

# Comp 521 Final Exam

8 December 2018

## Instructions

### Don't PANIC!

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** four 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.**

## Database Schema

Here are the tables you'll find in the database.

```
create table if not exists Books (  
    id integer primary key,  
    slug text,  
    title text,  
    author integer references Authors(id),  
    pseudonym text,  
    audience text, -- C|E|U C==Caution E==Everybody U==Not rated  
    reviewed integer, -- 1 == reviewed, 0 == not  
    rating_avg float,  
    rating_count integer,  
    cover integer references Pictures(id),  
    language text, -- iso-code  
    created timestamp,  
    modified timestamp  
)
```

```
create table if not exists Authors (  
    id integer primary key,  
    login text,  
    password text,  
    lastname text,  
    firstname text,  
    email text  
)  
  
create table if not exists Pictures (  
    id integer primary key,  
    url text,  
    width integer,  
    height integer  
)  
  
create table if not exists Pages (  
    book integer references Books(id),  
    page integer,  
    caption text,  
    picture integer references Pictures(id)  
)  
  
create table if not exists Words (  
    word text,  
    book integer references Books(id),  
    count integer -- how many times the word occurs in the book  
)
```

## Section 1

In this section there are questions worth 80 points total. The maximum score for this section is 60 points. Excess points will **not** carry over to section 2.

You will write SQL queries. They will be tested against two databases. The small db has about 0.1% of the books in Tar Heel Reader, the larger db has about 1%.

I'll scatter several of these SQL input fields through the exam for you to use for experiments and scratch space. They will **NOT** contribute in any way to your grade.

### Scratch space

Execute

#### 1. How many reviewed, English-language books are in the db?

The iso-code for English is 'en'. Your query should produce an integer.

Your result should resemble this:

count
-------

count
24

```
SELECT COUNT(B.id)
FROM Books B
WHERE B.reviewed == 1 AND B.language == 'en'
```

Execute 4 points

## 2. Who are the most prolific authors?

List the Author's last name, first name and the number of books they wrote for all authors who have written **more than 2 books**. Order the results first by the number of books descending, then by the last name in ascending order and finally by the first name in ascending order.

Your result should resemble this:

lastname	firstname	C
Fenne	Emmaline	3

```
SELECT A.lastname, A.firstname, COUNT(B.id) as ct
FROM Authors A, Books B
WHERE A.id = B.author
GROUP BY A.id
HAVING ct > 2
ORDER BY ct DESC , A.lastname, A.firstname
```

Execute 8 points

## Scratch space

Execute

## 3. Which reviewed English-language books use a repeated refrain?

List the book id, title and author last name for books that have the same caption text on at least 3 different pages. List the books in alphabetical order first by book title and then by authors last name. The iso-code for English is 'en'.

Your result should resemble this:

id	title	lastname
74335	Devin, What Are You Doing?	Depper
128828	Where's Elmo?	Fenne

```
SELECT DISTINCT B.id, B.title, A.lastname
FROM Books B, Authors A, Pages P
WHERE A.id = B.author AND P.book = B.id AND B.reviewed = 1 and B.language = 'en'
```

```
GROUP BY P.caption, B.id
HAVING COUNT(DISTINCT P.page) > 2
ORDER BY B.title, A.lastname
```

Execute 8 points

#### 4. Which books use the word "new" at least 2 times?

Your query should return the book id, book title, and number of pages in the book. I expect you to use the Words table here, **not** a *like* query on the caption.

Return them in ascending order by the book id.

Your result should resemble this:

id	title	pages
74345	ocean city new jersey	4

```
SELECT B.id, B.title, COUNT(P.page) as ct
FROM Books B, Pages P, Words W
WHERE P.book = B.id AND W.book = B.id AND W.word == "new" AND W.count > 1
GROUP BY B.id
ORDER BY B.id ASC
```

Execute 6 points

#### Scratch space

Execute

#### 5. Which books include both the words "see" and "you"?

I'm talking about words from the Words table not a *like* query.

List the book title, pseudonym and author login for each book. List each book only once. Order the results in alphabetical order by title and then by book id.

