Carnegie Mellon University
Department of Computer Science
15-415/615- Database Applications
C. Faloutsos & A. Pavlo, Spring 2015
Prepared by Vinay & Jiayu
DUE DATE: Tuesday, 3/24/2015, 1:30pm

Homework 6

IMPORTANT

- **Deposit hard copy** of your answers in **class at 1:30pm on Tuesday, 3/24/2015**.
- 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 : 2015/04/06 12:25

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

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

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

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

1. `Movies(title, imdb_number, year)`, where the primary key is `imdb_number`;
2. `Reviewers(userid, username)`, where the primary key is `userid`;
3. `Reviews(userid, imdb_number, rating, comment)`, where the primary key is `(userid, imdb_number)`, `userid` is foreign key referencing `Reviewers`, and `imdb_number` is foreign key referencing `Movies`.

For these tables we know the following statistics:

- `Movies` consists of $N_1 = 60,000$ tuples, there are:
  - 40,000 distinct movie titles,
  - 90 distinct years, 1925–2014 inclusive.
- `Reviewers` consists of $N_2 = 50,000$ tuples, there are:
  - 50,000 distinct userids,
  - 50,000 distinct usernames
- `Reviews` consists of $N_3 = 1,300,000$ tuples, there are:
  - 38,000 distinct userids,
  - 33,000 distinct movie titles,
  - 5 distinct 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) **[3 points]** **SELECT * FROM** `Movies` **WHERE** `year` = 2001;

(a) **666.7**

Solution: $N_1/90$

(b) **[4 points]** **SELECT * FROM** `Movies` **WHERE** `year` = 1999 **AND** `title` = "Fight Club"

(b) **0.01667**

Solution: $N_2/90/40000$

(c) **[4 points]** **SELECT * FROM** `Reviews` **WHERE** `rating` > 3

(c) **520000**

Solution: $N_3 \cdot 0.4$

(d) **[4 points]** **SELECT** `year`, `count(*)` **FROM** `Movies` **GROUP BY** `year`

(d) **90**

Solution: number of years

(e) **[5 points]** **SELECT** `count(*)`
FROM `Movies` **JOIN** `Reviews` **ON** `Movies.imdb_number` = `Reviews.imdb_number`
GROUP BY `Movies.title`

(e) **33000**

Solution: number of distinct movie titles

note: No partial points were given on this question

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

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

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

S

A	B	C
α	9	T
α	16	F
β	20	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 : $A \rightarrow B$ is violated.
ii. **[1 point]** □ Yes ■ **No** : $B \rightarrow A$ is violated.
iii. **[1 point]** ■ **Yes** □ No : $C \rightarrow A$ is violated.
iv. **[1 point]** □ Yes ■ **No** : $AC \rightarrow B$ is violated.
v. **[1 point]** □ Yes ■ **No** : $B \rightarrow AC$ 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:

$$P \rightarrow Q \quad (1)$$
$$Q \rightarrow R \quad (2)$$
$$PS \rightarrow TRV \quad (3)$$
$$QT \rightarrow UR \quad (4)$$
$$S \rightarrow V \quad (5)$$

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

(a) The given FD is a minimum cover.
(b) $\{P \rightarrow Q, Q \rightarrow R, PS \rightarrow T, QT \rightarrow UR, S \rightarrow V\}$
(c) $\{P \rightarrow Q, Q \rightarrow R, P \rightarrow T, Q \rightarrow U, S \rightarrow V\}$
(d) $\{P \rightarrow Q, Q \rightarrow R, PS \rightarrow T, QT \rightarrow U, S \rightarrow V\}$
(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 : $P \rightarrow R$

ii. **[2 points]** ■ **Yes** □ No : $PS \rightarrow U$

iii. **[2 points]** □ Yes ■ **No** : $QS \rightarrow U$

iv. **[2 points]** □ Yes ■ **No** : $QST \rightarrow P$

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

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

Question 3: Decompositions ..... [30 points]

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

3.1 For this set of questions, Consider the relation with attributes, $\mathcal{S} = \{A, B, C, D, E, F\}$, Let the following functional dependencies FD be defined over the relation $\mathcal{S}$:

$$A \to D$$
$$A \to E$$
$$D \to C$$
$$D \to F$$

(a) **[6 points]** Provide the attribute closure of $\{AB\}$.

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

(b) Identify whether the decomposition ABC, CDE, EFA is lossless and dependency-preserving?

i. **[3 points]** □ Yes ■ **No** : Lossless?

Solution: Not lossless because the intersection of CDE and EFA (i.e E) is not the candidate key for either of the relations.

ii. **[3 points]** □ Yes ■ **No** : Dependency Preserving?

Solution: Not dependency preserving because $A \to D$ is not present in the closure of the unions of projection of FD on the three decomposed relations.

(c) Identify whether the decomposition $ABCE$, ADC, $ADEF$ is lossless and dependency-preserving?

i. **[3 points]** ■ **Yes** □ No : Lossless?

Solution: Lossless join!

ii. **[3 points]** ■ **Yes** □ No : Dependency Preserving?

Solution: Dependency preserving!

3.2 Consider the relation with attributes, $\mathcal{A} = \{P, Q, R, S, T, U, V\}$. The following

functional dependencies hold over $\mathcal{A}$.

$$RT \to P$$
$$QS \to T$$
$$R \to Q$$

(a) Identify whether the decomposition QST, QR, $PRSUV$ is lossless and dependency-preserving?

i. **[3 points]** ■ **Yes** □ No : Lossless?

Solution: Lossless join!

ii. **[3 points]** □ Yes ■ **No** : Dependency Preserving?

Solution: Not dependency preserving

(b) Identify whether the decomposition QST, QR, $PRSUV$, PRT is lossless and dependency-preserving?

i. **[3 points]** ■ **Yes** □ No : Lossless?

Solution: Lossless join!

ii. **[3 points]** ■ **Yes** □ No : Dependency Preserving?

Solution: Dependency preserving!

Question 4: Normal Forms ..... [30 points]

Submit on separate page
Course: 15-415/615; HW: ; Q:
Name: ____; andrew-id: ____; late days: Graded by: Hong Bin Shim

Consider the relation with attributes, $\mathcal{S} = \{A, B, C, D, E\}$. Let the following functional dependencies be defined over the relation $\mathcal{S}$,

$$\begin{aligned} A &\rightarrow BC \\ CD &\rightarrow E \\ B &\rightarrow D \\ E &\rightarrow A \end{aligned}$$

(a) Identify whether this relationship in 3NF and/or BCNF?

i. **[2 points]** ■ **Yes** □ No : 3NF

ii. **[2 points]** □ Yes ■ **No** : BCNF

(b) **[5 points]** Provide the projection of FDs for the subset of attributes $\{ABD\}$

Solution: $A \rightarrow B$, $B \rightarrow D$

(c) **[8 points]** Give a BCNF decomposition of $\mathcal{S}$ that is lossless.

Solution: (A, B, C, E) and (B, D)

(d) **[3 points]** Is your BCNF decomposition dependency preserving?

□ True ■ **False**

(e) **[10 points]** Give a 3NF decomposition of $\mathcal{S}$ that is lossless and dependency preserving.

Solution: $(A, B, C), (C, D, E), (B, D), (E, A)$
I gave this one for free