

Your name: \_\_\_\_\_

## Database Design: Midterm 1

### CMSC424, Section 201, Fall 2015

- Numbers in square brackets indicate points (total 50 points = 12.5% of the course grade).
- **Show your reasoning. Write partial solutions.** You will get a fair amount of the credit if I think that you know the concepts.
- Unless otherwise specified, a *Yes/No* answer without accompanying explanation will not get any points.
- Do not write anything other than your name above the line at top. Try to keep your answers within the allotted space.

1. **[10 pts]** Circle true or false: **1 pt** for correct answer. **0.5 pts** will be deducted for incorrect.

(i) SQL EXCEPT operation eliminates duplicates. <b>Answer:</b> True	TRUE FALSE
(ii) FALSE AND UNKNOWN = FALSE. <b>Answer:</b> True	TRUE FALSE
(iii) Dynamic SQL typically requires use of a language-specific preprocessor to compile/prepare SQL statements, whereas Embedded SQL doesn't. <b>Answer:</b> False	TRUE FALSE
(iv) Using Prepared Statements leaves the system vulnerable to SQL Injection attacks. <b>Answer:</b> False	TRUE FALSE
(v) In an E/R model, attributes can also be attached to relationships. <b>Answer:</b> True	TRUE FALSE
(vi) Under the "multi-set/bag" semantics, $\{a, b\} \times \{c, c\} = \{(a, b), (c, c), (a, c), (b, c)\}$ <b>Answer:</b> False	TRUE FALSE
(vii) UNKNOWN OR TRUE = UNKNOWN. <b>Answer:</b> False	TRUE FALSE
(viii) On table instructor( <u>ID</u> , name, dept name, salary), ID (which is the primary key) can take the value of null if ID of the instructor is unknown/undisclosed. <b>Answer:</b> False	TRUE FALSE
(ix) Relational algebra is a non-procedural language. <b>Answer:</b> False	TRUE FALSE
(x) The keyword 'where' in SQL maps to the $\sigma$ operation in Relational Algebra. <b>Answer:</b> True	TRUE FALSE

2. **[2 pts]** What is a good primary key for the following relation, that represents information about US presidents ? Explain your reasoning.

President(name, start\_date, end\_date, vice\_president, preceded\_by, succeeded\_by)

**Answer:** (start\_date) is a good primary key for this table, especially if it includes the time.

3. **[3 pts]** Consider two create table statements: (a) create table R (a integer primary key);  
(b) create table S (b integer primary key, c integer references R(a) on update set to null); .

What will happen when a tuple in R is updated or deleted?

**Answer:** If a tuple in R is updated, for the corresponding tuples in S, the foreign key is set to null. Deletes to R are rejected if there are corresponding tuples in S.

Your name: \_\_\_\_\_

4. **[3 pts]** Consider a view  $V$  defined on relation  $R(a, b)$  where  $b$  is a numeric attribute, as follows:

**create view V as select a, sum(b) as sumb from R group by a;**

Discuss whether we should allow inserts, deletes, and updates into the view, and how to propagate them to underlying relation (the answer may be different for different operations).

**Answer:** We can allow deletes without any semantic ambiguity - just delete all the corresponding rows in  $R$ . Updates should not be allowed. Inserts may be allowed but are somewhat ambiguous – inserting a row  $(x, y)$  into  $V$  will essentially result in insertion of row  $(x, y)$  in  $R$ .

5. **[3 pts]** Briefly explain the notion of “triggers” in SQL, including an example of where it might be used.

**Answer:** See slides from class or the book.

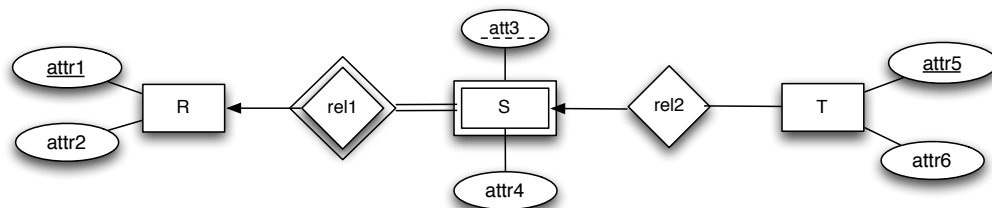
6. **[3 pts]** Given a table called  $T$  with three integer attributes called  $a$ ,  $b$ , and  $c$ , write a SQL query that returns the sum of all the values of  $c$  for every unique combination of  $a$  and  $b$ .

