# Stanford University
# Computer Science Department

## Fall 2004 Comprehensive Exam in Databases

1. **Open Book & Notes / No Laptops.**
2. **The exam is timed for 60 minutes.**
3. **Write your Magic Number on this Honor Code statement sheet and the cover sheet of the exam.**

---

The following is a statement of the Stanford University Honor Code:

*A. The Honor Code is an undertaking of the students, individually and collectively:*

*1. that they will not give or receive aid in examinations; that they will not give or receive un-permitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;*

*2. that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.*

*B. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.*

*C. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.*

By writing my Magic Number below, I certify that I acknowledge and accept the Honor Code.

***Magic Number*----------**

# Stanford University Computer Science Department
# 2004 Comprehensive Exam in Databases

- The exam is *open book and notes*.
- There are 8 problems on the exam, with a varying number of points for each problem and subproblem for a total of 60 points (i.e., one point per minute). It is suggested that you look through the entire exam before getting started, in order to plan your strategy.
- Please write your solutions in the spaces provided on the exam. Make sure your solutions are neat and clearly marked.
- *Simplicity and clarity of solutions will count.* You may get as few as 0 points for a problem if your solution is far more complicated than necessary, or if we cannot understand your solution.

MAGIC NUMBER: _____

| Problem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| Max. points | 15 | 10 | 5 | 4 | 5 | 6 | 5 | 10 | 60 |
| Points | | | | | | | | | |

1. (15 points) The following facts must be converted to an E/R diagram:

   - All motors have a manufacturer, a model number that is unique for that manufacturer, and a number of horsepower.
   - Some motors are gasoline engines, and have a number of miles-per-gallon.
   - Some motors are electric motors, and have a voltage.
   - All automobiles have a manufacturer, a model name that is unique among all automobiles, and a price. All automobiles have one gasoline engine.
   - Some automobiles are hybrid and also have an electric motor. For all hybrid cars, there is a waiting time for delivery.

   (a) (10 pts.) Draw an E/R diagram that represents these facts. Make sure there is no redundancy in your design, and entity sets are not used where an attribute will do. Indicate keys for entity sets. Indicate when a relationship is many-one, and also whether it is "onto"; i.e., there is guaranteed to be an entity on the "one" side associated with each entity on the "many" side.

   (b) (5 pts.) Convert your E/R diagram from part (a) into a relational database schema. For motors, use the object-oriented style, where each motor entity is represented in exactly one relation. For automobiles, use the E/R style, where an automobile entity is represented in as many relations as it belongs to entity sets in the automobile hierarchy. Your schema can be informal, i.e., just the names of the relations and their attributes; it does not have to be SQL `CREATE TABLE` statements. However, indicate keys for your relations.

2. (10 points) Consider the following relational schema:

```
Student(ID, SSN, Name)   \\ ID is a key; SSN is a key
StudVoter(SSN, State)    \\ (SSN,State) is a key
```

You may assume the `StudVoter` relation only contains students—that is, any `SSN` in `StudVoter` also appears in `Student`.

Write a SQL query that lists the ID's (no duplicates please) of all students registered to vote in three or more states. *Your query will be graded on simplicity as well as correctness.*

3. (5 points) Start with the schema from the previous problem. Add to the schema one or more SQL99 tuple-based `CHECK` constraints that ensure no student is registered to vote in more than one state.

4. (4 points) Is there an easier way to modify the original schema to express the "no student is registered to vote in more than one state" requirement from the previous problem? If so, show it. If not, briefly explain why not.

5. (5 points) Continuing from the previous problem, assume the "no student is registered to vote in more than one state" requirement is being enforced, one way or another. Now suppose in addition you want to ensure that every student is indeed registered to vote. Can you express this requirement using SQL99 CHECK constraints? If so, add the CHECK constraints to the schema. If not, briefly explain why not, and suggest other mechanisms that could be used to enforce the desired requirement.

6. (6 points) Continuing with the same schema, consider the following two transactions:

```
T1: select count(*) from StudVoter;
    select count(*) from StudVoter;

T2: insert into StudVoter values ('123-45-6789', 'California');
```

(a) Do the possible behaviors of these two transactions differ if we choose isolation level `repeatable-read` versus isolation level `serializable`? Answer "Yes" or "No" with a brief explanation.

(b) Now answer the same question, except replace transaction `T2` with:

```
T2: delete from StudVoter
    where SSN='123-45-6789' and state='California'
```

Do the possible behaviors of these two transactions differ if we use isolation level `repeatable-read` versus isolation level `serializable`? Answer "Yes" or "No" with a brief explanation.

7. (5 points) Continuing with the same schema, using standard SQL capabilities is it possible to authorize a database user to query only those students registered in California or Nevada? If so, briefly explain how. If not, briefly explain why not.

8. (10 points) Consider a relation $R(A, B, C, D)$ with the functional dependencies $ABC \rightarrow D$, $D \rightarrow C$, and $C \rightarrow B$.

   (a) (3 pts.) Find all the keys for $R$.

   (b) (1 pt.) Which of the given functional dependencies violate Boyce-Codd Normal Form?

   (c) (1 pt.) Which, if any, of the given functional dependencies violate Third Normal Form. Explain why, if so.

   (d) (2 pts.) How many superkeys does $R$ have?

   (e) (3 pts.) Explain why the multivalued dependency $D \twoheadrightarrow A$ holds in $R$.