CISC 7510X Final Exam

For the below questions, use the following schema definition.

```
customer(custid,fname,lname,ccn)
driver(driverid,fname,lname,licno,seatcapacity)
trip(tripid,tim,custid,driverid,dist,price,numseats)
pickup(tripid,tim)
dropoff(tripid,tim)
```

It’s a schema for a ride-sharing company. Customers use a phone app to order a trip. They specify number of seats, and agree on a price. A driver (who has a certain capacity in the car) picks up the customer and subsequently drops them off at the destination. The driver could potentially have multiple ongoing trips (upto the driver’s seat capacity). A trip is considered complete after the customer has been picked up and dropped off.

Pick the best answer that fits the question. Not all of the answers may be correct. If none of the answers fit, write your own answer.

1. (5 points) Find name of driver with license ABC-123

   (a) `select fname,lname from customer where licno= ’ABC-123’`

   (b) `select fname,lname from driver where licno= ’ABC-123’`

   (c) `select driverid from driver where licno= ’ABC-123’`

   (d) `select fname,lname from trip where licno= ’ABC-123’`

   (e) Other:

2. (5 points) Who was John Doe’s driver on April 1st, 2018?

   (a)
   ```
   select a.fname,a.lname from customer a inner join trip b
   on a.custid=b.custid inner join driver c on b.driverid=c.driverid
   where c.fname=’John’ and c.lname=’Doe’ and b.tim>=’2018-04-01’
   and b.tim<’2018-04-02’
   ```

   (b)
   ```
   select c.fname,c.lname from customer a inner join trip b
   on a.custid=b.custid inner join driver c on b.driverid=c.driverid
   where a.fname=’John’ and a.lname=’Doe’ and b.tim>=’2018-04-01’
   and b.tim<’2018-04-02’
   ```

   (c)
   ```
   select b.fname,b.lname from customer a inner join trip b
   on a.custid=b.custid inner join driver c on b.driverid=c.driverid
   where c.fname=’John’ and c.lname=’Doe’ and b.tim>=’2018-04-01’
   and b.tim<’2018-04-02’
   ```

   (d)
   ```
   select d.fname,d.lname from customer a inner join trip b
   on a.custid=b.custid inner join driver c on b.driverid=c.driverid
   where a.fname=’John’ and c.lname=’Doe’ and b.tim>=’2018-04-01’
   and b.tim<’2018-04-02’
   ```

   (e) Other:

3. (5 points) What’s the total number of *completed* trips for driver 12345?

(a) `select count(*) from trip where driverid=12345`

(b) `select count(*) from dropoff where driverid=12345`

(c) `select count(*) from trip natural inner join dropoff where driverid=12345`

(d) `select count(*) from driver natural inner join dropoff where driverid=12345`

(e) Other:

4. (5 points) What’s the average price of a trip for customer John Doe?

(a) `select avg(price) from trip where fname=’John’ and lname=’Doe’`

(b) `select avg(price) from customer inner join trip where fname=’John’ and lname=’Doe’`

(c) `select avg(price) from customer inner join trip using (custid) where fname=’John’ and lname=’Doe’`

(d) `select avg(price) from driver natural inner join trip where fname=’John’ and lname=’Doe’`

(e) Other:

5. (5 points) Has John Jackson ever been a passenger of Jack Johnson?

(a) `select * from customer natural inner join trip natural inner join driver where (customer.fname,customer.lname)= (’John’,’Jackson’) and (driver.fname,driver.lname)= (’Jack’,’Johnson’)`

(b) `select count(*)>0 from customer a inner join trip using (custid) inner join driver c using (driverid) where (a.fname,a.lname)= (’John’,’Jackson’) and (c.fname,c.lname)= (’Jack’,’Johnson’)`

(c) `select case when count(*)>0 then true else false end from customer a inner join trip using (custid) inner join driver c using (driverid) where a.fname=c.fname and a.lname=c.lname and a.fname=’John’ and a.lname=’Jackson’ and c.fname=’Jack’ and c.lname=’Johnson’`

(d) `select count(*) from customer a inner join trip using (custid) inner join driver c using (driverid) where (c.fname,c.lname)= (’John’,’Jackson’) and (a.fname,a.lname)= (’Jack’,’Johnson’)`

(e) Other:

6. (5 points) What’s the maximum trip distance for driver Jane Doe?

