CMU - SCS
15-415/15-615 Database Applications
Spring 2013, C. Faloutsos
Homework 6: Query Optimization + Schema Refinement
Deadline: 1:30pm on Tuesday, 3/26/2013

Reminders - IMPORTANT:

- Like all homeworks, it has to be done **individually**.
- Please **typeset** your answers.
- Please submit your answers in **hard copy, in class**, 1:30pm, on Tuesday, 03/26/2013 .
- As before, for ease of grading, please solve each of the three questions on a **separate** page, and type your **name and andrew ID on each** of the three pages.

Reminders - FYI:

- Weight: 5% of homework grade.
- The points of this homework add up to 100.
- Rough time estimates: 2-4 hours.

Q1. Query Optimization - Selectivities, 30 pts - SUBMIT ON SEPARATE PAGE

Consider the tables `WORKS_AT(SSN, gymID)` and `GYM(gymID, name)`. *Notice that* `gymID` *is* not *a candidate key for the table* `GYM`.

`WORKS_AT(SSN, gymID)` consists of $N_1 = 100,000$ tuples and has

- V(`SSN`,`WORKS_AT`) = 50,000 distinct values of SSN
- V(`gymID`,`WORKS_AT`) = 20,000 distinct values of gymID.

`GYM(gymID, name)` consists of $N_1 = 40,000$ tuples and has

- V(`gymID`,`GYM`) = 20,000 distinct values of gymID
- V(`name`,`GYM`) = 30,000 distinct values of name.

For all the computations below:

- Please give numerical answers, accurate up to the **fourth** significant digit.
- No need for explanations, unless explicitly requested.

Q1.1 [5 pts] Estimate the number of qualifying tuples of the query:

```
SELECT *
  FROM WORKS_AT
  WHERE SSN = 123456789;
```

Q1.2 [5 pts] Can `SSN` be a candidate key for the table `WORKS_AT`? Give a short explanation for your answer.

Q1.3 [5 pts] Estimate the number of qualifying tuples of the query:

```
SELECT *
  FROM GYM
  WHERE name = "Gym planet";
```

Q1.4 [5 points] Estimate the number of qualifying tuples of the query:

```
SELECT *
  FROM WORKS_AT
  WHERE SSN = 123456789 AND gymID=101;
```

Q1.5 [5 points] Notice that `gymID` is *not* a candidate key for the table `GYM`. Estimate the number of qualifying tuples of the query:

```
SELECT SSN, GYM.gymID, name
  FROM WORKS_AT JOIN GYM
  WHERE GYM.gymID = WORKS_AT.gymID;
```

Q1.6 [5 points] Estimate the number of qualifying tuples of the query:

```
SELECT WA1.SSN, WA2.SSN
  FROM WORKS_AT AS WA1 JOIN WORKS_AT AS WA2
  WHERE WA1.gymID = WA2.gymID;
```

Q2. Functional Dependencies, 30pts - SUBMIT ON SEPARATE PAGE

Q2.1 Consider the relation schema R = {P,Q,S,T,U,V} and the set of functional dependencies FD =

$$PQ \rightarrow S \quad (1)$$
$$PS \rightarrow Q \quad (2)$$
$$PT \rightarrow U \quad (3)$$
$$Q \rightarrow T \quad (4)$$
$$QS \rightarrow P \quad (5)$$
$$U \rightarrow V \quad (6)$$

Answer the following questions. Notice that:

- For Yes/No or True/False questions, you may just give binary answers. Explanations are optional and will be used for partial credit. Wrong answers, with no, or wrong explanations, will get **negative** points.
- For the rest of the questions, please give *short* justifications.

