CS145 Final Exam – Dec. 6, 2004, 3:30–6:30PM

- This exam is open book and notes. You may use a laptop, but please keep your sound on mute so you do not disturb others. You may access class notes or other written materials on your computer, even using the Internet if you have wireless access. However, you must **not** use your computer to run a DBMS, either over the Internet, or locally, should you have a DBMS that runs on your own computer.
- This exam consists of 35 multiple-choice questions, so you have 5 minutes per question. Questions count 3 points each, with 1 point deducted for wrong answers (nothing deducted if you choose not to answer a question). The maximum score is 105.
- Please circle your choice on each question: (a), (b), (c), or (d).

Print your name:_____

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 unpermitted 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.

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.

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.

I acknowledge and accept the Honor Code.

Signed:_____

****

In each of the first 10 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.
- SQL relations may have `NULL`'s, although in relational algebra, you should assume no `NULL`'s unless stated otherwise.
- SQL queries should be assumed to be in standard SQL-99.
- In SQL it is possible that there may be duplicate tuples, but in relational algebra assume the relations are sets unless stated otherwise.

****

1. In the following relational algebra expressions, assume union and difference are **bag** operators, and the relations $R(x)$, $S(x)$, and $T(x)$ are bags.

   Q_1: $(R \cup S) - T$
   Q_2: $(R - T) \cup (S - T)$

   (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.

2. The following SQL queries refer to relations $R(a, b)$ and $S(b, c)$.

   Q_1:
   ```
   SELECT * FROM R NATURAL JOIN S;
   ```
   Q_2:
   ```
   SELECT * FROM R LEFT NATURAL OUTER JOIN 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.

3. In the following queries, the schema of relation R can be arbitrary.

```
Q1:  (SELECT * FROM R)
        UNION
     (SELECT * FROM R);
Q2:  SELECT * 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.

4. The following SQL queries refer to a relation $R(a, b)$.

```
Q1:  SELECT a FROM R r1
     WHERE EXISTS(SELECT * FROM R WHERE a = r1.b);
Q2:  SELECT a FROM R
     WHERE b = ANY(SELECT a 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.

5. The following SQL queries refer to a relation $R(a, b, c)$.

```
Q1:  SELECT DISTINCT a, b FROM R;
Q2:  SELECT a, b FROM R GROUP BY a, b;
```

   (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.

6. In the following queries, the schema of R is arbitrary, although it must include a.

```
Q1:  SELECT * FROM R;
Q2:  SELECT * FROM R
     ORDER BY a;
```

   (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.

7. In the following queries, the schema is $R(a, b)$.

Q_1: `SELECT MIN(a), MIN(b) FROM R;`
Q_2: `SELECT * 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.

8. The following are expressions in relational algebra about a relation $R(a, b, c)$.

$$\begin{array}{lll} Q_1: & S & := \rho_{S(a,b1,c1)}(R) \\ & T & := S \bowtie R \\ & Ans & := \pi_{b1,c}(\sigma_{b1=c}(T)) \\ Q_2: & Ans & := \pi_{b,c}(\sigma_{b=c}(R)) \end{array}$$

   (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.

9. In the following relational algebra expressions, R and S have the same schema, which includes attribute a, but the schemas are otherwise arbitrary.

$$\begin{array}{ll} Q_1: & \pi_a(R) - \pi_a(S) \\ Q_2: & \pi_a(R - S) \end{array}$$

   (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.

10. In the following, R has attribute a, but its schema is otherwise not specified, nor is it relevant.

Q_1: `SELECT COUNT(a) FROM R;`
Q_2: `SELECT COUNT(*) 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.

The remaining questions have no particular assumptions, except as stated. However, you must still assume all SQL is according to the SQL-99 standard, unless stated otherwise.

****

11. Suppose we have relations $R(a, b)$ and $S(a, c)$, and the assertion:

```
CREATE ASSERTION Mystery CHECK (
    NOT EXISTS (SELECT * FROM R, S WHERE R.a=S.a AND R.b<>S.c)
);
```

Which of the following combinations of tuples are prohibited from existing at the same time?

(a) $(1, 1)$ in R and $(2, 2)$ in S.

(b) $(1, 2)$ in R and $(2, 1)$ in S.

(c) $(1, 2)$ in R and $(1, 2)$ in S.

(d) $(1, 1)$ in R and $(1, 2)$ in S.

12. Consider the following trigger, which applies to a relation $R(a, b)$. Attribute a is the primary key for R.