(a) `select max(dist) from driver inner join trip using(tripid) where lname=’Doe’ and fname=’Jane’`

(b) `select max(dist) from driver inner join trip on(tripid) where lname=’Doe’ and fname=’Jane’`

(c) `select max(dist) from driver inner join trip where driverid=trip.driverid and lname=’Doe’ and fname=’Jane’`

(d) `select max(dist) from driver natural inner join trip where lname=’Doe’ and fname=’Jane’`

(e) Other:

7. (5 points) Find customers who never ordered a trip.

   (a) `select * from customer left outer join trip on tripid where tripid is null`

   (b) `select custid,fname,lname from customer join on cistid=trip.customerid where tripid is null`

   (c) `select custid,fname,lname from customer natural left outer join trip where tripid is null`

   (d) `select custid,fname,lname from customer natural inner outer join trip where tripid is null`

   (e) Other:

8. (5 points) Find drivers who have never *completed* a trip.

   (a) `select a.driverid from driver a left outer join trip b on a.driverid=b.driverid where b.tripid is null`

   (b) `select a.driverid from driver a left outer join trip b on a.driverid=b.driverid left outer join dropoff c on b.tripid=c.tripid group by a.driverid having count(c.tripid)=0`

   (c) `select a.driverid from driver a inner join trip b on a.driverid=b.driverid left outer join dropoff c on b.tripid=c.tripid group by a.driverid having count(c.tripid)=0`

   (d) `select a.driverid from driver a left outer join trip b on a.driverid=b.driverid left outer join dropoff c on b.tripid=c.tripid where c.tripid is null`

   (e) Other:

9. (5 points) Find average speed that drivers drive at. (tip: EPOCH extracts number of seconds in interval)

   (a) `select avg( dist/(extract(epoch from c.tim-b.tim)/ 3600.0)) from trip a inner join pickup b using (tripid) inner join dropoff c using (tripid)`

   (b) `select avg( dist/(extract(epoch from c.tim-b.tim)/ 3600.0)) from trip a left join pickup b using (custid) left join dropoff c using (driverid)`

   (c) `select avg( dist/(extract(epoch from c.tim-b.tim)/ 3600.0)) from trip a inner join pickup b using (tripid) inner join dropoff c using (driverid)`

   (d) `select avg( dist/(c.tim-b.tim)) from trip a inner join pickup b using (tripid) inner join dropoff c using (tripid)`

   (e) Other:

10. (5 points) What’s the average and standard deviation of time between trip order and pickup?

(a)
```
select avg(extract(epoch from b.tim- coalesce(a.tim,0))),stddev(
extract( epoch from b.tim-coalesce(a.tim,0))) from trip a
left outer join pickup b using (tripid)
```

(b)
```
select avg(extract(epoch from b.tim- a.tim)),stddev( extract(
epoch from b.tim-a.tim)) from trip a inner join pickup b
using (tripid)
```

(c)
```
with int as (select extract(epoch from b.tim-a.tim) s from
trip a inner join pickup b using (tripid)) select avg(s),stddev(s)
from pickup natural inner join int
```

(d)
```
with int as (select extract(epoch from b.tim-a.tim) s from
trip a inner join pickup b using (tripid)) select avg(s),stddev(s)
from pickup inner join int using (s)
```

(e) Other:

11. (5 points) Who is currently in driver=12345 car?

(a)
```
select fname,lname from trip a inner join pickup b using(tripid)
left outer join dropoff c using (tripid) inner join driver
d using (driverid) where a.driverid=12345 and c.tripid is
null
```

(b)
```
select d.custid,d.fname,d.lname from trip a inner join pickup
b using(tripid) inner join dropoff c using (tripid) inner
join customer d using (custid) where a.driverid=12345 and
c.tripid is null
```

(c)
```
select d.custid,d.fname,d.lname from trip a left join pickup
b using(tripid) inner join dropoff c using (tripid) inner
join customer d using (custid) where a.driverid=12345 and
b.tripid is null
```

(d)
```
select d.custid,d.fname,d.lname from trip a inner join pickup
b using(tripid) left outer join dropoff c using (tripid)
inner join customer d using (custid) where a.driverid=12345
and c.tripid is null
```

(e) Other:

12. (5 points) Jane Doe is currently a passenger (has been picked up, but not dropped off). Besides the driver, who else is in the car?

