

# CSE 344 Midterm

Wednesday, November 7, 2012, 9:30-10:20

Name: \_\_\_\_\_

Question	Points	Score
1	45	
2	30	
3	25	
Total:	100	

- This exam is open book and open notes but NO laptops or other portable devices.
- You have 50 minutes; budget time carefully.
- Please read all questions carefully before answering them.
- Some questions are easier, others harder. Plan to answer all questions, do not get stuck on one question. If you have no idea how to answer a question, write your thoughts about the question for partial credit.
- Good luck!

# 1 SQL and Physical Tuning

1. (45 points)

You have been analyzing the data from a social networking site and have derived the following relation, which captures topics discussed by various users.

Discussion(*user1*,*user2*,*topic*)

The relation contains a tuple (*u1*,*u2*,*t*) every time a user *u1* discussed a topic *t* with user *u2*. To avoid duplicate entries, *user1* always precedes *user2* in alphabetical order.

(a) (15 points) Write a SQL query that returns all topics discussed by **Alice** and **Bob** but not discussed by **Alice** and **Chuck**.

`Discussion(user1,user2,topic)`

- (b) (15 points) Write a SQL query that returns the number of topics discussed by more than 10 pairs of users.

- (c) (5 points) Give two reasons why database administrators typically do **NOT** create an index on every single attribute of every single relation. You do not need to discuss the reasons. Just state them.

- (d) (10 points) Explain how a database administrator should proceed in order to select a good set of indexes for a relational database. Note that more complete answers will receive more points.

## 2 Relational Algebra, Datalog, and Relational Calculus

2. (30 points)

Consider the following database schema. Relation **Clinic** lists medical clinics with their unique identifiers, names, street addresses, and states. Relation **Equipment** lists the unique identifiers, types, and models of various pieces of equipment. Finally, relation **Assignment** indicates the equipment available in each clinic.

**Clinic**(cid, name, street, state)

**Equipment**(eid, type, model)

**Assignment**(cid, eid)

- (a) (10 points) Write a Relational Algebra expression in the form of a logical query plan that is equivalent to the SQL query below:

```
select count(*)
from Clinic C
where not exists
  (select *
   from Assignment A, Equipment E
   where C.cid = A.cid
        and A.eid = E.eid
        and E.type = 'Fridge'
        and E.model = 1004
  )
```

```
Clinic(cid, name, street, state)
Equipment(eid, type, model)
Assignment(cid, eid)
```

(b) (10 points) Write a Datalog query equivalent to the following SQL query:

```
select C.name
from Clinic C
where not exists
  (select *
   from Assignment A, Equipment E
   where C.cid = A.cid
        and A.eid = E.eid
        and E.type = 'Fridge'
        and E.model = 1004
  )
```

```
Clinic(cid, name, street, state)
Equipment(eid, type, model)
Assignment(cid, eid)
```

- (c) (10 points) Write a relational calculus query that returns the types of equipment assigned to clinics in the state of WA:

### 3 XML and XPath

3. (25 points)

(a) (15 points) Consider the following XML document stored in a file called trips.xml:

```
<trips>
  <business reason='Meeting at CompanyX' destination='Baltimore'>
    <airline>American</airline>
    <stops>
      <location>Houston</location>
      <location>Boston</location>
    </stops>
  </business>

  <personal destination='Boston'>
    <airline>American</airline>
    <stops>
      <location>Chicago</location>
    </stops>
  </personal>

  <personal destination='Hawaii'>
    <airline>Alaska</airline>
    <stops>
    </stops>
  </personal>
</trips>
```

Write an XQuery expression that will transform it into the following document:

```
<trips>
  <airline>
    <name>American</name>
    <trip destination="Baltimore">
      <stops>2</stops>
    </trip>
    <trip destination="Boston">
      <stops>1</stops>
    </trip>
  </airline>
  <airline>
    <name>Alaska</name>
    <trip destination="Hawaii">
      <stops>0</stops>
    </trip>
  </airline>
</trips>
```

(b) (10 points) Write a possible DTD for the document used as **input** above: