CSE-3421M Test #2

"Queries"

Family Name: ______________________________

Given Name: ______________________________

Student#: ______________________________

CS&E Account: ______________________________

- **Instructor:** Parke Godfrey
- **Exam Duration:** 75 minutes
- **Term:** Winter 2013

Answer the following questions to the best of your knowledge. Your answers may be brief, but be precise and be careful. The exam is closed-book and closed-notes. Calculators, etc., are fine to use. Write any assumptions you need to make along with your answers, whenever necessary.

There are four major questions, each with parts. Points for each question and sub-question are as indicated. In total, the test is out of 50 points.

In schemas, the underlined attributes denote a table's key. Attributes that are in *italics* are not nullable. Foreign keys are indicated by FK.

If you need additional space for an answer, just indicate clearly where you are continuing.

Marking Box	
1.	/15
2.	/15
3.	/10
4.	/10
Total	/50

1. (15 points) **SQL.** *San Carlos Municipal Airport.* [Exercise]

Congratulations! You have just made the major leagues by landing a job with **Amazing.com**, the famous Internet bookstore.

Customer(cno, *cname*, *address*)
Author(ano, *aname*, birth, country)
Book(isbn, *title*, *year*, *publisher*, *language*)
Wrote(ano, isbn)
    FK (ano) refs **Author**
    FK (isbn) refs **Book**

Stock(isbn, from, *qnty*, *price*)
    FK (isbn) refs **Book**
Purchase(cno, isbn, when, *from*, *qnty*)
    FK (cno) refs **Customer**
    FK (isbn, from) refs **Stock**
Payment(cno, when, *amount*)
    FK (cno) refs **Customer**

Figure 1: **Amazing.com** Schema.

The basic schema of the main database at **Amazing.com** is shown in Figure 1. The underlined attributes indicate a table's primary key (and are hence not nullable). Attributes marked in *italics* are not nullable. Foreign keys are indicated by FK.

The attribute price in **Stock** tells how much a customer pays for that book. The attribute from in **Purchase** indicates where the stock was pulled from.

a. (6 points) Write an SQL query to report customers by cno, name, and address who have ever bought (**Purchase**) a book authored by an author whose nationality (country) is *Canadian.*

Do not report duplicate rows.

b. (6 points) Write an SQL query to report customers by `cno`, name, and address who have ever bought (**Purchase**) a book that has two or more authors for more than $200.

Do not report duplicate rows.

c. (3 points) Consider the SQL query

```
select cname, cno, aname, ano
    from Customer, Author
except
select cname, cno, aname, ano
    from (  select cno, ano, isbn
                from Customer, Wrote
            except
            select P.cno, W.ano, W.isbn
                from Purchase P, Wrote W
                where P.isbn = W.isbn
        ) as X,
        Customer C, Author A
    where X.cno = C.cno and X.ano = A.ano;
```

State in *plain, concise English* what the query does.

You get *zero* credit if you use database terms in your answer! (E.g., “Well, the query first *joins* two tables, taking the *projection* of. . . ” does not count!)

2. (15 points) **General.** *All the beautiful choices!* [Multiple Choice]

Choose *one* best answer for each of the following. Each is worth one point. There is no negative penalty for a wrong answer.

a. The SQL "ORDER BY" clause

A. can be used to order the results of any sub-query.
B. can only be used as the last clause of a query to order the tuples of the answer.
C. may affect the results of a join.
D. may only be used when "GROUP BY" is used.
E. is required in any SQL query.

b. The SQL statement "DELETE FROM R;"

A. will remove all the tuples from **R**.
B. will remove no tuples from **R**.
C. may remove just some tuples from **R**.
D. is syntactically incorrect because it is missing a WHERE clause.
E. *Not enough information to determine.*

c. In a *real* relational database system, if you try to join (natural join) tables **R** and **S** *and* **R** is empty (that is, it has no tuples),

A. the system reports an error.
B. the answer set is an empty table.
C. the answer set is the same as table **S**.
D. the answer set consists of just one row.
E. an answer set is returned; however, the results are system dependent.

For Questions 2d and 2e, consider a table **T** with attributes A and B.

d. Which of the following SQL queries is illegal?