(a)
```
select d.custid,d.fname,d.lname from customer a inner join
trip b on (custid) inner join trip c on (driverid) inner
join customer d on (custid) where a.fname=’Jane’ and f.lname=’Doe’
```

(b)
```
with trips as (select a.custid cust1,b.custid cust2 from
trip a inner join trip b using (driverid)) select c.lname,c.fname
from trips a inner join customer b on a.cust1=b.custid inner
join customer c on a.cust2=c.custid where (b.lname,b.fname)=
(’Doe’,’Jane’)
```

(c)
```
with currtrips as (select d.custid,d.fname,d.lname,a.driverid
from trip a inner join pickup b using(tripid) left outer
join dropoff c using (tripid) inner join customer d using
(custid) where c.tripid is null) select b.custid,b.fname,b.lname
from currtrips a inner join currtrips b using(driverid) where
(a.fname,a.lname)= ('Jane','Doe') and a.custid!=b.custid
```

(d)
```
select othercust.fname,othercust.lname from customer jane
inner join customer othercust on a.custid=b.custid where
jane.lname='Doe' and jane.fname='Jane'
```

(e) Other:

13. (5 points) Has Jack Johnson ever shared a ride with John Jackson? (inside the same car at the same time)

(a)
```
with trips as (select a.custid,b.driverid from customer a
inner join trip b using (custid) inner join pickup c using
(tripid) left outer join dropoff d using (tripid) where (lname,fname)=
('Jack','Johnson') or (lname,fname)= ('John','Jackson'))
select count(*)>0 from trips a inner join trips b using (driverid)
where a.custid!=b.custid
```

(b)
```
with trips as (select a.custid,b.driverid,c.tim tstart, coalesce(d.tim,
now()) tend from customer a inner join trip b using (custid)
inner join pickup c using (tripid) left outer join dropoff
d using (tripid) where (lname,fname)= ('Jack','Johnson')
or (lname,fname)= ('John','Jackson')) select count(*)>0 from
trips a inner join trips b using (driverid) where a.custid!=b.custid
and (a.tstart between b.tstart and b.tend or b.tstart between
a.tstart and a.tend)
```

(c)
```
select count(*)>0 from customer a inner join trip b using
(custid) inner join trip c using (driverid) inner join customer
d using (custid) where a.fname='John' a.lname='Jackson' and
d.fname='Jack' and d.lname='Johnson' and extract(epoch from
b.tim-c.tim) = 0
```

(d)
```
select count(*)>0 from customer a inner join trip b using
(custid) inner join pickup c using (tripid) inner join dropoff
d using (tripid) inner join trip b2 on b.driverid=b2.driverid
inner join pickup c2 using (tripid) inner join dropoff d2
using (tripid) inner join customer a2 using (custid) where
a.fname='John' and a.lname='Jackson' and a2.fname='Jack'
and a2.lname='Johnson'
```

(e) Other:

14. (5 points) What percentage of trips cost over $20?

(a)
```
select row_number()/count(*) prcnt from trip where price>20
```

(b)
```
select percentage(price) prcnt from trip where price>20
```

(c)
```
select sum(case when price > 20 then 1.0 else 0.0 end)/sum(1.0)*100.0
prcnt from trip
```

(d)
```
select case when price > 20 then 1.0 else 0.0 end*100.0 prcnt
from trip
```

(e) Other:

15. (5 points) Find outlier trips: those that cost more than 4 stddev above the mean.

(a)
```
with stats as (select a.*, avg(price) over () m, stddev(price)
over () sd from trip a) select * from stats where price>m+sd*4
```

(b)
```
select a.* from (select price, avg(price) m, stddev(price)
sd from trip a group by a.price) a where price>m+sd*4
```

(c)
```
with stats as (select a.driverid, avg(price) over (order
by tim) m, stddev(price) over (order by tim) sd from trip
a group by a.driverid) select * from stats where price>m+sd*4
```

(d)
```
select * from trip where price > 4*mean
```

(e) Other:

16. (5 points) For drivers who completed at least 2000 hours of driving in 2017, what's the average revenue per driver for 2017?

