Comp 521 Practice Midterm 2

1 November 2018

Instructions

This is for you. Take it just like you would the real exam but it won't count for anything. The exam will be similar to this.

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.

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 in your browser before you submit the exam. I will know if you use other browser pages or programs.

You may **NOT** leave the room until you submit your exam. When you submit on exam day you will include a code just as on the first exam.

No submissions will be accepted after the allotted time.

Questions

1. (5 points) Of the following disk timing characteristics, which accounts for moving the head?

[Choose an answer ⬍] 5 points

2. (5 points) Which file organization offers the best average performance for database inserts?

[Choose an answer ⬍] 5 points

3. (2 points each) Which of these statements are True about why a DBMS might sort data?

- ☐ As the first step in bulk-loading a B+ tree.
- ☐ To eliminate duplicate copies.
- ☐ When the user requests the records in some order.
- ☐ When implementing a join operation.

4. (6 points) What is the most likely interpretation of the following SQL statement's execution?

```
DELETE FROM Professor
WHERE Id IN (
    SELECT T.ProfID
        FROM Teaching T, Course C
        WHERE T.Course = C.Name
        AND      T.Semester = C.Semester
        AND      C.FailRate > 0.25)
```

Choose an answer ⬦ 6 points

    a. All Courses with a failure rates greater than 25% are removed from the catalog.
    b. All Professors who have taught courses with failure rates greater than 25% are deleted.
    c. list of Professors who have taught courses with failure rates greater than 25% is generated.
    d. The statement has no impact.
    e. All Professors are removed from the database.

5. (5 points) Of the following disk-transfer timing characteristics, which is likely to be greatest when reading a random block?

Choose an answer ⬦ 5 points

6. (3 points each) How big is the natural join?

If relation R_1 has degree D_1 and cardinality C_1 and relation R_2 has degree D_2 and cardinality C_2, what is the degree D_j and cardinality C_j of relational algebra expression $R_1 \bowtie R_2$? In each case choose the answer that best describes the result.

Choose an answer ⬦ 3 points

    a. $C_1 < C_j < C_2$
    b. $C_j = C_1 + C_2$
    c. $min(C_1, C_2) < C_j < max(C_1, C_2)$
    d. $0 \leq C_j \leq C_1 C_2$

Choose an answer ⬦ 6 points

    A. $D_1 < D_j < D_2$
    B. $D_1 + D_2 < D_j \leq D_1 D_2$
    C. $max(D_1, D_2) \leq D_j \leq D_1 + D_2$
    D. $min(D_1, D_2) \leq D_j \leq D_1 D_2$

7. (4 points each) How many disk IO operation should we expect to sort a file with the given specifications using an external merge sort?

I don't like this question.

The file has **Nfile** records of 250 bytes each. Each disk block holds **Cblock** bytes and no record spans two blocks. We have **Nbuffer** memory buffers, each capable of holding one disk block.

The numeric values of the variables Nfile, Cblock, and Nbuffer are defined for you. You must write a formula using these names to produce the answer. I will check your answer with multiple different values for the variables. The field will turn green when your answer is correct.

You can use the usual mathematical operators and these functions: Math.ceil, Math.exp, Math.floor, Math.log. Note: $log_B(x) = log(x)/log(B)$

I'll guide you through the calculation with credit for each step.

$$NumberOfPasses = 1 + \lceil log_{B-1} \lceil N/B \rceil \rceil$$

First, determine the number of records that fit in a block. `Enter an expression` 4 points

Now assuming the previous value is available as **RpB** determine the number of blocks in the file.
`Enter an expression` 4 points

Determine the number of passes assuming the previous value is avalable as **Nblocks**.
`2`
4 points

Finally determine the number of IO operations assuming the previous value is available as **Npasses**.
`2*Npasses*Nblocks` 4 points

8. (5 points) What is the product of sums form of the following where condition?

```
where reviewed = 1 or (aid < 100 and title like '%fun%')
```

```
A. (reviewed = 1 and aid < 100) and (reviewed = 1 and title like '%fun')
B. (reviewed = 1 and aid < 100) or  (reviewed = 1 and title like '%fun')
C. (reviewed = 1 or  aid < 100) and (reviewed = 1 or  title like '%fun')
D. (reviewed = 1 or  aid < 100) or  (reviewed = 1 or  title like '%fun')
```

`Choose an answer` ⬍ 5 points

9. (5 points) What is the rotation rate of a disk that has average rotational delay of Tar?

