

CISC 7510X Final Exam

For the below questions, use the following schema definition.

```
employee(eid, name, email, phone, position, mid)
client(cid, name, email, phone, street, city, state, zip)
presentation(pid, cid, timestamp, description)
pteam(pid, eid)
sale(sid, cid, pid, amnt, timestamp)
plog(lid, pid, notes, timestamp)
```

It is a schema for a marketing company. There are **employees** and **clients**. One or more employees visit the client site to do a **presentation**. The **pteam** tracks which employees are part of which presentations. The **sale** object tracks resulting sales linked to a presentation. There could be multiple sales for each presentation. The **plog** object records notes related to a presentation.

Each question is worth 5 points. You get 1 point for leaving an answer blank. You get no points for a wrong answer.

1. Find client 'ABC PHARMACY' zip code.
 - (a) select zip from employee where name='ABC PHARMACY'
 - (b) select zip from client where name='ABC PHARMACY'
 - (c) select zipcode from sale where name='ABC PHARMACY'
 - (d) select * from employee inner join presentation on employee.eid and presentation.eid
 - (e) Other:
2. Find all presentations for client 123.
 - (a) select * from presentation where cid=123
 - (b) select * from pteam inner join presentation on pid
 - (c) select * from employee natural inner join presentation natural inner join client
 - (d) select * from plog where notes='Client 123 presentation'
 - (e) Other:
3. Find all presentations that had more than 5 employees.
 - (a) select pid from presentation group by pid having count(*) > 5
 - (b) select pid from presentation natural inner join pteam where count(*) > 5
 - (c) select a.pid from presentation a inner join pteam b on a.pid=b.pid group by a.pid having count(*) > 5
 - (d) select pid from pteam group by pid having count(*) > 5
 - (e) Other:

4. Find all presentations that had less than 5 employees.
 - (a) `select pid from presentation group by pid having count(*) < 5`
 - (b) `select pid from presentation natural inner join pteam where count(*) < 5`
 - (c) `select a.pid from presentation a left outer join pteam b on a.pid=b.pid group by a.pid having count(*) < 5`
 - (d) `select pid from pteam group by pid having count(*) < 5`
 - (e) Other:

5. List all the clients that John Doe visited.
 - (a) `select * from client where name='John Doe'`
 - (b) `select cid from presentation natural inner join pteam natural inner join employee natural inner join client where name='John Doe'`
 - (c) `select d.cid from employee a inner join pteam b on a.eid=b.eid inner join presentation c on b.pid=c.pid inner join client d on c.cid=d.cid where a.name='John Doe'`
 - (d) `select c.cid from employee a inner join pteam b on a.eid=b.eid inner join presentation c on b.pid=c.pid where a.name='John Doe'`
 - (e) Other:

6. How many different types of employee positions are there?
 - (a) `select count(*) from employee`
 - (b) `select count(distinct position) from employee`
 - (c) `select count(*) from employee group by position`
 - (d) `select position, count(*) from employee`
 - (e) Other:

7. What percentage of clients are in NY tri-state area (NY,NJ,CT)?
 - (a) `select 100*sum(
 case when state in ('NY','NJ','CT') then 1.0 else 0.0 end)/sum(1.0)
 from client
 where state in ('NY','NJ','CT')
 group by state
 having state in ('NY','NJ','CT')`
 - (b) `select 100*sum(
 case when state in ('NY','NJ','CT') then 1.0 else 0.0 end)/sum(1.0)
 from client
 group by state`
 - (c) `select 100*sum(
 case when state in ('NY','NJ','CT') then 1.0 else 0.0 end)/sum(1.0)
 from client`

(d) `select 100*sum(
 case when state in ('NY','NJ','CT') then 1.0 else 0.0 end)/sum(1.0)
 from client
 where (case when state in ('NY','NJ','CT') then 'NYTRI' else 'NOT' end)='NYTRI'
 group by case when state in ('NY','NJ','CT') then 'NYTRI' else 'NOT' end
 having count(*)>0`

(e) Other:

8. Find potential client dups, street,city,state,zip are the same for different clients. List the street,city,state,zip and number of dups.

(a) `select a.street,a.city,a.state,a.zip, 'DUP' as lbl from client a
 inner join contact b
 on a.street=b.street and a.city=b.city and a.state=b.state and a.zip=b.zip
 where a.cid < b.cid`

(b) `select street,city,state,zip, count(*) from client
 group by street,city,state,zip having count(*) > 1`

(c) `with cnts as (select street,city,state,zip,count(*) from client group by 1,2,3)
 select * from cnts where cnt>1`

(d) `with cnts as (select street,city,state,zip,count()
 over (partition by street,city,state,zip) cnt from client)
 select * from cnts where cnt>1`

(e) Other:

9. Find potential client dups, street,city,state,zip are the same for different clients. List potential duplicate client names.