**Answer:** SELECT  $a, b, \text{sum}(c)$  FROM  $T$  GROUP BY  $a, b$ ;

7. **[3 pts]** What is “specialization” in E/R modeling? Illustrate through an example.

**Answer:** See slides from class on ER models or the textbook.

8. **[3 pts]** Which set of relations do we end up with when converting the converting the following E/R diagram to a relational schema? **Answer:**



- $R(\text{attr1}, \text{attr2})$
- $S(\text{attr1}, \text{attr3}, \text{attr4})$
- $T(\text{attr5}, \text{attr6}, \text{attr3}, \text{attr1})$

9. **[4 pts]** On the relation shown below, compute the answers to the two queries listed.

A	B	C
$\alpha$	a	1
$\beta$	b	2
$\alpha$	c	1
$\alpha$	a	3
$\gamma$	c	3
$\gamma$	a	2
$\beta$	b	2
$\alpha$	c	2
$\beta$	c	2

- (i) select  $A, \text{max}(C)$  from  $R$  where  $C \neq \text{null}$  group by  $A$ ;  
 (ii) select  $A$  from  $R$  r1 where exists  
 (select \* from  $R$  r2 where  $r2.A = r1.A$  and  $r2.C < r1.C$ );

**Answer:** (i) empty set / nothing at all (ii)  $(\alpha), (\gamma), (\alpha)$

Your name: \_\_\_\_\_

10. **[4 pts]** Given the relations,  $R(a, b)$ ,  $S(b, c)$ , where  $a, b, c$  are integer attributes, describe in words what are the results of following relational algebra expressions.

(a)  $\sigma_{R.b \neq S.b}(R \times S) - (R \times S)$

**Answer:**  $\phi$  (Empty).

(b)  $R - \pi_{R.a, R.b}(\sigma_{R.b = S.b}(R \times S))$

**Answer:** It finds tuples of  $R$  that have no matching tuples in  $S$  (i.e. it computes  $R$  ANTIJOIN  $S$ ).

11. **[4 pts]** Give an example of a relation and a SQL query that uses a self-join over that relation that returns exactly 3 rows.

**Answer:** Relation  $T$ , two columns called  $a$  and  $b$ , contents:  $(1,1), (1,2), (1,3)$ ;

**Query:** `select * from T as t1, T as t2 where t1.a = t2.b;`

12. **[4 pts]** The following two queries are not equivalent (they don't always produce identical results) because of NULLs. Identify and explain the problem. Schemas are:  $R(a, b, d)$ ,  $S(c, d)$ . Assume  $a$  is the primary key for  $R$ .

<u>Query I</u>	<u>Query II</u>
<b>select a</b> <b>from R</b> <b>where R.b = (select count(S.c)</b> <b>from S</b> <b>where R.d = S.d)</b>	<b>select a</b> <b>from R, S</b> <b>where R.d = S.d</b> <b>group by R.a</b> <b>having R.b = count(S.c);</b>

**Answer:** Consider a tuple in  $R$ :  $(R.a = \alpha, R.b = 0, R.d = \beta)$ , and say there is no tuple in  $S$  such that  $S.d = \beta$ . The first query will generate the answer tuple  $\alpha$ , whereas the second query will not.

13. **[4 pts]** Suppose we have three relations  $r(A, B)$ ,  $s(B, C)$ , and  $t(B, D)$ , with all attributes declared as not null. Consider the expressions

- $r$  natural left outer join  $(s$  natural left outer join  $t)$ , and
- $(r$  natural left outer join  $s)$  natural left outer join  $t$

(a) Give instances of relations  $r$ ,  $s$  and  $t$  such that in the result of the second expression, attribute  $C$  has a null value but attribute  $D$  has a non-null value.

**Answer:**  $r = \{(1, 1)\}$ ,  $s = \phi$  (empty),  $t = \{(1, 1)\}$

(b) Is the above pattern, with  $C$  null and  $D$  not null, possible in the result of the first expression? Explain why or why not.

**Answer:** Because the input relations cannot have any nulls: for any tuple in  $(s$  natural left outer join  $t)$ ,  $C$  cannot be null ( $D$  could be null). When joining that with  $r$ , the only way to get  $C = \text{null}$  is if there is no match, in which case  $D$  will also be null.