

[Home](#)[Grades](#)[View submissions](#)[View feedback](#)[Logout](#)

jmajikes

Comp 521 Spring 2021 Midterm Version A

Instructions

We *strongly* recommend you read the whole exam and begin with questions you know how to solve quickly. Some questions will be harder or take longer than others; don't spend all your time on one question worth only a few points! Also, don't worry if you don't have time to solve every question - that's expected.

There are 100 points total, divided into two sections. The first section contains multiple choice / short response questions and is worth 55 points; the second contains SQL questions and is worth 45 points.

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

We've tested this most in **Google Chrome** so we recommend using that. That being said, it should work fine in any reasonably modern browser (I won't make any promises about Internet Explorer though).

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.

Tar Heel Reader Database Schema

Here are the tables you'll find for the database used primarily in Section 2 but also referenced in some questions in Section 1. If you want to reference this without scrolling, I'd recommend copying the schema into a text editor (e.g. notepad) rather than opening the exam in another tab (this could cause issues with inconsistencies in saved answers).

Some important notes:

- Primary key and foreign key reference constraints are specified as part of the column declaration.
- Comments on some columns are included (the -- indicates a column comment).
- You can assume that no two pages from the same book will be given the same page number and that the highest page number

associated with a book can be understood as the total number of pages in that book.

- The iso-code for English is 'en'

In case you are interested, these table definitions happen to come from the implementation of Prof. Gary Bishop's Tar Heel Reader project.

```
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 L  
    reviewed integer, -- 1 == reviewed, 0 == not  
    rating_avg float, -- average rating  
    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)
```

```
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 i
)
```

Section 1 (55 points)

1. 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 $/$.

4 points

2. Which of these is most related to the idea of data independence?

- A) Using a relational model
- B) Specification of views
- C) Efficient access to data
- D) Crash recovery and concurrency control
- E) All of the above

 4 points

3. ER Modeling

Suppose you are constructing an ER model to capture information about doctors, patients, and office visits. Each patient has a primary doctor. An office visit represents a particular patient seeing a particular doctor which may or may not be their primary doctor on a particular day for a particular reason.

Which of these is most likely true of an ER model that captures this situation?

- A) Office visits are best represented as an entity
- B) A key constraint exists between doctors and patients
- C) A patient's primary doctor is an attribute of a patient
- D) Office visits create a many-to-many relationship between doctors and patients
- E) A and D
- F) C and D
- G) A, C, and D
- H) B, C, and D
- I) All of A, B, C, and D
- J) None of A, B, C, and D

▼ 4 points

4. In the relational model, which of these is true of primary and candidate keys:

- A) A candidate key must be a primary key
- B) A foreign key in one relation is a candidate key in another relation
- C) Primary keys are always specified as the first fields of a relation
- D) A set of two or more fields may be used as a primary key.
- E) A and B
- F) B and C
- G) B and D
- H) C and D
- I) B, C, and D
- J) All of A, B, C, and D

▼ 4 points

5. 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 fields can you infer are not candidate keys based on this instance?

- A) sid only
- B) name only
- C) userid only
- D) age only
- E) gpa only
- F) name and age
- G) name and gpa
- H) name, age, and gpa

☐ 4 points

Note: The next 3 questions reference the Schema at the top of the exam.

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

How could we change the DB schema used in the Tar Heel Reader to prevent an author 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.

▼ 4 points

7. Which of the statements below are true of the following query?

```
SELECT A.id, A.lastname, A.firstname, B.title  
FROM Authors A, Books B, (SELECT max(page), book FROM Pages)  
WHERE B.id = C.book and A.id = B.author
```

A) An author that is included in the result is included at most once.

B) The nested query is correlated with the outer query

C) The outer query relies on an aggregate function

D) A and B

E) B and C

F) A and C

G) All of A, B, and C

H) None of A, B, and C

▼ 4 points

8. JOIN Behavior

Which of these statements using the schema above will produce a result relation that is guaranteed to have the same cardinality as the Pages relation?