A. `select * from T;`
B. `select count(*) from T;`
C. `select count(*) from T group by A;`
D. `select count(*), max (B) from T group by A;`
E. `select max(count(*)) from T group by A;`

e. Which of the following SQL queries is illegal?

A. `select A from T;`
B. `select A, count(*) from T;`
C. `select A, count(*) from T group by A;`
D. `select A, count(*) from T group by A, B;`
E. `select A, B, count(*) from T group by A, B;`

f. In relational algebra, the intersection operator ($\cap$) is *not* logically redundant if we only have additionally

A. join ($\bowtie$)
B. crossproduct ($\times$), select (σ), and project (π)
C. difference ($-$) and union ($\cup$)
D. crossproduct ($\times$) and difference ($-$)
E. crossproduct ($\times$) and union ($\cup$)

g. $\mathbf{R} \cap \mathbf{S}$ is equivalent to

A. $\mathbf{R} - (\mathbf{R} - \mathbf{S})$
B. $\mathbf{R} - (\mathbf{S} - \mathbf{R})$
C. $(\mathbf{R} - \mathbf{S}) - \mathbf{R}$
D. $-\,((-\,\mathbf{R}) \cup (-\,\mathbf{S}))$
E. *There is not enough information to answer this.*

h. Consider the schema

$\mathbf{R}(\underline{\mathsf{A}}, B)$ FK (B) refs $\mathbf{S}$
$\mathbf{S}(A, \underline{\mathsf{B}})$ FK (A) refs $\mathbf{R}$

Note that none of the attributes are nullable.

Which of the following is guaranteed to produce as many as, or more, tuples than each of the others?

A. $\mathbf{R} \bowtie \mathbf{S}$
B. $\pi_{\mathsf{A}}(\mathbf{R}) \bowtie \mathbf{S}$
C. $\mathbf{R} \bowtie \pi_{\mathsf{B}}(\mathbf{S})$
D. $\pi_{\mathsf{A}}(\mathbf{R}) \bowtie \pi_{\mathsf{B}}(\mathbf{S})$
E. *There is not enough information to answer this.*

i. Consider the schema from Question 2h again.

Which of the following is guaranteed to produce as many as, or more, tuples than each of the others?

A. $\mathbf{R} \bowtie \mathbf{S}$
B. $\mathbf{R} \times \mathbf{S}$
C. $\mathbf{R} \cap \mathbf{S}$
D. $\mathbf{R} \cup \mathbf{S}$
E. *There is not enough information to answer this.*

j. Say that A is the primary key of table $\mathbf{R}$, and that B is another, numeric attribute of $\mathbf{R}$. Let r be the number of tuples in table $\mathbf{R}$.

$$(\pi_{\mathsf{A}}(\sigma_{\mathsf{B}<5}(\mathbf{R}))) \bowtie (\pi_{\mathsf{A}}(\sigma_{\mathsf{B}\geq 5}(\mathbf{R})))$$

How many tuples will result from the above RA expression?

A. 0 **B.** 1 **C.** $\frac{1}{2}r$ **D.** r **E.** $2r$

k. Consider the following schema.

```
create table T (
    c integer primary key,
    d integer);
create table S (
    b integer primary key,
    c integer references T(c) on delete cascade);
create table R (
    a integer primary key,
    b integer references S(b) on delete set null);
```

Let tables **R**, **S**, and **T** contain r, s, and t tuples, respectively. Let n be the number of tuples in the result of the following query:

```
select *
    from R, S, T
    where R.B = S.B and S.C = T.C;
```

What is the lowest bound on n that you can say with certainty?

A. $n \leq r$

B. $n \leq t$

C. $n \leq r * s * t$

D. $n \leq min(r, s, t)$

E. *It is impossible to establish a bound with the information provided.*

For Questions 2l to 2n, one of the choices is not like the others. That is, one of the choices could evaluate to a different answer than the others do. Choose the one that may evaluate differently.

l. Consider the relations $\mathbf{R}(\mathsf{A},\mathsf{B})$, $\mathbf{S}(\mathsf{B},\mathsf{C})$, and $\mathbf{T}(\mathsf{C},\mathsf{A})$.

