UCLA
Computer Science Department
Winter 2013

Instructor: J. Cho

Student Name and ID: ______________________________

# CS143 Final: Closed Book, 110 minutes

**(** IMPORTANT PLEASE READ **):**

- There are 4 problems on the exam for a total of 64 points to be completed in 110 minutes. *You should look through the entire exam before getting started, in order to plan your strategy.*
- *Simplicity and clarity of your solutions will count.* You may get as few as 0 point for a problem if your solution is far more complicated than necessary, or if we cannot understand your solution.
- If you need to make any assumption to solve a question, please write down your assumptions.
- To get partial credits, you may want to write down how you arrived at your answer step by step.
- You may use two double-sided cheat-sheets during exam. You are also allowed to use a calculator.
- Please, write your answers neatly. Attach extra pages as needed. Write your name and ID on the extra pages.

| **Problem** | **Score** | |
|---|---|---|
| 1 | 20 | |
| 2 | 12 | |
| 3 | 14 | |
| 4 | 18 | |
| Total | 64 | |

## Problem 1 (Index): 20 points

1. Consider a B+tree of order (or branching factor) $n$.

   (a) What is the minimum number of records that the tree can index, when it has 2 levels? (3 points)

   **ANSWER:**
   $2 \times \lceil (n-1)/2 \rceil$

   (b) What is the minimum number of records that the tree can index, when it has $k$ levels ($k \geq 2$)? (3 points)

   **ANSWER:**
   $2 \times \lceil n/2 \rceil^{k-2} \times \lceil (n-1)/2 \rceil$

   (c) What is the maximum number of records that the tree can index, when it has 2 levels? (3 points)

   **ANSWER:**
   $n \times (n-1)$

   (d) What is the maximum number of records that the tree can index, when it has $k$ levels ($k \geq 2$)? (3 points)

   **ANSWER:**

$n^{k-1} \times (n-1)$

(e) Suppose that we are using a B+tree with $n = 99$ to index a table with 300,000 records. In the worst case, how many nodes in the index will one have to examine while looking up a record? (3 points)

**ANSWER:**
$2 \times \lceil n/2 \rceil^{k-2} \times \lceil (n-1)/2 \rceil \leq 300,000$ `where` $n = 99$. `So` $2 \times 50^{k-2} \times 49 \leq 300,000$. `The maximum such` $k$ `is 4.`

2. Consider the following movie-ratings database with two tables:

```
Movie(title, director)        // every movie title is unique
Rating(person, title, score) // every person name is unique
```

Suppose there are three types of queries commonly asked on this schema:

- Given a movie title, find the director of the movie.
- Match each person with the directors of movies the person has rated. (Given a person, list all directors of movies he has rated)
- Given a person, find the titles of all movies the person has rated.

(a) What is the *minimum* number of indexes needed to speed up *all three* types of queries? (2 points)

**ANSWER:**
`2`

(b) On which attributes should these indexes be created? (3 points)

**ANSWER:**
`title, person`

## Problem 2: 12 points

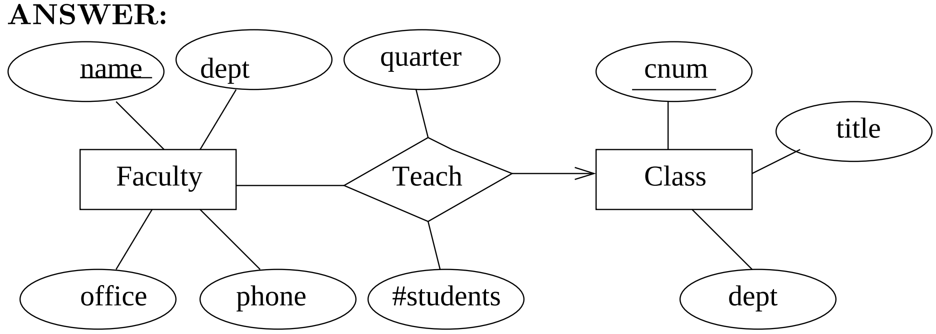
Consider the following relational schema describing faculty, classes, and the classes taught:

Faculty(name, dept, office, phone)
Class(cnum, title, dept)
Teach(name, cnum, quarter, #students)

In the above schema, Teach.name refers to the name of the faculty who teaches the class. Underlined attributes are the keys of the relations.