Your result should resemble this:

title	pseudonym	login
Rudimenta Grammaticae 2 - Basics of Grammar 2	Evan Millner	kborthramcv
What Do You See?	73imagine	wickovitzcz
Yukon, Yukon, What Do You See?	Alison	ghamleyd8

```
WITH youtable as (
SELECT B.id
FROM Books B, Words W
WHERE W.book = B.id and W.word == "you"
), seetable as (
SELECT B.id
FROM Books B, Words W
```

```
WHERE W.book = B.id and W.word == "see"
)
SELECT B.title, B.pseudonym, A.login
FROM youtable, seetable, Books B, Authors A
WHERE youtable.id = seetable.id AND B.id = youtable.id AND B.author = A.id
ORDER BY B.title, B.id
```

Execute 8 points

## 6. Which books have missing pages?

Imagine that we have discovered that a program bug has caused some pages to be deleted from books. The pages of every book should be numbered sequentially. Consider a book with pages numbered 1, 2, 4. This book appears to be missing page 3.

List the title and book id for each book that appears to have a page missing. List them in order by id.

Of course, if the last page was missing you couldn't detect it, don't worry about that case.

Example result:

title	id
Nick, Marielaina and Aimen's Book About Animals	18502
I Like...	38168
Para mantener la salud	62976
Soup Is The Best!	173034

```
SELECT B.title , B.id
FROM Books B, Pages P
WHERE P.book = B.id
GROUP BY B.id
HAVING COUNT(P.page) != MAX(P.page)
ORDER BY B.id ASC
```

Execute 8 points

## Scratch space

Execute

## 7. Which pairs of books have at least 3 different pictures in common?

Your query should produce pairs of book titles. List each pair of books only once. List each pair with the lower numbered book id first. List the pairs in ascending order by their book ids.

Your result should resemble this:

title	title
Max the Cat	This Cat Is...

```

SELECT B1.title, B2.title
FROM Books B1, Books B2, Pictures I1, Pictures I2, Pages P1, Pages P2
WHERE B1.id != B2.id AND B1.id = P1.book AND B1.id = P1.book AND B2.id = P2.book
AND P1.picture = I1.id AND P2.picture = I2.id AND I1.id = I2.id AND B1.id < B2.id
GROUP BY B1.id, B2.id
HAVING COUNT(DISTINCT I1.id) >2
ORDER BY B1.id ASC, B2.id ASC

```

Execute 8 points

## 8. Which books, longer than 10 pages, were written by authors from universities?

Assume an author is from a university if their email address ends with '.edu'.

Only include books with more than 10 pages.

For each book, your query should return 3 values: book title, author lastname, author email. Your results should be ordered by increasing book id.

Your result should resemble this:

title	lastname	email
The Guatamalan Adventure!	Guittet	vguittetd4@msu.edu
Gatsby and Daisy Drink Tea	Humphrey	hhumphreyd5@umn.edu
Melly and I are going to Jamaica	Heinl	nheinldw@uiuc.edu

```

SELECT B.title, A.lastname, A.email
FROM Books B, Authors A, Pages P
WHERE A.id = B.author AND P.book = B.id and A.email LIKE "%.edu"
GROUP BY B.id
HAVING COUNT(P.page) > 10

```

Execute 8 points

## Scratch space

Execute

## 9. Which books have the most words per page?

List the book title, author last name, word count, page count, and average number of words per page. Order your results by average words per page in descending order and author last name in ascending order.

Limit your answer to the top 5.

**Hints:** Use the Words table to get the number of words in a book. Remember you can do simple arithmetic in a select, so get the number of words and the number of pages and simply divide to get the number of words per page. Use integer division so we don't get into precision problems.

Your result should resemble this:

title	lastname	words	pages	wordsperpage
Animal Dads	Apps	66	6	11
What Do You See?	Ickovitz	112	11	10
Yukon, Yukon, What Do You See?	Hamley	98	10	9
Orientation and mobility	Upjohn	77	9	8
Who's in Love with Arthur? Chapter 4	Dome	47	6	7

Execute 10 points

### 10. What SQL query corresponds to this relational algebra expression?

$\pi_{pseudonym} \sigma_{rating\_count > 200} Books$

Write the SQL query that produces the same result.

