© 2010 Sudarshan S. Chawathe

Name: _____

1. (1 pt.)
   - **Read the following, and all questions, very carefully.**
   - You may refer to the textbook, your notes, the assigned reading, and any other material you wish during this exam. Material cannot be shared.
   - COS 480 students must answer all questions that are not marked with a ⋆. Questions marked with a ⋆ may also be answered, for extra credit.
   - COS 580 students must answer all questions, including those marked with a ⋆. The points will be scaled by 3/4.

   Write your name in the space provided above.

2. (9 pts.) Consider $R(A, B, C, D, E)$ with dependencies

$$\begin{aligned} A &\rightarrow BD \\ BC &\rightarrow A \\ D &\rightarrow E \\ E &\rightarrow AB \end{aligned}$$

   List **all** keys of R. **Explain** clearly both (1) why the keys you list are valid and (2) why there are no other keys.

[additional space for answering the earlier question]

3. (5 pts.) Project the dependencies of Question 2 onto the relation $R'(A, B, C)$ using the algorithm described in class. Show enough of the working steps to clarify why the projected dependencies you list are the only ones generated by the algorithm.

4. (10 pts.) Decompose the schema of Question 2 as necessary to generate a BCNF schema. *For each decomposition, clearly indicate* the used dependency and the projected dependencies for the decomposed relations. List the relations in the final schema and explain why it is in BCNF.

[additional space for answering the earlier question]

5. (10 pts.) *Prove or disprove* the soundness of each of the following inference rules for functional dependencies *from first principles*.

(a)
$$\begin{array}{rcl} AB & \rightarrow & CD \\ AD & \rightarrow & BC \\ \hline A & \rightarrow & C \end{array}$$

(b)
$$\begin{array}{rcl} CD & \rightarrow & AB \\ BC & \rightarrow & AD \\ \hline C & \rightarrow & A \end{array}$$

6. (5 pts.) Consider a database with the following familiar schema, as discussed in class. (Ask for clarifications if the semantics are unclear.) The attributes `year` and `credits` have integer types while the rest have string types.

```
Students(sid, name, year, major)
Courses(cid, title, ta)
Enrolls(sid, cid, credits)
```

Provide a *standard SQL* (as described in class) query to find all students who are enrolled in all classes whose TA is 'Turing'. For each such student the query result must contain exactly one tuple, with that's student's id and name as attributes. Explain why your answer is correct.

7. (5 pts.) Provide an extended algebra query that is equivalent to the query of Question 6. Explain why your answer is correct.

8. (10 pts.) Provide a *standard SQL* query to find, for each major, the students of that major who have enrolled for the largest total number of credits. For each major, the query result must contain one or more tuples of the form (m, c, i, n), where m is a major, i and n are the ID and name of a student with major m who is enrolled for a total of c credits, and c is the maximum such value for major m. The result should be sorted lexicographically by descending m, ascending n, and ascending i.

9. (5 pts.) Provide an extended algebra query that is equivalent to the query of Question 8. Explain why your answer is correct.

10. (5 pts.) $\star$ How many superkeys does the relation R of Question 2 have? Explain your answer.

11. (10 pts.) $\star$ Given a relation $S \subset \mathbb{Z} \times \mathbb{Z}$ write a *basic relational algebra* expression that returns the tuples that have the third-largest value in as second attribute, or explain clearly why such a query cannot be written. Recall that $\mathbb{Z}$ denotes the set of integers. A basic relational algebra expression is one that uses no operators other than these five: $\sigma, \pi, \times, \cup, -$.