A. $\pi_{\mathsf{A,B}}((\mathbf{R} \bowtie \mathbf{S}) \bowtie \mathbf{T})$

B. $\pi_{\mathsf{A,B}}(\mathbf{R} \bowtie (\mathbf{T} \bowtie \mathbf{S}))$

C. $\mathbf{R} \bowtie \pi_{\mathsf{A,B}}(\mathbf{S} \bowtie \mathbf{T})$

D. $\pi_{\mathsf{A,B}}(\mathbf{R} \bowtie \mathbf{S}) \bowtie \pi_{\mathsf{A,B}}(\mathbf{S} \bowtie \mathbf{T})$

E. $\pi_{\mathsf{A,B}}(\mathbf{R} \bowtie \mathbf{T}) \bowtie \pi_{\mathsf{A,B}}(\mathbf{R} \bowtie \mathbf{S})$

m. Consider the schema $\mathbf{R}(\underline{\mathsf{A}},\mathsf{B},\mathsf{C})$, $\mathbf{S}(\underline{\mathsf{A}},\underline{\mathsf{D}})$, and $\mathbf{T}(\underline{\mathsf{D}},\mathsf{B},\mathsf{E})$.

A. $\pi_{\mathsf{B}}((\sigma_{\mathsf{C}=5}(\mathbf{R}) \bowtie \sigma_{\mathsf{E}=7}(\mathbf{T})) \bowtie \mathbf{S})$

B. $\pi_{\mathsf{B}}(\pi_{\mathsf{A,B,D}}(\sigma_{(\mathsf{C}=5)\wedge(\mathsf{E}=7)}(\mathbf{R} \bowtie \mathbf{T})) \cap (\mathbf{S} \times \pi_{\mathsf{B}}(\mathbf{T})))$

C. $\pi_{\mathsf{B}}((\sigma_{\mathsf{C}=5}(\mathbf{R} \bowtie \mathbf{S})) \bowtie (\sigma_{\mathsf{E}=7}(\mathbf{S} \bowtie \mathbf{T})))$

D. $\pi_{\mathsf{B}}(\sigma_{\mathsf{C}=5}(\mathbf{R} \bowtie \mathbf{S})) \cap \pi_{\mathsf{B}}(\sigma_{\mathsf{E}=7}(\mathbf{S} \bowtie \mathbf{T}))$

E. $\pi_{\mathsf{B}}(\pi_{\mathsf{A,B}}(\sigma_{\mathsf{C}=5}(\mathbf{R})) \bowtie (\mathbf{S} \bowtie \pi_{\mathsf{D,B}}(\sigma_{\mathsf{E}=7}(\mathbf{T}))))$

n. Consider the schema $\mathbf{R}(\underline{\mathsf{A}},\underline{\mathsf{B}})$, and $\mathbf{S}(\underline{\mathsf{B}},\underline{\mathsf{C}})$. (There are no FKs!)

A. $\pi_{\mathsf{A,C}}(\mathbf{R} \bowtie \mathbf{S})$

B. $\{\langle A, C\rangle \mid \exists B(\langle A, B\rangle \in \mathbf{R} \rightarrow \langle B, C\rangle \in \mathbf{S})\}$

C. $\{\langle A, C\rangle \mid \neg\, \mathsf{B}(\langle A, B\rangle \in \mathbf{R} \rightarrow \langle B, C\rangle \not\in \mathbf{S})\}$

D. `select distinct R.A, S.C from R, S where R.B = S.B;`

E.
```
select distinct R.A, S.C from R, S s1
     where R.B in (select s2.B from S s2 where s1.C = s2.C);
```

o. Consider the schema $\mathbf{R}(\mathsf{A},\mathsf{B})$, $\mathbf{S}(\mathsf{B},\mathsf{C})$, and $\mathbf{T}(\mathsf{C},\mathsf{A})$. (There are no FKs!)

