECT * **FROM** `Movies` **WHERE** `year` $>$ 1960;

iii. **30000**

Solution: (Optional) justification: $N_1 \frac{54}{90}$
Because not well specified, will also accept $30555.5555 = N_1 \frac{55}{90}$

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

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

iv. **100000**

Solution: (Optional) justification: `actorID` is a primary key in `Actors` and a foreign key in `Acted_in`. Therefore, $N_2 \cdot N_3/200000$

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

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

v. **18518.5185 or 1,000,000**

Solution: (Optional) justification: There are two ways to view this question. If you assume that (title,year) is the primary key of Movies and are foreign keys in Acted_in then the answer is 1,000,000. However, we do not explicitly make these assumptions so the estimated number of tuples could be calculated as $N_1 \cdot N_3 \cdot \frac{1}{90} \cdot \frac{1}{30000}$.

Question 2: Functional Dependencies . . . . . [20 points]

Submit on separate page
Course: 15-415/615; HW: ; Q:
Name: _____; andrew-id: _____; late days:

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:

S

A	B	C
1	a	X
4	a	Y
5	b	X

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** : $A \to B$ is violated.
ii. **[1 point]** ■ **Yes** ☐ No : $B \to A$ is violated.
iii. **[1 point]** ☐ Yes ■ **No** : $BC \to A$ is violated.
iv. **[1 point]** ■ **Yes** ☐ No : $B \to C$ is violated.
v. **[1 point]** ■ **Yes** ☐ No : $C \to AB$ 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**

Solution: No, because we can only see an instance.

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:

$$P \to S \quad (1)$$
$$PQ \to ST \quad (2)$$
$$S \to RU \quad (3)$$
$$RU \to S \quad (4)$$
$$PT \to V \quad (5)$$

(a) **[3 points]** Which of the following is a minimum cover of the FD?
(a) The given FD is a minimum cover.
(b) $\{P \to S; PQ \to T; PQ \to S; S \to R; S \to U; PT \to V; RU \to S\}$
(c) $\{P \to R; P \to U; PQ \to T; PT \to V\}$
(d) $\{P \to S; PQ \to T; S \to R; S \to U; PT \to V; RU \to S\}$
(e) none of the above - the cover is _____

Solution: (d)

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

i. **[1 point]** ■ **Yes** □ No : $P \rightarrow U$

ii. **[2 points]** ■ **Yes** □ No : $PT \rightarrow SV$

iii. **[1 point]** □ Yes ■ **No** : $SQ \rightarrow V$

iv. **[1 point]** □ Yes ■ **No** : $PS \rightarrow RV$

v. **[1 point]** ■ **Yes** □ No : $PQ \rightarrow V$

vi. **[1 point]** □ Yes ■ **No** : $PSRU \rightarrow QT$

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

Solution: It should include P, ie., $\{P, R, S, U\}$.

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

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 $S = \{A, B, C, D, E, F, G\}$:

$$A \to D$$
$$AB \to E$$
$$D \to C$$
$$D \to F$$
$$AE \to G$$
$$CF \to D$$

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

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

Solution: Optional Justification: A is the candidate key in AD and ACF

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

Solution: Yes: D is a candidate key in DCF, for the join AD and DCF; and then A is the candidate key in $ADCF$, for the join with $ABEG$

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

Solution: No: while $ADCF$ and $ABDE$ can be joined on A which is a candidate key for $ADCF$, the joining attributes BE are not a candidate key in either BEG, nor $ABDCFE$

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

Solution: We lost both $CF \to D$, as well as $D \to CF$

(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**

Solution: We lost $AE \to G$

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:

$$\begin{aligned} PR &\rightarrow S \\ P &\rightarrow T \\ PT &\rightarrow R \\ S &\rightarrow U \\ ST &\rightarrow V \\ TV &\rightarrow S \\ QT &\rightarrow V \\ V &\rightarrow Q \end{aligned}$$

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?

Solution: $\{PR \rightarrow S, P \rightarrow T, PT \rightarrow R\}$

(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? ■ **Yes** □ No

(d) **[3 points]** Is r_1 BCNF? ■ **Yes** □ No

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

Solution: $\{QT \rightarrow V, V \rightarrow Q\}$

(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? ■ **Yes** □ No

(h) **[3 points]** Is r_2 BCNF. □ Yes ■ **No**

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

Solution: $\{S \to U, ST \to V, TV \to S\}$

(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) $\{S, U\}$, $\{S, T, V\}$

(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

Solution: The earlier answer, SU and STV, are all in BCNF