1. Draw an ER diagram from which this relational schema could have been produced. From the information you have, please make sure to indicate any relevant constraints in the diagram, including keys and cardinalities. (6 points)

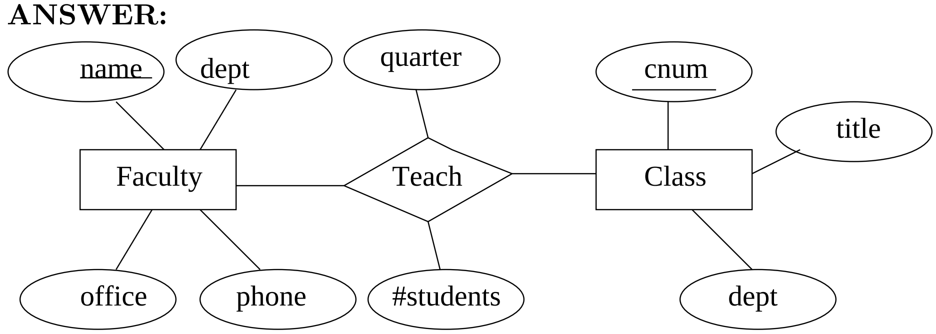
**ANSWER:**




`Every faculty teaches at most one class when name is the key of Teach.`

2. Now suppose for relation Teach we instead have: Teach(name, cnum, quarter, #students). If your ER diagram changes as a result, draw the the changed ER diagram. (3 points)

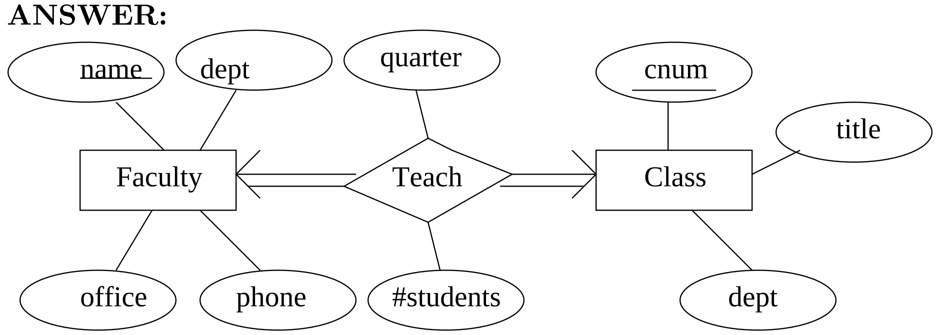
**ANSWER:**




`A faculty may teach multiple classes and a class can be taught by multiple faculty if (name, cnum) together is a key.`

3. Now suppose that we do not have any explicit information about the keys of the Teach relation, but we know that there is a SQL referential integrity constraint from Faculty.name to Teach.name, and another SQL referential integrity constraint from Class.cnum to Teach.cnum. If your ER diagram changes as a result, draw the changed ER diagram. (3 points)

**ANSWER:**




```
All faculty has to teach a class from the first constraint.  From the second
constraint, every class has to be taught by faculty.  Also, the two SQL
referential constraints imply that each of Teach.name and Teach.cnum is
a key.  When Teach.name is a key, a faculty can teach at most one class.
Similarly when Class.cnum is a key, a class can be taught by at most one
faculty.
```

## Problem 3: 14 points

All relations in Problem 3 is based on the set semantic and no null values are allowed. We use the notation $|S|$ to denote the number of tuples in the relation $S$.

1. Given a relation R(A, B, C, D, E) and functional dependencies (FD's) AB → CE, B → D, and D → E, which of the following functional dependency cannot be inferred? (3 points)

   (a) AD → CE

   (b) BC → D

   (c) AB → A

   (d) B → E

   (e) None. All above FD's can be inferred.