2.1.1 [2 pts] Yes/No. Is FD a minimum cover?
2.1.1 [4 pts] Yes/No. Is the decomposition {PQ, QS, PQTU, UV} lossless?
2.1.2 [4 pts] Somebody claims that the decomposition {PQ, QS, PQTU, UV} is not dependency-preserving. If you agree with the statement, give all the missing dependencies. If you disagree, just state so.
2.1.3 [5 pts] Yes/No. Is the decomposition {PQS, PSTU, PTV} lossless?
2.1.4 [1 pts] True/False. The decomposition {PQS, PSTU, PTV} is not dependency-preserving, because it does not preserve U $\rightarrow$ V.
2.1.5 [2 pts] True/False. The decomposition {PQS, PSTU, PTV} is not dependency-preserving, because it does not preserve U $\rightarrow$ V nor Q $\rightarrow$ T.
2.1.6 [2 pts] True/False. The decomposition {PQS, PSTU, PTV} is dependency-preserving.

Q2.2 Consider now the same relation schema R = {P,Q,S,T,U,V} with **different**, simpler, set of functional dependencies FD' =

$$Q \rightarrow ST \quad (7)$$
$$P \rightarrow T \quad (8)$$
$$PS \rightarrow T \quad (9)$$
$$QU \rightarrow V \quad (10)$$

Answer the following questions. Again, **negative** points for wrong, binary answers.

2.2.1 [1 pts] True/False. The attribute closure $\{P\}^+$ is {P,S,T}.
2.2.2 [1 pts] True/False. The attribute closure $\{P\}^+$ is {P,T}.

2.2.3 [1 pts] True/False. The attribute closure $\{P,Q\}^+$ is $\{P,T,Q,S\}$.
2.2.4 [1 pts] True/False. The attribute closure $\{P,Q\}^+$ is $\{P,S,T\}$.
2.2.4 [1 pts] True/False. The attribute closure $\{P,Q\}^+$ is $\{P,T,Q,S,U,V\}$.
2.2.5 [1 pts] True/False. The dependency $Q \rightarrow S$ can be deduced from FD'.
2.2.6 [2 pts] True/False. The dependency $QU \rightarrow TUV$ can be deduced from FD'.
2.2.7 [2 pts] True/False. All the candidate keys of R are $\{P,Q\}$.

Q3. BCNF and 3NF, 40pts - SUBMIT ON SEPARATE PAGE

Consider the relation schema R = {P,Q,S,T,U,V} and the functional dependencies FD =

$$PQ \rightarrow S \tag{11}$$
$$PS \rightarrow Q \tag{12}$$
$$PT \rightarrow U \tag{13}$$
$$Q \rightarrow T \tag{14}$$
$$QS \rightarrow P \tag{15}$$
$$U \rightarrow V \tag{16}$$

Consider also the relation schemas

- R1={P,Q,S}
- R2={P,Q,S,U,V} and
- R3={P,Q,S,T}

As before, **negative points** for wrong, binary answers; explanations are optional, unless explicitly requested.

Q3.1 [2 pts] Write the projection of the FDs on R1.
Q3.2 [2 pts] True/False. The set of dependencies FD given above (11-16) is a minimal cover.
Q3.3 [4 pts] True/False. R1 is in 3NF.
Q3.4 [4 pts] True/False. R1 is in BCNF.
Q3.5 [2 pts]. Write the projection of the FDs on R2.
Q3.6 [4 pts] True/False. All the canditate keys of R2 are {PQU, QSU}.
Q3.7 [4 pts]. True/False. R2 is in BCNF.
Q3.8 [4 pts] True/False. Consider the decomposition of R2 {PQU, PQS, UV}. The new relations are in BCNF.
Q3.9 [2 pts]. Write the projection of the FDs on R3.
Q3.10 [2 pts] True/False. The candidate keys of R3 are {PQ,QS,PS}.
Q3.11 [4 pts] R3 is *not* in BCNF. Give all the dependencies of FD that violate the BCNF.
Q3.12 [2 pts] True/False. R3 is in 1NF.
Q3.13 [4 pts] True/False. Consider the decomposition of R3 to {PQS, QT}. The new relations are in BCNF.

End of homework questions