Comp 521 Midterm 2

15 November 2018

Instructions

Don't PANIC!. Focus on getting your answer as close to correct as you can. There will be an opportunity for partial credit on the SQL and equation questions that will depend on how much you have to change to make your answer correct.

We will **not** answer questions about course content, SQL syntax, etc. We will only deal with issues related to the exam implementation.

You should use the latest **Google Chrome** browser to take this exam; it may work in other browsers but I can't help you if it doesn't.

If your browser hangs, for example because of a bad SQL query, simply kill the page and refresh. It should restore all of your work.

You may use **only** two 8.5 by 11 sheets of paper, possibly double sided, as a cheatsheet.

You **must expand your web browser window to cover the full screen** and keep it that way for the duration of the exam. You must **not** access anything besides this page before you submit the exam. I will know if you use other browser pages or programs.

You may **NOT** leave before you submit your exam. When you submit you must enter the code displayed on the screen at the front of the class or given to you by ARS. Only your **first** submission with a correct submit code will be graded.

After you submit you must leave the room. Do not use your phone or computer until after you leave the room.

No submissions will be accepted after the allotted time.

Questions

1. What is the equivalent of this SQL in relational algebra?

Consider the following SQL statement.

```
select distinct a from b, c where d
```

Think of `a`, `b`, `c`, and `d` as variables, tables, or conditions as we would have in a SQL statement.

Identify the relational algebra expression below that is most similar to the SQL.

A. $\pi_a(b \times c)\sigma_d$
B. $\pi_a\sigma_d(b \times c)$
C. $\pi_d\sigma_{b,c}d$

D. $\pi_a \sigma_{b,c} a$
E. $\pi_{b,c} \sigma_a d$

B 6 points

2. What is the point of data striping in RAID configurations?

A. It provides redundancy
B. It supports the recovery from a failed drive
C. It reduces the mean time between failures.
D. It increases the effective disk transfer rate
E. All of these.

D 6 points

3. Why might sequentially scanning a table stored as a clustered B+ tree take about 50% longer than sequentially scanning a table stored in ISAM format?

A. Because the typical occupancy for a B+ tree is only about 2/3.
B. Because we have to follow the links in the B+ tree for every record.
C. Because we have compute the hash for each key.
D. Because successive records may be far apart on disk.

A 6 points

4. How tall is this B+ tree?

If a B+ tree has average fanout of F and contains N nodes, approximately how many levels should we expect?

A. FN
B. $F \log_2 N$
C. $F\sqrt{N}$
D. $\log_F N$

D 6 points

5. Why is an SQL query that uses the distinct keyword likely to be slower than one that doesn't even though it likely produces fewer results?

A. Processing that additional word in the query takes longer.
B. Detecting and eliminating duplicates requires extra computation.
C. Relational algebra does not allow duplicates.
D. Redundancy is bad for performance.

B 6 points

6. An index contains a collection of data entries and supports efficient retrieval of all data entires k* with a given key value k. Given a data entry k*, what is the maximum number of disk I/O operations required to fetch the record from a file with N records?

A. 0
B. 1

C. N
D. $\sqrt{N}$
E. $2N$

B 6 points

7. Why may we only have one clustered index on a relation?

A. It is a limit of relational algebra.
B. We can only store the data in one sorted order without introducing redundancy.
C. It is a limit of SQL.
D. To do otherwise would violate the third law of thermodynamics.

B 6 points

8. Does using an index always make a query faster?

No 3 points

9. How many cylinders does this disk have?

Consider a disk with total disk capacity of **Cdisk** TB (2^{40} bytes), surface capacity of **Csurf** GB (2^{30} bytes), sector size of **Csect** bytes and **St** sectors per track.

How many cylinders are there? Write an equation involving any of these variables needed to calculate the number of cylinders.

Csurf*2**30/(Csect*St) 6 points

10. How long will this external merge sort take?

Make the following assumptions:

- You have 1 MB (2^{20} bytes) of RAM available for buffers.
- Your disk block and internal buffer size is 4 kB (2^{12} bytes).
- Your disk can read or write 256 MB (2^{28} bytes) per second.
- The table to be sorted is T MB where $4 <= T <= 32$.

Write a forumla using the value T that will produce the estimated time in seconds.

Hint: This problem does not require a fancy formula or functions like floor or ceiling; only simple arithmetic is required. If you insist on using them, the avaliable functions include: `Math.log`, `Math.floor`, `Math.ceil`, `Math.sqrt`. But I repeat **You do not need these functions.**

4*T*2**20/(2**28) 9 points

Database Info

```
create table Sailors
    (id integer primary key,
     name text,
     rating integer,
     age integer)
```

```
create table Boats
    (id integer primary key,
     name text,
     color text)

create table Reserves
    (sid integer,
     bid integer,
     day date,
     primary key (sid, day)
     foreign key (sid) references Sailors(id)
     foreign key (bid) references Boats(id))
```

I'll dump the first few rows from each table.

```
select * from Sailors limit 3;
108|Jason Cabrera|7|59
110|Steven Douglas|9|48
112|Brittany Mcclure|1|63

select * from Boats limit 3;
6|Marine|black
9|Clipper|yellow
22|Marine|blue

select * from Reserves limit 3;
179|9|2017-08-17
108|48|2017-06-06
136|9|2017-06-06
```

Database questions

11. How many boats do we have in each color?

For each color, list the color and count of boats of that color. List them in alphabetical order by color.

```
SELECT B.color, COUNT(*)
FROM Boats B
GROUP BY (B.color)
ORDER BY(B.color)
```

Execute 6 points

12. How many reservations did we have for each color?

For each color, list the color and the number of times boats of that color were reserved. List them in decending order by the count then by color in alphabetical order.

```
SELECT B.color, COUNT(*) as c
FROM Reserves R, Boats B
WHERE R.bid = B.id
GROUP BY B.color
ORDER BY c DESC, B.color
```

Execute 8 points

13. How many different sailors reserved each color?

For each color, list the color and the number of different sailors who reserved a boat of that color. List them in decending order by the count then by color in alphabetical order.

```
SELECT B.color, COUNT(DISTINCT S.id) as c
FROM Sailors S, Reserves R, Boats B
WHERE S.id = R.sid AND B.id = R.bid
GROUP BY B.color
ORDER BY c DESC, B.color
```

Execute 8 points

14. Which sailors reserved boats of at least 3 different colors?

List the sailors by name alphabetically.

```
SELECT S.name
FROM Sailors S, Boats B, Reserves R
WHERE S.id = R.sid AND B.id = R.bid
GROUP BY S.id
HAVING COUNT(DISTINCT B.color) >2
ORDER BY S.name
```

Execute 8 points

15. Which pairs of sailors have the same age?

List pairs of sailors who are the same age. List the two names in alphabetical order followed by their age. List the pairs in order of increasing age.

I'm looking for rows like:

```
Allen  Barry  20
```

```
SELECT S1.name, S2.name, S1.age
FROM Sailors S1, Sailors S2
WHERE S1.age = S2.age AND S1.name < S2.name
Order BY S1.age ASC
```

Execute 10 points

Honor Pledge

I certify that no unauthorized assistance has been received or given in the completion of this work. Fill in your full name here: Armaan Sethi

Submit your exam

You must enter the exam submit code displayed on the screen at the front of the room (or given to you by ARS) immediately before submitting. Your submission **will not be graded** if you use an incorrect or old code. The system will warn you if your code is invalid. If that happens simply use the back button on your browser to go back and enter the correct code.

Only your **first submission** with the correct code will be counted. Do not enter the code before you are ready to make your final submission.

Enter submit code: 6376

Submit