**CS145 Final Examination**

Wednesday, December 12, 2001, 12:15 – 3:15PM

## Directions

The exam is *open book/notes*; any written materials may be used.

For each of the 35 questions, circle the letter (a), (b), (c), or (d) of your chosen answer. Do not circle more than one answer. If you wish to change your answer, please indicate clearly what your "final answer" is.

Score = 3 times number-right minus number-wrong, so random guessing nets you nothing on the average, and 105 is a perfect score.

If you wish to explain or demonstrate your solution to a problem for partial credit, you may use page bottoms or the backs of the pages (but warn us on the front). Please use this option sparingly, e.g., if you think the question is flawed or open to multiple interpretations, because we shall only be awarding partial credit in rare situations.

You have about 5 minutes per question. Use your time wisely, and do not spend too much time on any one question.

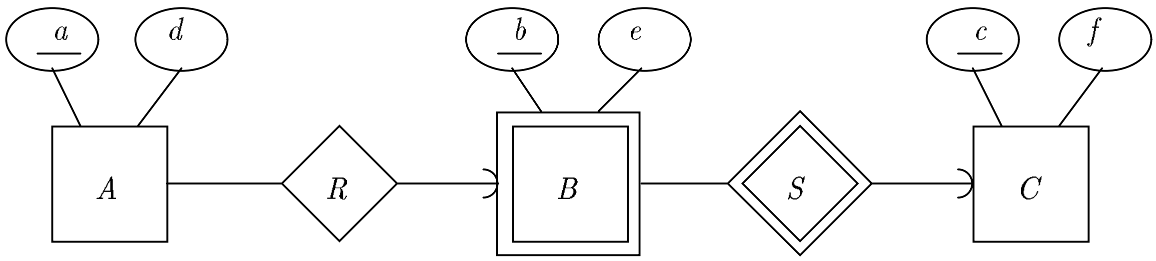
Do not forget to **sign the pledge** below.

I acknowledge and accept the honor code.

______________________________

Print your name here: ________________________________________

The following three questions are based on the E/R diagram below.




**Question 1:** If entity set $A$ currently has 100 entities, which of the following could be the number of $B$ entities?

| | |
|---|---|
| I. | 1 |
| II. | 100 |
| III. | 200 |

(a) I or II only (b) II or III only (c) II only (d) I, II, or III

**Question 2:** If we convert the E/R diagram to relations in the standard way described in the text, which set of attributes would *not* appear in the schema of some relation?
(a) $(b, c, e)$ (b) $(a, b)$ (c) $(a, d)$ (d) $(c, f)$

**Question 3:** Suppose the following are the current sets of entities for the three entity sets: $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, and $C = \{c_1, c_2\}$. Which of the following are possible relationship sets for $R$?

I. $\{(a_1, b_1),\ (a_2, b_1)\}$
II. $\{(a_1, b_1, c_1),\ (a_2, b_2, c_2)\}$
III. $\{(a_1, b_1)\}$

(a) I only (b) II only (c) I or II only (d) neither I, nor II, nor III

---

**Question 4:** Suppose $R(a, b)$ contains the tuples $\{(1, 2),\ (3, 4)\}$ and $S(b, c)$ contains the tuples $\{(2, 5),\ (2, 6),\ (7, 8)\}$. The natural outerjoin of $R$ and $S$ contains how many tuples?
(a) 2 (b) 3 (c) 4 (d) 5

**Question 5:** Initially, user $A$ is the owner of relation $R$, and no other user holds privileges on $R$. The following are executed:

by $A$: `GRANT UPDATE ON R TO B`
by $A$: `GRANT UPDATE(a) ON R TO C WITH GRANT OPTION`
by $C$: `GRANT UPDATE(a) ON R TO B WITH GRANT OPTION`
by $A$: `REVOKE UPDATE(a) ON R FROM C CASCADE`

Which of the following best describes the status of $B$'s privileges on $R$?
(a) $B$ can update any attribute of $R$ except $a$ but cannot grant that privilege.
(b) $B$ has no privileges on $R$ and cannot grant any.
(c) $B$ can update any attribute of $R$ except $a$, but can grant others the privilege to update $R.a$ .
(d) $B$ can perform any update on $R$ but cannot grant that privilege.

