Carnegie Mellon University
Department of Computer Science
15-415/615- Database Applications
C. Faloutsos & A. Pavlo, Fall 2016

Homework 6 (by Prashanth Menon) - Solutions

Due: **hard copy, in class** at **3:00pm, on Wednesday, Nov. 2/2016**

IMPORTANT

- **Deposit hard copy** of your answers in **class at 3:00pm on Nov. 2/2016**.
- 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)

Revision : 2016/11/11 18:23

Question	Points	Score
Query Optimization	20	
Functional Dependencies I	10	
Functional Dependencies II	20	
Decompositions	20	
Normal Forms	30	
Total:	100	

Question 1: Query Optimization . . . . . [20 points]

GRADED BY: Huanchen Zhang

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

For this problem we consider a modified and simplified version of the bike sharing database from Homework 2. The bike sharing database has the following three tables:

1. `Bike(bid, model, year)`, where the primary key is `bid`;
2. `Station(sid, name, city, zip_code)`, where the primary key is `sid`;
3. `Trip(tid, date, start_sid, end_sid, bid)`, where the primary key is `tid`, where `start_sid` and `end_sid` are foreign keys to `Station`, and where `bid` is a foreign key to `Bike`.

For these tables we are given the following statistics:

- `Bike` consists of $N_1 = 50,000$ tuples, and there are:
  - $V(Bike, model)$= 4000 distinct bike models.
  - $V(Bike, year)$= 20 distinct years, 1997-2016 inclusive.
- `Station` consists of $N_2 = 24,000$ tuples, and there are:
  - $V(Station, name) = 4000$ distinct station names.
  - $V(Station, city) = 600$ distinct cities.
  - $V(Station, zip\_code) = 50$ distinct station zip codes.
- `Trip` consists of $N_3 = 2,000,000$ tuples, and there are:
  - $V(Trip, start\_sid)$= 24,000 distinct `start_sid`'s.
  - $V(Trip, end\_sid)$= 24,000 distinct `end_sid`'s.
  - $V(Trip, bid)$= 15,000 distinct `bid`'s.
  - $V(Trip, date)$: not needed - not given.

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. E.g., if the answer is $1,234,567$, you may round it to $1,234,000$. We will accept either rounding half up or down.

(a) **[3 points]** **SELECT * FROM** `Bike` **WHERE** `year` $>$ 2004 **AND** `year` $<=$ 2009;

(a) **12500.0**

Solution: $N_1 \cdot 0.25$

Grading info: -3 for any incorrect answer

(b) **[2 points]** **SELECT * FROM** `Bike` **WHERE** `model` = 'Super Bike';

(b) **12.5**

Solution: $N_1/V(Bike, model)$

Grading info: -2 for any incorrect answer

(c) **[3 points]** **SELECT** * **FROM** `Station` **WHERE** `city` = 'Pittsburgh' **AND** `name` = 'Carnegie Mellon University';

(c) **0.01**

Solution: $N_2/V(Station, city)/V(Station, name)$

Grading info: -1.5 for off-by-factor-of-10 error, -3 for any other incorrect answer

(d) **[3 points]** **SELECT** `zip_code`, `count(*)` **FROM** `Station` **GROUP BY** `zip_code`;

(d) **50.0**

Solution: V(Station, zipcode)

Grading info: -3 for any incorrect answer

(e) **[4 points]** **SELECT** * **FROM** `Trip` **JOIN** `Bike` **ON** `Trip.bid` = `Bike.bid` **WHERE** `year` > 2014;

(e) **200,000.0**

Solution: $N_3 \cdot 0.1$

Grading info: -4 for any other incorrect answer

(f) **[5 points]** **SELECT** `Station.city`, `count(*)` **FROM** `Trip` **JOIN** `Station` **ON** `Station.sid` = `Trip.start_sid` **GROUP BY** `Station.city`;

(f) **600.0**

Solution: $V(Station, city)$

Grading info: -2 for off-by-factor-of-10 error, -5 for any incorrect answer

Question 2: Functional Dependencies I . . . . . [10 points]

GRADED BY: Jiexi Lin

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:

S	A	B	C
	a	10	T
	b	20	T
	c	20	F
	c	30	F

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 : $A \to B$ is violated.

ii. **[2 points]** ■ **Yes** □ No : $B \to A$ is violated.

iii. **[2 points]** ■ **Yes** □ No : $C \to A$ is violated.

iv. **[2 points]** ■ **Yes** □ No : $AC \to B$ is violated.

v. **[2 points]** □ Yes ■ **No** : $BC \to A$ 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.