A. $\{\langle \mathsf{A}\rangle \mid \neg\exists B, C(\langle \mathsf{A, B}\rangle \in \mathbf{T} \wedge \langle \mathsf{B, C}\rangle \in \mathbf{S} \wedge \langle \mathsf{C, A}\rangle \in \mathbf{T})\}$

B. $\{\langle \mathsf{A}\rangle \mid \neg\exists B(\langle \mathsf{A, B}\rangle \in \mathbf{T} \wedge \exists C(\langle \mathsf{B, C}\rangle \in \mathbf{S} \wedge \langle \mathsf{C, A}\rangle \in \mathbf{T}))\}$

C. $\{\langle \mathsf{A}\rangle \mid \forall \mathsf{B}(\langle \mathsf{A, B}\rangle \in \mathbf{R} \rightarrow$
$\qquad\forall \mathsf{C}(\langle \mathsf{C, A}\rangle \not\in \mathbf{T} \vee \langle \mathsf{B, C}\rangle \not\in \mathbf{S}))\}$

D. $\{\langle \mathsf{A}\rangle \mid \forall \mathsf{B}(\langle \mathsf{A, B}\rangle \in \mathbf{R} \wedge$
$\qquad\forall \mathsf{C}(\langle \mathsf{B, C}\rangle \in \mathbf{S} \wedge$
$\qquad\qquad\forall \mathsf{A2}(\langle \mathsf{C, A2}\rangle \in \mathbf{T}$
$\qquad\qquad\qquad\rightarrow \mathsf{A} \neq \mathsf{A2})))\}$

E. $\{\langle \mathsf{A}\rangle \mid \forall \mathsf{C}(\langle \mathsf{C, A}\rangle \in \mathbf{T} \wedge$
$\qquad\forall \mathsf{B}(\langle \mathsf{B, C}\rangle \in \mathbf{S} \wedge$
$\qquad\qquad\forall \mathsf{A2}(\langle \mathsf{A2, B}\rangle \in \mathbf{R}$
$\qquad\qquad\qquad\rightarrow \mathsf{A} \neq \mathsf{A2})))\}$

3. (10 points) **Relational Algebra & Calculus.** *The area under what?!* [Short Answer]

R	
A	B
1	2
3	2
5	6
7	8
9	8

S	
B	C
6	2
2	4
8	1
8	3
2	5

T	
A	C
7	1
1	2
9	3
5	4
3	5

Three tables: **R**, **S**, & **T**.

a. (3 points) Show the results of **R** $\bowtie$ **S**.

b. (3 points) Show the results of **R** $\bowtie$ (**S** $\bowtie$ **T**).

c. (2 points) Consider any three tables **J**, **K**, and **L**. Is it possible that $\mathbf{J} \bowtie (\mathbf{K} \bowtie \mathbf{L})$ and $(\mathbf{J} \bowtie \mathbf{K}) \bowtie \mathbf{L}$ might evaluate to different tables?

If so, construct a *small* example that demonstrates this.

If not, *briefly* explain why.

d. (2 points) Which is SQL based upon, the *domain* relational calculus or the *tuple* relational calculus?

Briefly justify your answer.

Extra space.

4. (10 points) **Aggregation.** *Groups for a thousand, Alex.* [Exercise]

Person(sin, name, gender, job, studio, title, year)
    FK (studio) references **Studio**,
    FK (title, year) references **Movie**
Category(cat)
Movie(title, year, length, *director*, *studio*, *cat*)
    FK (director) references **Person** (sin),
    FK (studio) references **Studio**,
    FK (cat) references **Category**
Cast(sin, title, year, character)
    FK (sin) references **Person**,
    FK (title, year) references **Movie**
Studio(studio, city)

Figure 2: Movies and Casts.

Write SQL queries for the following with respect to the schema in Figure 2.

a. (3 points) For each *director* (someone who has *directed* at least one movie), report the person's sin, name, gender, and the number of movies that he or she has directed. Call this last column #movies.

b. (4 points) For each actor (someone who has been *cast* in a movie) and movie category (cat), report the sin, name, gender, and title of the actor and the number of movies of *that* category in which he or she has been cast. Call this last column #movies.

Only count a given actor or actress *once* per movie; that a person may have played several *characters* in the movie does *not* count multiple times.

Extra space.

c. (3 points) For each movie, report how many male actors and how many female actresses—gender = 'M' for *male* and gender = 'F' for *female*—were *cast* in the movie in columns male and female, respectively.

   Again, only count a given actor or actress *once* per movie; that a person may have played several *characters* in the movie does *not* count multiple times.

   The column male or female should report 0 (zero) if there were no actors or actresses, respectively, in the movie.

Extra space.

Happy Pi Day!

Relax. Breathe. Turn in your test. Go home.