---

**Question 6:** Initially, $R(a, b)$ contains the tuples $\{(1, 2),\ (3, 4)\}$ only, and two transactions run at about the same time. Transaction $T_1$ consists of the steps:

$1a$) `SELECT * FROM R;`
$1b$) `SELECT * FROM R;`

and transaction $T_2$ consists of the steps:

$2a$) `DELETE FROM R WHERE a = 1;`
$2b$) `INSERT INTO R VALUES(5,6);`

$T_1$ runs with isolation level `REPEATABLE READ`, and $T_2$ runs with isolation level `SERIALIZ-ABLE`. Each transaction commits after performing both its steps. The possible numbers of tuples that the second selection of $T_1$ (step $1b$) can retrieve are:
(a) 1 or 2 only (b) 2 or 3 only (c) 2 only (d) 1, 2, or 3 only

---

**Question 7:** Suppose $R$ is declared:

```
CREATE TABLE R (a INT REFERENCES S(b) ON UPDATE CASCADE);
```

Of the following modifications to $S$:

```
I.  UPDATE S SET b = 10;
II. DELETE FROM S;
```

which will guarantee that the tuple (2) will no longer be in $R$, assuming it was there before the modification?
(a) both I and II (b) I only (c) II only (d) neither I nor II

**Question 8:** Let relation `Employees` and assertion $A$ be declared by the following:

```
CREATE TABLE Employees (
    name CHAR(50) PRIMARY KEY,
    dept CHAR(20),
    salary INT
);

CREATE ASSERTION A CHECK ( 'Toy' IN (
    SELECT dept
    FROM Employees
    GROUP BY dept
    HAVING AVG(salary) >= 50000
));
```

Which of the following best describes the constraint enforced by this assertion?

(a) Every employee making at least \$50,000 must be in the Toy Department.

(b) Only the Toy Department may have an average salary of \$50,000 or more.

(c) The average salary of employees in the Toy Department is at least \$50,000.

(d) The average salary in each department other than the Toy Department is less than \$50,000.

---

**Question 9:** Relation $R(a, b, c)$ currently has the following instance:

$$\{(1,2,3),\ (3,4,2),\ (2,6,1)\}$$

We make the following view definitions:

```
CREATE VIEW V AS
    SELECT a*b AS d, c FROM R;

CREATE VIEW W AS
    SELECT d, SUM(c) AS e FROM V GROUP BY d;
```

What is the sum of all the components of all the tuples of the following query?

```
SELECT AVG(d), e FROM W GROUP BY e;
```

(a) 10 (b) 17 (c) 23 (d) 28

**Question 10:** Given a relation `Emps(name, salary)` we want to find the names of the employees whose salary is greatest; i.e., no other employee makes a strictly larger salary. Recall that core relational algebra includes only union, intersection, difference, select, project, rename, product, natural join and theta-join. Extended relational algebra includes these operators, plus extended projection, $\delta$, $\gamma$, $\tau$, and outerjoin. Which of the following best describes the languages in which we can express the query above?

(a) It can be expressed in SQL and extended relational algebra, but not in the core relational algebra or Datalog.

(b) It can be expressed in extended relational algebra and Datalog, but not in SQL or core relational algebra.

(c) It can be expressed in SQL, Datalog, and core relational algebra.

(d) It can be expressed in SQL, but not in any form of relational algebra or Datalog.

---

**Question 11:** The relation $R(A, B, C, D)$ has dependencies $A \twoheadrightarrow B$, $C \rightarrow A$, and $D \rightarrow B$. Which of the following dependencies *does not* hold in $R$?

(a) $A \rightarrow B$ (b) $C \rightarrow B$ (c) $C \twoheadrightarrow D$ (d) None of the above; i.e., they all hold in $R$.

---

**Question 12:** Relations $R(a)$ and $S(b)$ contain, respectively, only the tuple (1) and only the tuple (2). Then, we add the following constraint to $R$:

```
CHECK(a IN (SELECT 2*b FROM S))
```

and the following constraint to $S$:

```
CHECK(b IN SELECT a-1 FROM R))
```

(Ignore existing violations.) How many `INSERT`, `DELETE`, and `UPDATE` statements, in any combination, does it take to get the tuple (10) into $R$, without deferring the checks?

