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. Illegible handwriting may get no points, at the discretion of the graders. Only drawings may be hand-drawn, as long as they are neat and legible.
- Please submit your answers in **hard copy, in class**, 1:30pm, on Tuesday, 03/26/2013 .
- For ease of grading, please solve each of the three questions on a **separate** page, i.e., three pages in total for this homework. If you need more pages for one problem, please staple them together. Type your **name and andrew ID on each** of the three pages. FYI, we will have three labeled piles at the front of the classroom, one for each problem.

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)`, which are simplified versions of the tables in the Gym database of the 1st homework.

`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.

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] 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 functional dependencies FD = {PQ→ S, PS → Q, PT → U, Q → T, QS → P, U → V }. Answer the following Yes/No or True/False questions. Give short answers whenever asked to do so.

Negative points for wrong answers. Explanations are optional and will be used for partial credit.

2.1.1 [2 pts] Yes/No. Is FD 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 only U → V.
2.1.5 [2 pts] True/False. The decomposition {PQS, PSTU, PTV} is not dependency-preserving, because it does not preserve both U → V and Q → 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 simpler functional dependencies FD' = {Q→ ST, P → T, PS → T, QU → V }. Answer the following True/False questions.

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 → S can be deduced from FD'.
2.2.6 [2 pts] Show that the dependency QU → 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→ S, PS → Q, PT → U, Q → T, QS → P, U → V }. **Notice that these are different functional dependencies** than the ones given in Q2. Consider also the relation schemas R1={P,Q,S}, R2={P,Q,S,U,V}, and R3={P,Q,S,T}.

Negative points for wrong answers. Explanations are optional and will be used for partial credit.

Q3.1 [2 pts] Write the projection of the FDs on R1. .
Q3.2 [2 pts] True/False. The set of dependencies given in Q3.1 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 keys are {PQU, QSU}.
Q3.7 [4 pts]. True/False. R2 is in BCNF.
Q3.8 [4 pts] True/False. The decomposition of R2 {PQU, PQS, UV} is in BCNF.
Q3.9 [2 pts]. Write the projection of the FDs on R3.
Q3.10 [2 pts] True/False. The keys of R3 are {PQ,QS,PS}.
Q3.11 [4 pts] R3 is *not* in BCNF. Give all the dependencies that violate the BCNF.
Q3.12 [2 pts] True/False. R3 is in 1NF.
Q3.13 [4 pts] True/False. The decomposition of R3 to {PQS, QT} is in BCNF.