   **ANSWER:**
   `(a)`

2. Suppose a relation R(A, B, C, D, E) and the functional dependencies A → DE, D → B, and E → C. Now project R (and therefore its functional dependencies) onto R'(A, B, C). Which of the following statements is true about the key(s) for R'? (3 points)

   (a) ABC is the only key for R'

   (b) A is the only key for R'

   (c) DE is the only key for R'

   (d) Each of A, B, and C is individually a key for R'

   (e) None of the above is true

   **ANSWER:**
   `(b). A → B and A → C are logically implied by the original set of FD's.`
   `Therefore, A is the key.`

3. Consider the relation $R(A, B, C)$, on which the multivalued dependency $A \twoheadrightarrow B$ holds. If $|R| = 12$, $|\pi_A(R)| = 1$ and $|\pi_C(R)| = 4$, what is/are the possible value(s) of $|\pi_B(R)|$? (4 points)

   **ANSWER:**
   `3`

4. Consider the relation $R(A, B, C, D)$, on which the functional dependencies $B \to C$ and $C \to D$ hold. If $|R| = 20$, $|\pi_B(R)| = 10$, and $|\pi_D(R)| = 10$, what is/are the possible value(s) of $|\pi_C(R)|$? (4 points)

   **ANSWER:**
   `10`

## Problem 4: 18 points

1. Given a table Employee(<u>name</u>,salary), suppose there are two tuples initially: ('A',30) and ('B',40). Consider the following two concurrent transactions, each of which runs once and commits. You may assume there are no other transactions in the system and individual statements will execute atomically.

```
T1: BEGIN TRANSACTION
    S1: INSERT INTO Employee VALUES ('C',50)
    S2: UPDATE Employee SET salary = salary+30 WHERE name='A'
    COMMIT

T2: BEGIN TRANSACTION
    S3: SELECT MIN(salary) AS p1 FROM Employee
    S4: SELECT MAX(salary) AS p2 FROM Employee
    COMMIT
```

Suppose that the transaction T1 is executed with the isolation level **READ COMMITTED**.

(a) If transaction T2 executes with the isolation level **SERIALIZABLE**, what is/are the possible pair(s) of values p1 and p2 returned by T2? Please carefully indicate all possible pair(s). For example, you may write (p1,p2)=(10,20) if you believe T2 returns 10 as the value of p1 and 20 as the value of p2. (3 points)

**ANSWER:**
`(p1,p2) = (30,40), (40,60)`

(b) If transaction T2 also executes with the isolation level **READ COMMITTED**, what is/are the possible pair(s) of values p1 and p2 returned by T2? Please carefully indicate all possible pair(s). (3 points)

**ANSWER:**
`(p1,p2) = (30,40), (40,60), (30,60)`

(c) If transaction T2 executes with isolation level read uncommitted, what is/are the possible pair(s) of values p1 and p2 returned by T2? Please carefully indicate all possible pair(s). (3 points)

**ANSWER:**
`(p1,p2) = (30,40), (40,60), (30,60), (30,50)`

2. Now consider a simple relation Employee(<u>name</u>, salary), and the following transaction T:

```
T: (Q1) SELECT salary FROM Employee WHERE name = 'John'
        <other SELECT statements that only READ data from the database>
   (Q2) SELECT salary FROM Employee WHERE name = 'John'
        COMMIT
```

(a) Suppose all other transactions in the system are declared as SERIALIZABLE and READ ONLY, indicating that they do not modify any tuples. What is the weakest isolation level needed for transaction T to ensure that queries Q1 and Q2 will always get the same result? Circle one: (3 points)

   i. READ UNCOMMITTED
   ii. READ COMMITTED
   iii. REPEATABLE READ
   iv. SERIALIZABLE

**ANSWER:**
`(i) If no transaction modifies the table, there is nothing to worry about`

(b) Suppose all other transactions in the system are declared as SERIALIZABLE, and they involve only SELECT statements and some UPDATEs on salaries (i.e., no insertions or deletions). What is the weakest isolation level needed for transaction T to ensure that queries Q1 and Q2 will always get the same result? Circle one: (3 points)

   i. READ UNCOMMITTED
   ii. READ COMMITTED
   iii. REPEATABLE READ
   iv. SERIALIZABLE

**ANSWER:**
`(iii) Since there is no insertion, we do not have to worry about phantoms, but we do need to worry about non-repeatable reads`

(c) Suppose all other transactions in the system are declared as SERIALIZABLE, and we know nothing else about them. What is the weakest isolation level needed for transaction T to ensure that queries Q1 and Q2 will always get the same result? Circle one: (3 points)

   i. READ UNCOMMITTED
   ii. READ COMMITTED
   iii. REPEATABLE READ
   iv. SERIALIZABLE

**ANSWER:**

`(iii) because we do not perform any range queries in T.`