A) SELECT book, url from Pages JOIN Pictures

B) SELECT book, url from Pages NATURAL JOIN Pictures

C) SELECT book, url from Pages LEFT JOIN Pictures ON Pages.picture = Pictures.id

D) SELECT book, url from Pages RIGHT JOIN Pictures ON Pages.picture = Pictures.id

E) SELECT book, url from Pages FULL OUTER JOIN Pictures ON Pages.picture = Pictures.id

F) None of the above

✓ 4 points

9. The result relations of two SELECT statements are union-compatible if...

- A) They have the same number of fields
- B) The fields of one relation have the same names as the fields of the other relation.
- C) The i th field of one relation has the same domain constraint as the i th field of the other.
- D) The two relations have the same number of rows.
- E) A and B
- F) A and C
- G) B and C
- H) A, B, and C
- I) A, B, C, and D

✓ 4 points

10. I/O Cost Modeling

Note: the values used below are not realistic nor are they meant to be.

Suppose we have R records stored in a heap file. Each record is 128 bytes long (including its record id). Records are packed into 1024 byte disk pages. The average time to read or write a disk page is 5 ms. The file is indexed by an unclustered tree index using a search key that is 30 bytes long. Non-leaf index pages have a fan out of 100. Leaf index pages have an average occupancy of 50%. Each data entry on the index leaf pages is 20 bytes long.

What is the I/O cost in milliseconds of an equality search? Write the answer as an expression in terms of R . You may assume the function $\log(a,b)$ is available and computes the b -base log of a .

Enter an expression 8 points

11. Why are the leaves of a tree index maintained as a doubly linked list (assume unique search keys)?

- A) To facilitate insertions.
- B) To facilitate deletions.
- C) To facilitate equality searches
- D) To facilitate range searches
- E) To facilitate scanning all records
- F) C, D, and E
- G) All of A through E

4 points

12. Extendible Hashing

Suppose we have an initially empty extendible hash with an initial global depth of 2 (i.e., the last two bits of a hash value are used to determine which of 4 buckets to hash to). The directory has four initial entries corresponding to buckets A, B, C, and D in that order. Each bucket can contain a maximum of 3 entries before it needs to be split. When a bucket splits, assume that the original bucket keeps its name and a new bucket is created and given the next available letter as a name (for example, the first bucket created after the initial buckets will have name E).

The following hash values are inserted in the following order:

0, 3, 4, 7, 11, 1, 15, 8, 12

What is the global depth after these insertions? 1 point

What is the local depth associated with bucket A? 1 point

What is the local depth associated with bucket B? 1 point

What is the local depth associated with bucket C? 1 point

What is the local level associated with bucket D? 1 point

What does the directory look like after these insertions? Please list the bucket names present in the directory in the order they appear. For example, if no insertions were performed the expected answer would be 'ABCD' (without the quotes):

2 points

Section 2 (45 points)

In this section, you will write SQL queries for the Tar Heel Reader schema described at the beginning of the exam. They will be tested immediately against two databases. The small db has about 0.1% of the books in Tar Heel Reader, the larger db has about 1%. Your queries are immediately tested against both and if your queries output matches the expected output, the displayed answers will be outlined in green. Your actual score will be determined when your query is tested against a different database but green feedback should mean that you are on track to receive full credit.

The following scratch space can be used to help develop and test queries against these two databases if you want. Nothing about the scratch space contributes to any points on the midterm and anything in the scratch space is ignored.

Scratch space

Execute 0 points

13. List the titles of all books in English with an average rating between 2.0 and 2.5 inclusive.

Your result should resemble this (the correct answer will have more rows than what is shown here):

title
Pancakes
Shapes
Basetball

Execute 4 points

14. How many books (in any language) have a title that contains the word "animal"

Don't worry about case or plurality, you are just looking for titles with the sequence "animal" in it)?

count
3

Execute 4 points

15. 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

Execute 5 points

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

Your query should return the book id and title. Each book should only be included once in the result.

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	

Execute 6 points

17. 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

Execute 8 points

18. 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

Execute 8 points

19. 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 (the / operator) so we don't get into precision problems.

Your result should resemble this:

title	lastname	words	pages	wordspage
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

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