Express your answer in revolutions per minute; Tar is in seconds. `Enter an expression` 5 points

Database info

You are to write a single SQL query to answer each question. You will not write any Python code below.

The schema of the database is below. It represents book reading activity on Tar Heel Reader over a short period.

The **ip** address is assumed to correspond to a single location. For reasons you'll learn in Comp 431 they sometimes don't but for the purposes of the exam assume they do.

```
create table Reads ( -- a record for each book that was read
    time date,       -- time the book was read
    ip integer,      -- ip of the computer
    bid integer,     -- book id
    foreign key (bid) references Books,
    foreign key (ip) references Locations)

create table Books (
    bid integer primary key,
    title text,
    aid integer,         -- author id
    reviewed integer,    -- 1 if the book is reviewed, 0 otherwise
    foreign key (aid) references Authors)

create table Locations (
    ip integer primary key, -- ip address of the computer
    country text)           -- country determined from the ip address
```

```
create table Authors (
    aid integer primary key,
    login text,    -- user id like Gary
    birthday text) -- month and day of the author's birthday like May 29
```

I'll dump the first few rows of each table so you can get a better idea of the contents.

Reads

```
(time, ip, bid)

('2017-10-01 00:06:48', 3023614022, 124363)
('2017-10-01 00:07:35', 3023614022, 124884)
('2017-10-01 00:07:53', 3023614022, 124366)
```

Books

```
(bid, title, aid, reviewed)

(148, 'Lullaby', 9, 1)
(171, 'Balloons Everywhere!', 14, 1)
(202, 'Baa Baa Black Sheep', 61, 1)
```

Authors

```
(aid, login, birthday)

(1, 'Gary', 'May 29')
(2, 'DLM', 'February 24')
(5, 'Jenny', 'January 18')
```

Locations

```
(ip, country)

(32591211, 'United States')
(40693121, 'United States')
(60836422, 'Canada')
```

More questions

10. (6 points) How many different books where read by each country where more than 1 book was read?

I'm expecting tuples with the country name followed by the count. Sort your results first by number of books in descending order and then by country in ascending order.

```
SELECT A.login, COUNT(DISTINCT L.country)
FROM Locations L, Reads R, Books B, Authors A
WHERE L.ip = R.ip AND B.aid = A.aid AND R.bid = B.bid
GROUP BY A.login
HAVING COUNT(DISTINCT L.country) > 3
ORDER BY COUNT(DISTINCT L.country) DESC
```

Execute   6 points

11. (10 points) How many different countries read books by each author?

I want the author's login and the number of different countries in which their book was read. Sort your results first by the number of countries in descending order and then by the login in ascending order. Limit your list to those authors whose books were read in at least 4 different countries.

```
```
Execute   10 points

12. (10 points) What fraction of the reads were books that have been reviewed?

I'm looking for a single floating point number that reflects the portion of the read events referred to a reviewed book. For example if there were only 2 reads total and one was reviewed and the other wasn't I'd expect the answer to be 0.5. Remember the reviewed field is 1 if the book has been reviewed and 0 otherwise.

```
SELECT AVG(B.reviewed)
FROM Reads R, Books B
WHERE B.bid = R.bid
```
Execute   10 points

13. (10 points) Which reviewed books have the same title as another different reviewed book?

I'm looking for the titles of the reviewed books that share a title with at least one other different book that is also reviewed. Report only the titles. Sort your results first by the duplicate count in descending order and then by the title in ascending order.

```
SELECT b1.title
FROM Books b1, Books b2
WHERE b1.title = b2.title AND b1.bid != b2.bid AND b1.reviewed = 1 AND b2.reviewed = 1
GROUP BY (b1.title)
ORDER BY COUNT(*) DESC, b1.title
```
Execute   10 points

Submit