Your result should resemble this:

pseudonym
Melba Jane Fitzwater

```
SELECT B.pseudonym
FROM Books B
WHERE B.rating_count > 200
```

Execute 4 points

### 11. What is the average rotational delay of a disk that spins at $R$ RPM?

Write a single arithmetic expression. Express your result in seconds. RPM means Revolutions Per Minute. Your answer should involve only constants, the variable  $R$ , and operators such as  $+$ ,  $-$ ,  $*$ , and  $/$ .

Enter an expression 4 points

### 12. How long will this external merge sort take?

Make the following assumptions:

- You have  $2^{20}$  bytes of RAM available for buffers.
- Your disk block and internal buffer size is  $2^{13}$  bytes.
- Your disk can read or write  $2^{29}$  bytes per second.
- The table to be sorted is  $T$  MB where  $4 \leq T \leq 32$ .
- One MB =  $2^{20}$  bytes.

Write a formula using the variable  $T$  that will produce the estimated time in seconds.

4 points

**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 available functions include: `Math.log`, `Math.floor`, `Math.ceil`, `Math.sqrt`. But I repeat **You do not need these functions**. Recall that the exponentiation operator (which you do not need) is `**`.

## Section 2

There are questions totaling 40 points in this section. Points from section 1 do **not** carryover into this section.

13. Of the following, which best characterizes the objective of database table normalization?

- A) Evaluating queries more efficiently
- B) Efficient testing of functional dependencies
- C) Creating more independent relations
- D) Minimizing redundancy

 4 points

14. What is the primary technique used to normalize tables?

- A) Vertical decomposition
- B) Horizontal decomposition
- C) Table Merging
- D) SQL Assertions

 4 points

15. How does concurrency improve database performance?

- A) By making transactions durable
- B) By making transactions atomic
- C) By isolating transactions
- D) By allowing aborts of unnecessary transactions
- E) By scheduling CPU and disk resources for maximum utilization

 4 points

16. When a transaction is aborted, which of the following may be performed by the DBMS?

- A) Changes to pages are undone
- B) Its locks are freed
- C) It may be rescheduled
- D) Other transactions may be aborted
- E) All of the above

 4 points

17. Which of the following database properties states that, once committed, data should persist in the database, even if the system crashes?



- A) Atomicity
- B) Buffer Writes
- C) Consistency
- D) Durability
- E) Equality

D 4 points

18. Which of the following database enhancements created the possibility of deadlock?

- A) Error recovery
- B) Locks
- C) Aborting transactions
- D) Committing transactions
- E) All of the above

B 4 points

19. Which of the following are not consequences of database redundancy?

- A) Redundant storage
- B) Update anomalies
- C) Insertion anomalies
- D) Deletion anomalies
- E) Lossless-Join Decompositions

B 4 points

20. Which is not a candidate key based on this instance?

Consider the following valid instance of a relation:

sid	name	userid	age	gpa
50000	Dave	dave@cs	19	3.3
53666	Jones	jones@cs	18	3.4
53688	Smith	smith@ee	18	3.2
53650	Smith	smith@math	19	3.8
53831	Madayan	mada@music	19	1.8
53832	Guldu	guldu@music	22	2.0

Which of the following can you infer is not a candidate key based on this instance?

- A) sid
- B) name
- C) userid
- D) name, age

B 4 points

21. What is the result of this change to Strict Two-Phase locking?

Suppose that the Strict Two-Phase locking scheme is modified to use only exclusive locks, which are requested before both reads and writes and released on commit.

What is the result?

- A) Concurrent transactions always deadlock.
- B) The system functions correctly, but with lower performance.
- C) The resulting schedules are not serialized.
- D) It behaves identically to Strict Two-Phase locking.

B 4 points

## 22. How can we prevent users from creating multiple logins with the same email address?

How could we change the DB schema used in section 1 to prevent a user from signing up for multiple logins with the same email address?

- A) It can't be done.
- B) Add the `unique` constraint to the login field in the Authors table.
- C) Add the `unique` constraint to the email field in the Authors table.
- D) Add the `unique` constraint to the password field in the Authors table.

C 4 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:

## 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: