CS315: Principles of Database Systems, IIT Kanpur				**Quiz II** (25 Oct 2024)
Name	Deebo			**20 marks**
Roll No	24007	Dept.	AWSM	Page **1** of **2**

Instructions:
1. This question paper contains 1 page (2 sides of paper). Please verify.
2. Write your name, roll number, department above in **block letters neatly with ink**.
3. Write your final answers neatly **with a blue/black pen**. Pencil marks may get smudged.
4. Don't overwrite/scratch answers, **no hardcoding** allowed – ambiguous cases may get 0 marks.

Q1. The tables p and q share a schema (both have two columns, named A, B), and have m, n rows respectively with $m < n$. Neither table has any duplicate rows. Find out the minimum, maximum sizes (in terms of m, n) of the tables resulting from the following operations? Justify your answers. Suboptimal answers will get no credit. **(4 + 4 = 8 marks)**

$p \bowtie q$ (natural join) Minimum size **0** Maximum size m

Give justification here

Since the tables have identical schema, both columns will participate in the natural join and hence, $p \bowtie q$ is essentially $p \cap q$. Thus, the join will be the largest when $p \subset q$ and smallest when the tables are disjoint.

$p \cup q$ (union) Minimum size n Maximum size $m + n$

Give justification here

The union will be the smallest when $p \subset q$ and the largest when the tables are disjoint.

Q2. Given a table `tbl(A INTEGER, B INTEGER)`, Deebo wants to write a conditional SQLite query (of the kind given on the right) to print **YES** if the column `A` is *cute* and **NO** otherwise. Deebo calls a column cute if that column contains no NULL values and moreover, that column has unique values in all rows (cute columns are candidate keys). Complete the query by giving the Boolean expression for the **YES** case. Give only the Boolean expression and not the entire query. *Hint: put parenthesis around statements if comparing their results.* **Note**: We will type your answers as SQLite queries to actual DBs to give marks. If your query takes excessively long to execute, it will get a default zero score. (**6 marks**)

```
SELECT CASE
   WHEN [Boolean expression]
      THEN 'YES'
   ELSE 'NO'
END;
```

((SELECT COUNT(*) FROM tbl) = (SELECT COUNT (DISTINCT A) FROM tbl))
AND
((SELECT COUNT(*) FROM tbl WHERE A IS NULL) = 0)

Q3. Fill exactly one box on the right and give brief justification. **(2 x (1+2) = 6 marks)**

For a schema R with $R', R'' \subseteq R$ being **subsets** of the columns, the dependency $R' \to R''$ is known to hold. For which columns $C \in R$ does $R'C \to R''$ hold true? ($R'C \equiv R' \cup \{C\}$ and *iff* means if-and-only-if). Justify using Armstrong's axioms.

- [x] Every $C \in R$
- [ ] No $C \in R$
- [ ] Iff $C \in R'$
- [ ] Iff $C \in R''$

Give justification here

Reflexivity gives us $R'C \to R'$ after which applying transitivity with $R' \to R''$ gives $R'C \to R''$.

The schema $R(A, B, C, D, E)$ has 3 dependencies $AB \to C$, $CD \to E$ and $DE \to B$. Is AB a candidate key for this schema? Is ABD a candidate key? Justify by applying the algorithm for finding attribute set closure.

- [ ] AB but not ABD
- [x] ABD but not AB
- [ ] Both AB, ABD
- [ ] Neither AB nor ABD

Give justification here

Starting with $(AB)^+ \supseteq AB$, as $AB \subseteq AB$, the dependency $AB \to C$ allows us to update the closure to $(AB)^+ \supseteq ABC$. However, no other FDs can be applied hereon since CD, DE are not subsets of ABC which forces the algorithm to terminate. This tells us that $(AB)^+ \neq R$.

Starting with $(ABD)^+ \supseteq ABD$, as $AB \subseteq ABD$, the dependency $AB \to C$ allows us to update the closure to $(ABD)^+ \supseteq ABCD$ and, as $CD \subseteq ABCD$, the dependency $CD \to E$ allows us to update the closure to $(ABD)^+ \supseteq ABCDE = R$. Since $(ABD)^+ \subseteq R$, we get $(ABD)^+ = R$.