(a) 1 (b) 5 (c) 9 (d) We cannot get (10) into $R$ as long as these constraints are satisfied.

---

**Question 13:** Suppose $R$ is declared `CREATE TABLE R(a INT, b INT)` and currently contains only the tuple $(1, 2)$. We have just executed the JDBC statement

```
PreparedStatement myStat = myCon.createStatement(
    "UPDATE R SET b = ? WHERE a = 1");
```

A suitable next step, if we wish to replace the tuple $(1, 2)$ in $R$ by $(1, 5)$ would be:

(a) `myStat.setInt(1,5);`

(b) `myStat.setInt(2,5);`

(c) `myStat.executeUpdate(5);`

(d) `myStat.executeQuery();`

The next two questions are based on the following information. Suppose we declare the following user-defined types (UDT's) and relations:

```
CREATE TYPE PairType AS (
    a INT,
    b INT,
);

CREATE TYPE ProductType AS (
    arguments PairType,
    result INT
);

CREATE TABLE MultTable OF ProductType;
```

**Question 14:** Which of the following methods is *least* useful for inserting the fact that $2 \times 3 = 6$, i.e., the object `ProductType(PairType(2,3),6)` into `MultTable`?

(a) The observer method for `arguments`
(b) The mutator method for $a$
(c) The generator method for `PairType`
(d) The generator method for `ProductType`

**Question 15:** In order to use `MultTable` to find out what $2 \times 3$ equals, we can write the following query with the occurences of two important operators replaced by `?1` and `?2`:

```
SELECT mm ?1 result()
FROM MultTable mm
WHERE mm ?1 arguments() ?2 a() = 2 AND
    mm ?1 arguments() ?2 b() = 3;
```

The proper operators for `?1` and `?2` are:

(a) `?1 = ->` and `?2 = ->`
(b) `?1 = ->` and `?2 = .` (dot)
(c) `?1 = .` and `?2 = ->`
(d) `?1 = .` and `?2 = .`

---

**Question 16:** In the SQL 3-valued logic, the value of expression

```
R.a > R.b OR R.a <= 0 OR R.b >= 0
```

can be:

(a) Only `TRUE` or `FALSE`
(b) Only `FALSE` or `UNKNOWN`
(c) Only `TRUE` or `UNKNOWN`
(d) Any of `TRUE`, `FALSE`, or `UNKNOWN`.

**Question 17:** The relation $R(a, b)$ currently has no constraints whatsoever, and contains the tuples $(1, 2)$ and $(3, 4)$. We now create a trigger:

```
CREATE TRIGGER Foo
AFTER UPDATE OF R
REFERENCING OLD ROW AS OldTuple, NEW ROW AS NewTuple
FOR EACH ROW
WHEN (NewTuple.b <= 10)
INSERT INTO R VALUES(NewTuple.a, 1 + NewTuple.b);
```

Next, we execute the statement:

```
UPDATE R SET b = 3 WHERE a = 1;
```

What is the new value of $R$ as a result of the update and all executions of the trigger?

(a) $\{(1, 2),\ (3, 4)\}$
(b) $\{(1, 3),\ (1, 4),\ (3, 4)\}$
(c) $\{(1, 3),\ (1, 4),\ (1, 5),\ (1, 6),\ (1, 7),\ (1, 8),\ (1, 9),\ (1, 10)\}$
(d) $\{(1, 3),\ (1, 4),\ (1, 5),\ (1, 6),\ (1, 7),\ (1, 8),\ (1, 9),\ (1, 10),\ (1, 11)\}$

---

The next two questions are based on the relation $R(A, B, C, D, E)$ with functional dependencies: $ABC \to DE$ and $E \to BCD$.

**Question 18:** The number of keys of $R$ is:

(a) 1 (b) 2 (c) 3 (d) 10

**Question 19:** Which of the following is true of $R$?

(a) $R$ is not in 3NF. (b) $R$ is in 3NF but not in BCNF. (c) $R$ is in BCNF but not in 4NF. (d) $R$ is in 4NF.

---

**Question 20:** A set of attributes $X$ is said to be *closed* if $X^+ = X$. In the relation $R(A, B, C, D)$ with functional dependencies $AB \to C$ and $BC \to D$, how many closed sets of attributes are there?

