

CSE 344 Midterm

Monday, November 9th, 2015, 9:30-10:20

Name: _____

Question	Points	Score
1	30	
2	30	
3	10	
Total:	70	

- This exam is CLOSED book and CLOSED devices.
- You are allowed ONE letter-size page with notes (both sides).
- 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 Indexing

1. (30 points)

Consider the following database for a summer camp:

```
Camper(cid, name)
Team(gid, age)
Assignment(cid, gid)
```

Relation **Camper** contains a tuple for each camper with name **name** and unique id **cid**. Each tuple in relation **Team** holds the unique id of a team and the age of the campers assigned to that team. Finally, table **Assignment** captures the assignments of campers to teams. Primary keys are underlined.

(a) (10 points) For each query below, indicate if the query correctly returns the **name** of all campers who have not yet been assigned to a team. Anywhere between zero and all queries compute the correct answer. Each query is syntactically correct.

1.

```
SELECT C.name
FROM Camper C LEFT OUTER JOIN Assignment A
ON C.cid = A.cid
WHERE A.gid IS NULL
```

2.

```
SELECT C.name
FROM Camper C
WHERE C.cid NOT IN (SELECT A.cid FROM Assignment A)
```

3.

```
SELECT C.name
FROM Camper C INNER JOIN Assignment A
ON C.cid = A.cid
GROUP BY C.cid, C.name
HAVING count(*) = 0
```

Camper(cid,name)
Team(gid,age)
Assignment(cid,gid)

- (b) (10 points) Write a SQL query that computes the number of campers in each age group but only for age groups below 9 years old. Note that multiple teams can have campers with the same age.

Camper(cid,name)
Team(gid,age)
Assignment(cid,gid)

- (c) (10 points) Consider the following query. For each index below, indicate if it can potentially speed up the query if it is the only index that exists.

```
SELECT C.name  
FROM Camper C, Team T, Assignment A  
WHERE C.cid = A.cid  
AND T.gid = A.gid  
AND T.age > 6
```

1. Index on Camper(cid)
2. Index on Camper(name)
3. Index on Camper(cid,name)
4. Index on Camper(name,cid)
5. Index on Team(gid)
6. Index on Team(age)
7. Index on Team(gid,age)
8. Index on Team(age,gid)

2 Relational Algebra, Datalog, and Relational Calculus

2. (30 points)

Consider the following database schema. Relation `Item` lists objects with their unique id (`oid`), category, and price. Relation `Gift` has one tuple for every person `pid` that offered a gift to a recipient `rid`. `Gift.oid` references `Item.oid`.

```
Item(oid,category,price)
Gift(pid, rid, oid)
```

(a) (10 points) Write a Relational Algebra expression in the form of a logical query plan (i.e., draw a tree) that is equivalent to the SQL query below.

```
SELECT O1.category, MAX(ABS(O1.price - O2.price))
FROM Gift G1, Gift G2, Item O1, Item O2
WHERE G1.pid = G2.rid
AND G2.pid = G1.rid
AND O1.oid = G1.oid
AND O2.oid = G2.oid
AND O1.category = O2.category
GROUP BY O1.category
HAVING count(*) > 5
```

Item(oid,category,price)
Gift(pid, rid, oid)

- (b) (10 points) Write a Datalog query that returns the identifier of all people who offered or received a ring as a gift but never offered nor received a book.

Item(oid, category, price)
Gift(pid, rid, oid)
Person(pid)

(c) (10 points) State what each of the following relational calculus queries computes.

(a)

$$P(x) = \exists o. \exists c. \exists p. \exists r. (\text{Person}(x) \wedge \text{Item}(o, c, p) \wedge \text{Gift}(x, r, o))$$

(b)

$$P(x) = \text{Person}(x) \wedge (\forall o. \forall c. \forall p. (\text{Item}(o, c, p) \Rightarrow \exists r. \text{Gift}(x, r, o)))$$

(c)

$$P(x) = \text{Person}(x) \wedge (\forall o. \forall r. (\text{Gift}(x, r, o) \Rightarrow \exists p. \text{Gift}(p, x, o)))$$

3 Query Evaluation

3. (10 points)
 - (a) (10 points) Describe how a relational database management system (DBMS) evaluates a query. (A) What happens from the moment the DBMS receives a SQL statement in the form of a string to the moment it returns results to the user? More complete answers will receive more points. (B) How does the presence of an index affect what happens?