(a) `select a.street,a.city,a.state,a.zip, 'DUP' as lbl from client a
 inner join contact b
 on a.street=b.street and a.city=b.city and a.state=b.state and a.zip=b.zip
 where a.cid < b.cid`

(b) `select street,city,state,zip, count(*) from client group by
 street,city,state,zip having count(*) > 1`

(c) `with cnts as (select a.*, count(*) over (partition by street,city,state,zip) cnt from client a)
 select name from cnts where cnt>1`

(d) `with cnts as (select street,city,state,zip,count(*) over (partition by street,city,state,zip) cnt
 from client) select name from cnts where cnt>1`

(e) Other:

10. Find all presentations attended by Employee 123.

(a) `select pid from pteam where eid=123`

(b) `select a.pid from presentation a inner join pteam b on a.pid=b.pid where b.eid=123`

(c) `select a.pid from presentation a inner join pteam b on a.pid=b.pid inner join employee c
 on b.eid=c.eid where c.eid=123`

(d) `select pid from presentation where eid=123`

(e) Other:

11. Find all presentations *not* attended by Employee 123.
- select pid from presentation where pid not in (select pid from pteam where eid != 123)
 - with stat as (select pid,sum(case when eid=123 then 1 else 0 end) over (partition by pid) cnt from pteam) select pid from stat where cnt > 1
 - select a.pid from presentation a left outer join pteam b on a.pid=b.pid and b.eid=123 where b.pid is null
 - select a.pid from presentation a inner join pteam b on a.pid=b.pid and b.eid=123 where b.pid is null
 - Other:
12. Find all presentations that have both John Doe and Jane Doe.
- select a.pid from presentation a inner join pteam b on a.pid=b.pid inner join employee c on b.eid=c.eid group by a.pid having sum(case when c.name in ('John Doe','Jane Doe') then 1 else 0 end)>1
 - select a.pid from pteam a inner join employee b on a.eid=b.eid where b.name in ('Jane Doe','John Doe')
 - select a.pid from pteam a inner join employee b on a.eid=b.eid group by a.pid having max(case when name='John Doe' then 1 else 0 end)=1 and max(case when fname='Jane Doe' then 1 else 0 end)=1
 - select a.pid from pteam a inner join employee b using(eid) group by a.pid having max(case when name in ('John Doe','Jane Doe') then 1 else 0 end)=1
 - Other:
13. Find clients who have never been visited.
- select cid from presentation group by cid having count(*)=0
 - select a.cid from employee a inner join pteam b on a.eid=b.eid group by a.cid having count(*)=0
 - select a.cid from presentation a left outer join pteam b on a.pid=b.pid group by a.cid having count(*)=0
 - select a.cid from client a left outer join presentation b on a.cid=b.cid where b.cid is null
14. Find employees who have attended maximum number of presentations.
- select eid from pteam group by eid having count(*) > all(select count(*) from pteam group by eid)
 - with cnts as (select eid, sum(1) x, sum(sum(1)) over () y from pteam group by eid) select eid from cnts where x=y
 - with cnts as (select eid, sum(1) x from pteam group by eid), mx as (select max(x) y from cnts) select eid from cnts inner join mx on x > y
 - select eid from pteam group by eid having count(*) in (select count(*) from pteam group by eid)
 - Other:

19. Below query is identical to:

```
select a.*,b.val
from T1 a
left outer join T2 b
on a.key=b.key and a.val!=b.val
```

- (a) with TMP as (
select a.*,b.val from T1 a inner join T2 b on a.key=b.key
where a.val!=b.val)
select a.*,b.val from T1 a left outer join TMP b on a.key=b.key
- (b) select a.*,b.val
from T1 a inner join T2 b
on a.key=b.key and a.val!=b.val
- (c) with TMP as (
select a.*,b.val from T1 a left outer join T2 b on a.key=b.key
where a.val!=b.val)
select a.* from TMP where a.val!=b.val
- (d) All of the above queries are identical.
- (e) None of the queries are identical to the question.

20. When you write:

```
select * from T1 a inner join T2 b on a.tim between b.start and b.end
```

what is the expected performance?

- (a) Hash join, approximately $O(N \log N)$, where N is the number of records in both T1 and T2.
- (b) Inner loop join, approximately $O(N^2)$, where N is the number of records in both tables.
- (c) Sort merge join, approximately $O(N)$, where N is the number of records in both T1 and T2.
- (d) Distributed hash join, approximately $O(N)$ to distribute data, and $O(N \log N)$ after distribution.
- (e) Other:

Scrap Page:

```
employee(eid, name, email, phone, position, mid)
client(cid, name, email, phone, street, city, state, zip)
presentation(pid, cid, timestamp, description)
pteam(pid, eid)
sale(sid, cid, pid, amnt, timestamp)
plog(lid, pid, notes, timestamp)
```