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;
```

★ SOLUTION: $N_1/V(SSN, WORKS_AT) = 100,000/50,000 = 2$

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

★ SOLUTION: No, because it has multiple duplicates in the table.

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

```
SELECT *
FROM GYM
WHERE name = "Gym_planet";
```

★ SOLUTION: $N_2/V(name, GYM) = 40,000/30,000 = 1.33$

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 \textit{gymID} is not a candidate key for the table \textit{GYM}. Estimate the number of qualifying tuples of the query:

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

\[ \star \text{SOLUTION:} \quad \frac{N_1}{V(\text{SSN, WORKS_AT}) \times V(\text{gymID, WORKS_AT})} = \frac{100,000}{50,000 / 20,000} = 0.0001 \]

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

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

\[ \star \text{SOLUTION:} \quad \text{gymID is primary key in the table GYM and foreign key in the table WORKS_AT. So,} \]
\[ \frac{N_1 \times N_2}{V(\text{gymID, WORKS_AT})} = \frac{100,000 \times 40,000}{20,000} = 200,000 \]

\[ \star \text{SOLUTION:} \quad \text{gymID is not primary key in the table WORKS_AT. So,} \]
\[ \frac{N_1 \times N_1}{V(\text{gymID, WORKS_AT})} = \frac{100,000 \times 100,000}{20,000} = 500,000 \]
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 $FD = \{PQ \rightarrow S, PS \rightarrow Q, PT \rightarrow U, Q \rightarrow T, QS \rightarrow P, U \rightarrow V\}$

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?

★ SOLUTION: Yes.

2.1.1 [4 pts] Yes/No. Is the decomposition $\{PQ, QS, PQTU, UV\}$ lossless?

★ SOLUTION: No. Consider the instance of $R \{(p1,q,s1,t1,u1,v1), (p2,q,s2,t2,u2,v2)\}$. Because of the functional dependencies $QS \rightarrow P$ and $PQ \rightarrow S$, $p1 = p2$ if and only if $s1 = s2$. The join of $PQ$ and $QS$ has 4 tuples $\{(p1,q,s1), (p1,q,s2), (p2,q,s1), (p2,q,s2)\}$. So, the join of $PQ, QS, PQTU$ and $UV$ will contain at least 4 tuples - actually 8 - , and the decomposition is lossy.

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.

★ SOLUTION: The decomposition is not dependency-preserving. It does not preserve $PQ \rightarrow S$ and $PS \rightarrow Q$.

2.1.3 [5 pts] Yes/No. Is the decomposition $\{PQS, PSTU, PTV\}$ lossless?

★ SOLUTION: Yes. The join of $PQS, PSTU$ and $PTV$ can be constructed in two steps. First, we construct the join of $PQS$ and $PSTU$, which is lossless because their intersection is $PS$, which is a key for $PQSTU$. Next, we join the latter join with $PTV$. This is again lossless because the attribute intersection is $PT$ and $PT \rightarrow PTV$. 

4
2.1.4 [1 pts] True/False. The decomposition \{PQS, PSTU, PTV\} is not dependency-preserving, because it does not preserve \(U \rightarrow V\).

★ SOLUTION: True.

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 \(Q \rightarrow T\).

★ SOLUTION: True. The projection of the FDs onto PQS gives us: \(PQ \rightarrow S\), \(PS \rightarrow Q\) and \(QS \rightarrow P\). The projection of the FDs onto PSTU gives us: \(PT \rightarrow U\), while the projection onto PTV gives us: \(PT \rightarrow V\) (by transitivity on \(PT \rightarrow U\) and \(U \rightarrow V\)). The closure of this set of dependencies does not contain \(U \rightarrow V\) nor does it contain \(Q \rightarrow T\). So this decomposition is not dependency-preserving.

2.1.6 [2 pts] True/False. The decomposition \{PQS, PSTU, PTV\} is dependency-preserving.

★ SOLUTION: False.

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

\[
\begin{align*}
Q & \rightarrow ST \\
P & \rightarrow T \\
PS & \rightarrow T \\
QU & \rightarrow V
\end{align*}
\]

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 \(\{P,T\}\).

★ SOLUTION: True.

2.2.3 [1 pts] True/False. The attribute closure \(\{P,Q\}^+\) is \(\{P,T,Q,S\}\).

★ SOLUTION: True.

2.2.4 [1 pts] True/False. The attribute closure \(\{P,Q\}^+\) is \(\{P,S,T\}\).
★ SOLUTION: False.

2.2.4 [1 pts] True/False. The attribute closure \( \{P,Q\}^+ \) is \( \{P,T,Q,S,U,V\} \).

★ SOLUTION: False.

2.2.5 [1 pts] True/False. The dependency \( Q \rightarrow S \) can be deduced from FD’.

★ SOLUTION: True, it can be shown by using decomposition on \( Q \rightarrow ST \) (\( Q \rightarrow S, Q \rightarrow T \)).

2.2.6 [2 pts] True/False. The dependency \( QU \rightarrow TUV \) can be deduced from FD’.

★ SOLUTION: True. It can be shown by using decomposition on \( Q \rightarrow ST \) (\( Q \rightarrow S, Q \rightarrow T \)), augmentation (\( QU \rightarrow TU \)), and union with \( QU \rightarrow V \).

2.2.7 [2 pts] True/False. All the candidate keys of R are \( \{P,Q\} \).

★ SOLUTION: False. U should also be included in the candidate keys, since it does not appear in the right-hand side of any of the dependencies, and thus it is not possible to infer it from them.
Q3. **BCNF and 3NF, 40pts - SUBMIT ON SEPARATE PAGE**

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

$$PQ \rightarrow S$$  \hspace{1cm} (11)
$$PS \rightarrow Q$$  \hspace{1cm} (12)
$$PT \rightarrow U$$  \hspace{1cm} (13)
$$Q \rightarrow T$$  \hspace{1cm} (14)
$$QS \rightarrow P$$  \hspace{1cm} (15)
$$U \rightarrow V$$  \hspace{1cm} (16)

Consider also the relation schemas

- $R_1 = \{P,Q,S\}$
- $R_2 = \{P,Q,S,U,V\}$ and
- $R_3 = \{P,Q,S,T\}$

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 $R_1$.

★ **SOLUTION:** $\{PQ \rightarrow S, PS \rightarrow Q, QS \rightarrow P\}$

Q3.2 [2 pts] True/False. The set of dependencies $FD$ given above (11-16) is a minimal cover.

★ **SOLUTION:** True.

Q3.3 [4 pts] True/False. $R_1$ is in 3NF.

★ **SOLUTION:** True.

Q3.4 [4 pts] True/False. $R_1$ is in BCNF.

★ **SOLUTION:** True. $PQ, PS$ and $QS$ are the candidate keys for $R_1$.

Q3.5 [2 pts]. Write the projection of the FDs on $R_2$.

★ **SOLUTION:** $\{PQ \rightarrow S, PS \rightarrow Q, QS \rightarrow P, U \rightarrow V\}$

Q3.6 [4 pts] True/False. All the candidate keys of $R_2$ are $\{PQU, QSU\}$. 
**SOLUTION:** False. The keys are \{PQU, PSU, QSU\}.

Q3.7 [4 pts]. True/False. R2 is in BCNF.

**SOLUTION:** False. It is not even in 2NF as U is a proper subset of the subset keys and the FD U → V exists.

Q3.8 [4 pts] True/False. Consider the decomposition of R2 \{PQU, PQS, UV\}. The new relations are in BCNF.

**SOLUTION:** True.

Q3.9 [2 pts]. Write the projection of the FDs on R3.

**SOLUTION:** \{PQ → S, PS → Q, Q → T, QS → P\}

Q3.10 [2 pts] True/False. The candidate keys of R3 are \{PQ,QS,PS\}.

**SOLUTION:** True.

Q3.11 [4 pts] R3 is not in BCNF. Give all the dependencies of FD that violate the BCNF.

**SOLUTION:** Q is a proper subset of a key, but the dependency Q → T exists.

Q3.12 [2 pts] True/False. R3 is in 1NF.

**SOLUTION:** True.

Q3.13 [4 pts] True/False. Consider the decomposition of R3 to \{PQS, QT\}. The new relations are in BCNF.

**SOLUTION:** True.