```
CREATE TRIGGER Mystery1
AFTER DELETE ON R
REFERENCING OLD ROW AS OldRow
FOR EACH ROW
WHEN (OldRow.a = ANY(SELECT b FROM R))
INSERT INTO R VALUES(OldRow.a, OldRow.b);
```

Which statement describes the condition that is enforced by the trigger $Mystery1$ by itself?

(a) Every value that appears in $R.a$ also appears in $R.b$.

(b) Attribute $R.b$ is unique.

(c) Attribute $R.b$ is a foreign key, referencing $R.a$.

(d) None of the above.

13. Assume that relation $R(a)$ is initially empty. The following trigger exists on R:

```
CREATE TRIGGER Mystery2
AFTER INSERT ON R
REFERENCING NEW ROW AS NewRow
FOR EACH ROW
WHEN ( 10 > (SELECT AVG(a) FROM R))
BEGIN
    UPDATE R SET a = 2*a WHERE a = NewRow.a;
    INSERT INTO R VALUES(NewRow.a + 1)
END;
```

If we insert the tuple (1) into R, how many tuples will R have after the trigger is allowed to execute as many times as required?

(a) 8 (b) 9 (c) 10 (d) 11

14. In the relation $R(A, B, C, D, E)$ with functional dependencies $AB \rightarrow C$, $C \rightarrow B$, and $D \rightarrow E$, the number of superkeys is:

(a) 2 (b) 4 (c) 6 (d) 8

The next two questions refer to the following transactions and to a relation $R(a)$ that, before either transaction executes, contains only the tuple (1).

```
T1: INSERT INTO R VALUES(2);
    INSERT INTO R VALUES(3);
    DELETE FROM R WHERE a=1;

T2: SELECT SUM(a) FROM R;
```

15. If $T2$ runs with isolation level `SERIALIZABLE` while $T1$ runs with isolation level `READ UNCOMMITTED`, then the possible results returned by $T2$ are:

(a) 1 or 5 only.

(b) 1, 3, or 6 only.

(c) 1, 3, 5, or 6 only.

(d) None of the above.

16. If $T2$ runs with isolation level `READ UNCOMMITTED` while $T1$ runs with isolation level `SERIALIZABLE`, then the possible results returned by $T2$ are:

(a) 1 or 5 only.

(b) 1, 3, or 6 only.

(c) 1, 3, 5, or 6 only.

(d) None of the above.

17. Consider a database with relation R and users Alice, Bob, Carol,and Dave. Alice owns relation R. The following sequence of operations takes place:

```
Alice: GRANT SELECT ON R TO Bob WITH GRANT OPTION
Alice: GRANT SELECT ON R TO Carol WITH GRANT OPTION
Carol: GRANT SELECT ON R TO Bob WITH GRANT OPTION
Bob:   GRANT SELECT ON R TO Dave WITH GRANT OPTION
Carol: GRANT SELECT ON R TO Dave
Dave:  GRANT SELECT ON R TO Carol WITH GRANT OPTION
Alice: REVOKE SELECT ON R FROM Bob CASCADE
```

After these statements are executed, which of the following statements is true?

(a) Dave has the `SELECT ON R` privilege, but without the grant option.

(b) Dave has the `SELECT ON R` privilege with the grant option.

(c) Dave does not have the `SELECT ON R` privilege.

(d) Dave has the grant option for the `SELECT ON R` privilege, but does not have the privilege itself.

18. In SQL 3-valued logic, what is the value of `X < Y AND (X >= Y OR Z = 10)`?

(a) `TRUE` or `FALSE`, but never `UNKNOWN`

(b) `FALSE` or `UNKNOWN`, but never `TRUE`

(c) `TRUE` or `UNKNOWN`, but never `FALSE`

(d) `TRUE`, `FALSE`, or `UNKNOWN`

19. The relation $R(A, B, C, D)$ with functional dependencies $A \to B$, $B \to C$, and $BC \to A$ is:

(a) Not in 3NF.

(b) In 3NF but not in BCNF.

(c) In BCNF but not in 4NF.

(d) In 4NF.

20. The relation $R(A, B, C, D)$ with dependencies $AB \to C$, $ABC \to D$, and $AC \twoheadrightarrow B$ is:

(a) Not in 3NF.

(b) In 3NF but not in BCNF.

(c) In BCNF but not in 4NF.

(d) In 4NF.