Question 3: Functional Dependencies II ..... [20 points]

GRADED BY: Prashanth Menon

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 $\mathcal{R} = \{P, Q, R, S, T, U, V, W\}$ and the set of functional dependencies FD:

$$W \rightarrow U \quad (1)$$
$$RU \rightarrow VW \quad (2)$$
$$PQ \rightarrow ST \quad (3)$$
$$P \rightarrow R \quad (4)$$
$$SV \rightarrow WU \quad (5)$$
$$R \rightarrow S \quad (6)$$

(a) **[6 points]** Which of the following is a minimum cover of the FD? Mark all that qualify; if none, mark accordingly, and give your *own.* answer.

i. The given FDs (Eq 1-6), is a minimum cover already.
ii. $\{W \rightarrow U, RU \rightarrow V, PQ \rightarrow T, P \rightarrow R, SV \rightarrow W, R \rightarrow S\}$
iii. $\{W \rightarrow U, RU \rightarrow V, PQ \rightarrow S, PQ \rightarrow T, P \rightarrow R, SV \rightarrow U, R \rightarrow S\}$
iv. $\{W \rightarrow U, RU \rightarrow V, RU \rightarrow W, PQ \rightarrow S, P \rightarrow R, SV \rightarrow W, R \rightarrow S\}$
v. none of the above - the cover is _____

Solution: ii

(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 : $PV \rightarrow W$
ii. **[3 points]** □ Yes ■ **No** : $R \rightarrow T$
iii. **[3 points]** ■ **Yes** □ No : $P \rightarrow S$
iv. **[3 points]** □ Yes ■ **No** : $PQR \rightarrow U$

(c) **[1 point]** True or False: The attribute closure $\{P\}^+$ is $\{R\}$.
□ True ■ **False**
Grading info: *It is* $\{P, R, S\}$.

(d) **[1 point]** True or False: The attribute closure $\{PQ\}^+$ is $\{P, Q, R, S, T\}$.
■ **True** □ False

Question 4: Decompositions ..... [20 points]

GRADED BY: Jiexi Lin

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, $\mathcal{X} = \{A, B, C, D, E, F\}$, Let the following functional dependencies FD be defined over the relation $\mathcal{X}$:

$$A \rightarrow C$$
$$BC \rightarrow D$$
$$F \rightarrow E$$

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

Solution: $\{AB\}^+ = \{ABCD\}$

(b) Consider the decomposition ABC, $CDEF$. Mark 'True' or 'False':

i. **[3 points]** ☐ True ■ **False** : It is lossless

ii. **[3 points]** ☐ True ■ **False** : It is dependency-preserving

(c) Consider the decomposition $ABCD$, $ABDF$, EF. Mark 'True' or 'False':

i. **[3 points]** ■ **True** ☐ False : It is lossless

ii. **[3 points]** ■ **True** ☐ False : It is dependency-preserving

(d) Consider the decomposition $ABCDF$, EF. Mark 'True' or 'False':

i. **[3 points]** ■ **True** ☐ False : It is lossless

ii. **[3 points]** ■ **True** ☐ False : It is dependency-preserving

Question 5: Normal Forms ..... [30 points]

GRADED BY: Lu Zhang

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

Consider the relation with attributes, $\mathcal{E} = \{W, X, Y, Z\}$. Suppose that the following functional dependencies hold:

$$XY \rightarrow W \quad (7)$$
$$X \rightarrow Z \quad (8)$$
$$W \rightarrow X \quad (9)$$

(a) **[6 points]** List *all* the candidate key(s) for $\mathcal{E}$. A, possibly correct, answer may be: "$\{WX\}$ and $\{YZ\}$".

Solution: $\{WY\}$, $\{XY\}$

Grading info: -3: for each missing candidate key

(b) **[1 point]** Is the relation $\mathcal{E}$ in BCNF? □ Yes ■ **No**

(c) From the list below, select all applicable choices to justify whether $\mathcal{E}$ is (or is not) in BCNF.

Note: when we refer to the *main requirement* for BCNF, we mean: *every determinant is a super key.*

i. **[1 point]** □ True ■ **False** : All FD's satisfy the main requirement.
ii. **[1 point]** □ True ■ **False** : FD (7) violates the main requirement.
iii. **[1 point]** ■ **True** □ False : FD (8) violates the main requirement.
iv. **[1 point]** ■ **True** □ False : FD (9) violates the main requirement.

(d) **[1 point]** Is the relation $\mathcal{E}$ in 3NF? □ Yes ■ **No**

(e) From the list below, select all applicable choices to justify whether $\mathcal{E}$ is (or is not) in 3NF.

Note: when we refer to the *secondary requirement* for 3NF, we mean: *for every FD $X \rightarrow A$, A is part of a candidate key.*

i. **[1 point]** □ True ■ **False** : All FD's satisfy the secondary requirement.
ii. **[1 point]** □ True ■ **False** : FD (7) violates the secondary requirement.
iii. **[1 point]** ■ **True** □ False : FD (8) violates the secondary requirement.
iv. **[1 point]** □ True ■ **False** : FD (9) violates the secondary requirement.

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

Solution: $\mathcal{E}_{1,1}=(X,Z)$, $\mathcal{E}_{1,2}=(W,X,Y)$.

Grading info: -3: 3NF decomposition more than with 2 tables; -4: 3NF decomposition with more than 5 tables

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

Solution: $\mathcal{E}_{1,1}=(W,X)$, $\mathcal{E}_{1,2}=(W,Y)$, $\mathcal{E}_{1,3}=(X,Z)$

Grading info: -5: BCNF decomposition with more than 3 tables