(a)
```
select avg( sum(price) ) over () avgrev from trip having
count(*)>=2000 group by driverid limit 1
```

(b)
```
select driverid,sum(price), avg( sum(price) ) over () avgrev
from trip having count(*)>=2000 group by driverid
```

(c)
```
with stats as (select driverid,sum(a.price) totrev,sum(extract(epoch
from c.tim-b.tim)/3600.0) hours from trip a inner join pickup
b using (tripid) inner join dropoff c using (tripid) where
a.tim>= ’2017-01-01’ and a.tim< ’2018-01-01’ group by driverid)
select avg(totrev) from stats where hours>=2000
```

(d)
```
with trip2017 as (select driverid,tripid,price from trip
where tim>= ’2017-01-01’ and tim< ’2018-01-01’),
tripevents as (select driverid,tim,1 cnt from trip2017 natural
inner join pickup union all select driverid,tim,-1 cnt from
trip2017 natural inner join dropoff),
runtot as (select a.*, sum(cnt) over (partition by driverid
order by tim) rtot from tripevents a),
workint as (select a.*, lag(tim) over (partition by driverid)
lagtim from runtot a where rtot <= 1),
drvs2k as (select driverid from workint where rtot=0 group
by driverid having sum( extract(epoch from tim-lagtim)/3600.0)
>= 2000),
drvrev as (select driverid,sum(price) r from drvs2k natural
inner join trip2017 group by driverid)
select avg(r) from drvrev
```

(e) Other:

17. (5 points) What percentage of all trips have more than one customer in the car?

(a)
```
select sum(case when a.tripid!=b.tripid then 1.0 else 0.0
end)/sum(1.0)*100.0 prcnt from trip a inner join trip b using
(driverid)
```

(b)
```
select sum(case when c.custid is not null then 1.0 else 0.0
end)/sum(1.0)*100.0 prcnt from customer a inner join trip
b on a.custid=b.custid inner join trip c on b.driverid=c.driverid
left outer join customer d on c.custid=d.custid
```

(c)
```
with evt as (select driverid, tripid, pickup.tim,1 cnt from
trip inner join pickup using (tripid) union all select driverid,
tripid, dropoff.tim,-1 cnt from trip inner join dropoff using
(tripid)),
rtots as (select a.*, sum(cnt) over (partition by driverid)
rtot from evt a),
over2 as (select driverid, tripid from rtots where rtot>=2
group by driverid, tripid)
select sum(case when b.tripid is not null then 1.0 else 0.0
end)/sum(1.0)*100.0 prcnt from trip a left outer join over2
b using (tripid)
```

(d)
```
with stats as (select driverid, tripid, count(*) cnt from
trip group by driverid, tripid)
select sum(when cnt>2 then 1.0 else 0.0 end)/sum(1.0)*100.0
prcnt from stats
```

(e) Other:

18. (5 points) The below code (tip: write out the first few output numbers):

```
with recursive n(n) as (
    select 2 n union all
    select n+1 from n where n<1000
)
select a.n
from n a inner join n b on b.n < sqrt(a.n)+1
group by a.n
having a.n=2 or min(a.n % b.n) > 0 order by 1
```

(a) Is invalid

(b) Will generate a list of numbers 1 to 1000

(c) Will create a table with all primes between 1 and 1000

(d) Will produce all prime numbers between 1 and 1000

(e) Other:

19. (5 points) In general, on limited memory system, no indexes, and huge tables, what join type would perform best?

    (a) merge join.

    (b) hash join.

    (c) indexed lookup join.

    (d) inner loop join.

    (e) Other:

20. (5 points) 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 left outer join T2
    b on a.key=b.key where a.val!=b.val)
    select a.* from TMP where a.val!=b.val
    ```

    (b)
    ```
    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
    ```

    (c)
    ```
    select a.*,b.val from T1 a inner join T2 b on a.key=b.key
    and a.val!=b.val
    ```

    (d) All of the above queries are identical.

    (e) None of the queries are identical to the question.