(a) 5 (b) 7 (c) 10 (d) 12

---

**Question 21:** Which of the following is a safe Datalog rule?

(a) `Answer(x,y) ← P(x,z) AND NOT Q(y)`
(b) `Answer(x,y) ← P(x,z) AND x<y AND NOT P(y,z)`
(c) `Answer(x,y) ← P(x,z) AND P(y,z) AND x >= z`
(d) None of the above (i.e., none is safe).

**Question 22:** Suppose we have the following database schema: $P(A,B)$ and $Q(C,D)$. Then the Datalog rule

```
Answer(x,y) <- P(x,x) AND Q(x,y)
```

produces the same relation instance as which of the following expressions of relational algebra? Do not be concerned about the name of the resulting relation or its attributes.

(a) $\pi_{B,D}\big(\sigma_{P.A=Q.C}(P \times Q)\big)$

(b) $\pi_{C,D}\Big(\sigma_{B=D}\big(\rho_{P(B,C)}(P) \bowtie Q\big)\Big)$

(c) $\pi_{P.A,Q.D}\big(\sigma_{A=B}(P) \times Q\big)$

(d) $\Big(\rho_{R(C)}\big(\pi_A(\sigma_{A=B}(P))\big)\Big) \bowtie Q$

---

**Question 23:** Suppose we have the following database schema: $P(A,B,C)$ and $Q(C,D)$. Then the relational algebra expression $\pi_{P.A,Q.D}\big(\sigma_{P.A<Q.C}(P \underset{P.B=Q.C}{\bowtie} Q)\big)$ produces the same value as which of the following Datalog programs?

(a) `Answer(x,w)` $\leftarrow$ `P(x,y,z) AND Q(z,w) AND x < z`
(b) `Answer(x,w)` $\leftarrow$ `P(x,y,z) AND Q(z,z) AND x < z`
(c) `Answer(x,w)` $\leftarrow$ `P(x,y,z) AND Q(y,w) AND x < y`
(d) `Answer(x,w)` $\leftarrow$ `P(x,y,z) AND Q(z,z) AND x < y`

---

The next two questions are based on the following Datalog program and EDB:

```
Reach(x) ← Source(x)
Reach(x) ← Reach(y) AND Arc(y,x)

NoReach(x) ← Target(x) AND NOT Reach(x)
```

The EDB is: $Source = \{a\}$; $Target = \{c,d\}$; $Arc = \{(a,c),\ (b,b)\}$.

**Question 24:** The stratified model for this Datalog program and EDB is:
(a) $Reach = \{a,b,c\}$; $NoReach = \{d\}$
(b) $Reach = \{a,c\}$; $NoReach = \{d\}$
(c) $Reach = \{a,c,d\}$; $NoReach = \emptyset$
(d) $Reach = \{b,c\}$; $NoReach = \emptyset$

**Question 25:** The number of minimal models other than the stratified model is:
(a) 0 (b) 1 (c) 2 (d) 3

**Question 26:** Which of the following statements is *not* true about Persistent, Stored Modules (PSM) in SQL?

(a) In a client-server environment, PSM allows multiple SQL statements to be executed on the server without communicating with the client in between these statements.

(b) PSM can include branching logic not otherwise supported in a single SQL query or update statement.

(c) The `RETURN` statement causes the PSM to terminate and return the value supplied.

(d) None of the above. (All of these statements are true.)

---

**Question 27:** Consider the following table definition and SQL query.

```
CREATE TABLE R (
    a INT PRIMARY KEY,
    b INT,
    c INT,
    d INT,
    e INT,
);

SELECT a, MIN(b), SUM(c)
FROM R
WHERE b > 5;
GROUP BY a
HAVING condition;
```

Which of the following statements is *not* true?

(a) The *condition* can be `d = 5`.

(b) The *condition* can be `a = SUM(e)`.

(c) The value of `MIN(b)` must be 6 or more.

(d) None of the above. (That is, all of the above statements are true.)

---

In each of the following 8 questions, you are asked to compare two queries $Q_1$ and $Q_2$. You must tell whether the queries are:

1. The same [choice (a)], meaning that for every database the answers to the two queries are the same. That is, the same tuples are produced by each query, and a tuple is produced the same number of times by each query. The order in which tuples are produced is not to be considered.