```
<BIB>
   <BOOK isbn="1234">
      <TITLE>TCP/IP Illustrated</TITLE>
      <AUTHOR>Stevens</AUTHOR>
      <PUBLISHER>Addison-Wesley</PUBLISHER>
   </BOOK>
   <BOOK>
      <TITLE>Advanced Unix Programming</TITLE>
      <AUTHOR>Stevens</AUTHOR>
   </BOOK>
   <BOOK isbn="5678">
      <TITLE>Data on the Web</TITLE>
      <AUTHOR>Abiteboul</AUTHOR>
      <AUTHOR>Buneman</AUTHOR>
      <AUTHOR>Suciu</AUTHOR>
      <YEAR>2000</YEAR>
   </BOOK>
</BIB>
```

Figure 1: XML data contained in file books.xml

The next three questions refer to the document in Fig. 1.

21. The number of nodes produced by the XPath expression `/BIB/BOOK[@isbn]/AUTHOR` is:

    (a) 0 (b) 2 (c) 4 (d) 5

22. The result of the XQuery query:

```
let $d := document("books.xml")
for $b in $d/BIB/BOOK
where $b/AUTHOR = "Abiteboul"
return {$b/AUTHOR}
```

    is a sequence of how many `AUTHOR` nodes?

    (a) 0 (b) 1 (c) 3 (d) 9

23. Which of the following might appear in a DTD for which the document of Fig. 1 is valid?

```
(a)  <!ELEMENT BOOK (TITLE*, AUTHOR*, PUBLISHER?, YEAR*)>
(b)  <!ELEMENT BOOK (TITLE, AUTHOR+, (PUBLISHER | YEAR)>
(c)  <!ELEMENT BOOK (TITLE, AUTHOR*, PUBLISHER+, YEAR?)>
(d)  <!ELEMENT BOOK (TITLE?, AUTHOR?, PUBLISHER*, YEAR?)>
```

24. Consider the following table declarations:

```
CREATE TABLE R1 (
       a INT PRIMARY KEY,
       b INT
);
CREATE TABLE R2 (
       c INT PRIMARY KEY,
       d INT REFERENCES R1(a)
);
CREATE TABLE R3 (
       e INT PRIMARY KEY,
       f INT CHECK(f IN (SELECT a FROM R1))
);
```

Assume $R1$ contains the tuples (1,10), (2,10), and (3,20), while $R2$ and $R3$ are empty. Which of the following sequences of statements would **not** be allowed by SQL?

```
(a) INSERT INTO R3 VALUES(5,2);
    DELETE FROM R1 WHERE a=2;
    INSERT INTO R2 VALUES(1,1);

(b) INSERT INTO R2 VALUES(1,1);
    DELETE FROM R1 WHERE a=2;
    INSERT INTO R3 VALUES(6,1);

(c) INSERT INTO R3 VALUES(6,1);
    UPDATE R1 SET a=4 WHERE a=3;
    INSERT INTO R2 VALUES(10,2);

(d) DELETE FROM R1 WHERE a=2;
    INSERT INTO R2 VALUES(1,1);
    INSERT INTO R3 VALUES(5,2);
```

25. Which of the following is not a property of ACID transactions?

    (a) Atomicity
    (b) Concurrency
    (c) Isolation
    (d) Durability

26. If a document conforms to the following DTD:

```
<!DOCTYPE a [
    <!ELEMENT a (b, c?, d+)>
    <!ELEMENT b (c+, d)>
    <!ELEMENT c (d)>
         <!ATTLIST c e ID>
    <!ELEMENT d (#PCDATA)>
]>
```

What is the smallest number of elements that can appear in the document? Put another way, what is the smallest number of beginning tags (those without a "/") that can appear in the document?

(a) 1 (b) 3 (c) 6 (d) 9

27. A data cube for sales of pants has three dimensions:

i. Waist, with 8 different values.

ii. Inseam, with 6 different values.

iii. Color, with 10 different values.

What is the total number of marginals in the data cube?

(a) 24 (b) 188 (c) 213 (d) 693

28. If WalMart keeps information about its sales in a star schema, which of the following is **least** likely to be the schema of one of the dimension tables?

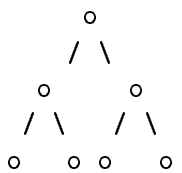
(a) Customers(name, address, phone, creditCard, email)

(b) Sales(customerName, item, date, quantity, price)

(c) Items(name, manufacturer, modelNo, color, size)

(d) Stores(storeID, city, state, address, phone)

29. Seven entity sets are arranged in a subclass hierarchy in the form of a balanced binary tree, as:



How many different subsets of these seven entity sets could be exactly the set of entity sets to which some entity of this hierarchy belongs?

(a) 7 (b) 8 (c) 17 (d) 25

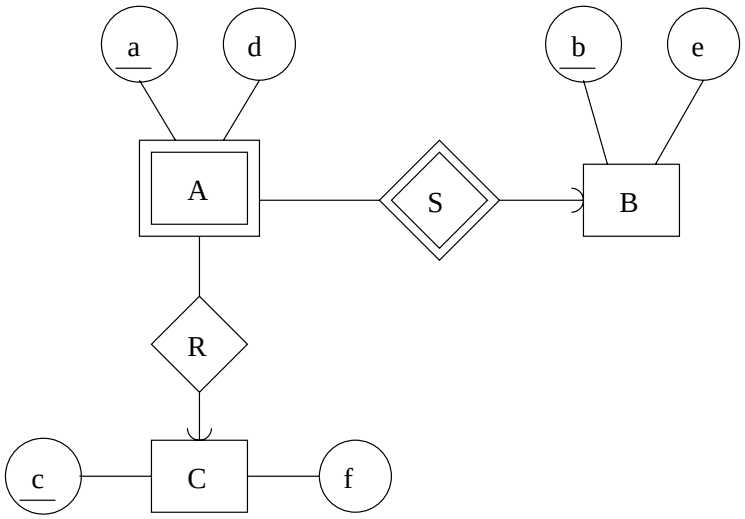


Figure 2: E/R diagram

The next two questions refer to the E/R diagram in Fig. 2 and its conversion to a relational database schema.

30. Which of the following entity sets or relationships does **not** get converted to a relation?

    (a) B (b) C (c) R (d) S

31. Which of the following relation schemas describes the relation for entity set A?

    (a) $A(a, b, d)$
    (b) $A(a, b, c, d)$
    (c) $A(a, d)$
    (d) $A(a, c, d)$

32. Relation $R(x)$ is a set of integers, and we want to test each one to see if any of the integers are prime. Presumably we have a Java function that tests primality of an integer; the manner of testing for primality is irrelevant. We are writing a program that gets each of the tuples of R and tests them. Our first steps are:

```
PreparedStatement s = myCon.createStatement(
    "SELECT * FROM R");
ResultSet r = s.executeQuery();
```

    where `myCon` is a connection to the database containing relation R. Which of the following is an appropriate next step?

    (a) `p = r.getInt(1);`
    (b) `p = r.getInt();`
    (c) `x = r.next();`
    (d) `while(r.next()) {p = r.getInt(1);}`

33. Suppose that in a C program using CLI, we have a statement s that has just been prepared by:

```
SQLPrepare(s, "SELECT a FROM R WHERE b = ?", SQL_NTS);
```

If we want to find those values of a such that tuple $(a, 10)$ is a member of the relation $R(a, b)$, which CLI function should we use next?

(a) `SQLExecute`

(b) `SQLExecDirect`

(c) `SQLBindCol`

(d) `SQLBindParameter`

34. Below is a table that summarizes how general select-from-where queries are dealt with in various "embedded SQL" approaches. The meanings of the second and third columns are for you to figure out.

PSM	Cursor	`FETCH`
CLI	Statement Handle	`SQLFetch(...)`
JDBC	xxx	`next()`

Notice that one of the entries appears as xxx. Which of the following is the most appropriate replacement for xxx?

(a) `Connection`

(b) `ResultSet`

(c) `Statement`

(d) `PreparedStatement`

35. Here are two user-defined types and a table declaration:

```
CREATE TYPE T1 AS (a INT, b INT);
CREATE TYPE T2 AS (c INT, d T1, e T1);
CREATE TABLE R OF T2;
```

The arity (number of columns) of R is:

(a) 1 (b) 3 (c) 5 (d) 7