2. Completely different [choice (d)], meaning that there are databases where $Q_1$ produces more of some particular tuple, and other databases where $Q_2$ produces more of some particular tuple. Note that the query producing the smaller number of copies of a tuple may produce zero copies of that tuple.

3. One is contained in the other but they are not the same [choice (b) or (c)]. For instance, $Q_1$ is contained in $Q_2$ if on every database, $Q_2$ produces at least as many copies of each tuple as $Q_1$ does. Note that it is possible $Q_2$ produces one or more copies of a tuple, while $Q_1$ produces none of that tuple.

General advice:

- Do not assume a query has a trivial syntactic error and therefore produces nothing.
- Relations mentioned in the queries may have attributes not mentioned, but their existence should not affect the answer.
- Relations may have `NULL`'s.
- SQL queries should be assumed to be in standard SQL unless stated otherwise.
- In SQL, it is possible that there may be duplicate tuples, but in relational algebra or Datalog assume the relations are sets unless we state otherwise.
- Although we did not cover it explicitly in class, SQL duplicate elimination and grouping treat `NULL` like any other value.

---

**Question 28:** In the following, assume $R$ and $S$ have the same schema.

$Q_1$: `(SELECT * FROM R) INTERSECT ALL (SELECT * FROM S);`
$Q_2$: `(SELECT * FROM R) NATURAL JOIN (SELECT * FROM S);`

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

---

**Question 29:** Here, the schema is $R(a, b)$.

$Q_1$: $\gamma_{a,MAX(b)\rightarrow c,MIN(b)\rightarrow d}(R)$
$Q_2$: $\left(\gamma_{a,MAX(b)\rightarrow c}(R)\right) \bowtie \left(\gamma_{a,MIN(b)\rightarrow d}(R)\right)$

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

**Question 30:**

$Q_1$:
```
Answer(x,y) ← Arc(x,y)
Answer(x,y) ← Arc(x,z) AND Answer(z,y)
```

$Q_2$:
```
Answer(x,y) ← Arc(x,y)
Answer(x,y) ← Answer(x,z) AND Answer(z,w) AND Answer(w,y)
```

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

---

The following two questions are based on this ODL schema:

```
class A (extent Aext key aName) {
    attribute string aName;
    relationship Set<B> myBs inverse B::myA;
}

class B (extent Bext key bName) {
    attribute string bName;
    relationship A myA inverse A::myBs;
    relationship C myC inverse C::myBs;
}

class C (extent Cext key cName) {
    attribute string cName;
    relationship Set<B> myBs inverse B::myC;
}
```

**Question 31:**

$Q_1$:
```
SELECT aa.aName
FROM Aext aa
WHERE FOR ALL bb IN aa.myBs:
    bb.myC.cName = "Joe"
```

$Q_2$:
```
SELECT bb.myA.aName
FROM Bext bb
WHERE bb.myC.cName = "Joe"
```

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

**Question 32:**

$Q_1$:
```
SELECT bb.bName
FROM Cext cc, cc.myBs bb
WHERE bb.myA.aName = "Joe"
```

$Q_2$:
```
SELECT bb.bName
FROM Aext aa, aa.myBs bb
WHERE aa.aName = "Joe"
```

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

---

**Question 33:** Assume the relation schemas are $R(a, b)$ and $S(a, b)$.

$Q_1$: $\pi_a(R - S)$
$Q_2$: $\pi_a(R) - \pi_a(S)$

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

---

**Question 34:**

$Q_1$:
```
SELECT R.a
FROM R
WHERE EXISTS (SELECT S.c FROM S WHERE S.d = R.b);
```

$Q_2$:
```
SELECT R.a
FROM R
WHERE R.b = ANY (SELECT S.d FROM S);
```

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.

**Question 35:**

$Q_1$:
```
SELECT a
FROM R
WHERE a > ALL (SELECT b FROM R);
```

$Q_2$:
```
SELECT a
FROM R
WHERE a <> ANY (SELECT b FROM R);
```

(a) $Q_1$ and $Q_2$ produce the same answer.
(b) The answer to $Q_1$ is always contained in the answer to $Q_2$.
(c) The answer to $Q_2$ is always contained in the answer to $Q_1$.
(d) $Q_1$ and $Q_2$ produce different answers.