

CS 1555 Practice Second Exam

Name: _____

Pitt email address: _____@pitt.edu

Peoplesoft number: _____

Score: _____ / 53

- The use of any notes or electronic devices will be considered cheating.
- You must turn in ALL pages of the exam for your exam to be graded. Failure to turn in all pages of this exam booklet will result in a 0 on the exam.
- Illegible answers will receive 0 points.

1. (10 points) Carefully read each prompt and select the best answer for each question.
- (a) Which of the following read anomalies are possible in the “Read uncommitted” ANSI isolation level?
- A** Phantom reads
 - B** Dirty reads
 - C** Non-repeatable reads
 - D** All of the above
- (b) Which if the following describes heap files?
- A** The tuple with the min/max value of some attribute is stored at the start of the first data block, but blocks are not totally sorted according to that attribute
 - B** They contain tuples placed in a arbitrary order
 - C** They contain tuples totally sorted according to some attribute with the min/max value easily accessible at the start of the first data block
 - D** None of the above
- (c) Column store databases are especially useful for...
- A** Analytical processing on stored data
 - B** Speeding up recovery operations
 - C** Speeding up query processing
 - D** Avoiding deadlocks
- (d) Which of the following stated relational algebra equivalences is **not** true?
- A** $\sigma_{a1=1}(\sigma_{a2=2}(R)) = \sigma_{a2=2}(\sigma_{a1=1}(R))$
 - B** $\pi_{a1,a3,a5}(\sigma_{a3=3}(R)) = \sigma_{a3=3}(\pi_{a1,a3,a5}(R))$
 - C** $\pi_{a3}(\pi_{a2,a3}(\pi_{a1,a2,a3,a5}(R))) = \pi_{a3}(R)$
 - D** $\sigma_{a1=1 \vee a2=2}(R) = \sigma_{a1=1}(\sigma_{a2=2}(R))$

2. (4 points) Define the blocking factor of a file.

3. (6 points) List the three phases of ARIES in the order they are performed.

4. (9 points) Consider the following order 3 B+tree.

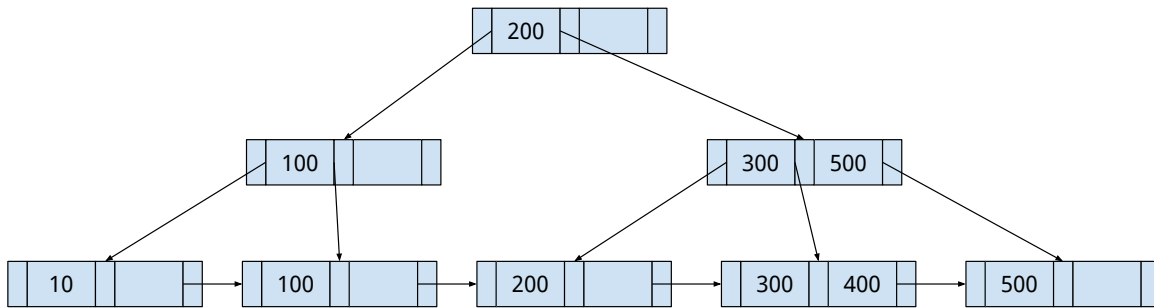


Figure 1: An example B+tree

(a) Draw the tree that will result from inserting 350 into the B+tree shown in Figure 1.

(b) Draw the tree that will result from removing 200 from the B+tree shown in Figure 1.

(c) Draw the tree that will result from removing 10 from the B+tree shown in Figure 1.

5. (10 points) Consider the schema shown in Figure 5 and the following query.

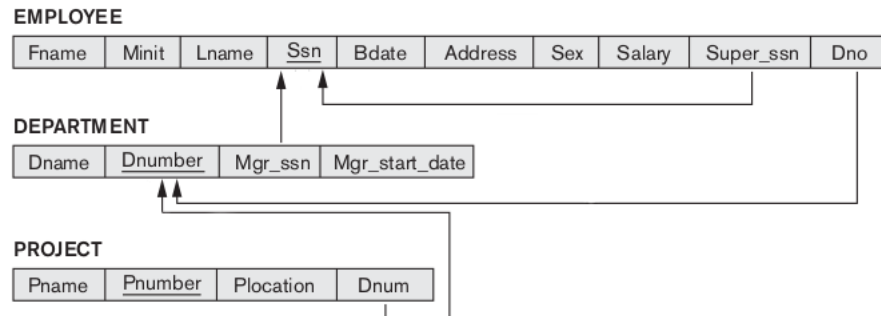


Figure 2: An example schema

```

SELECT E.Lname, D.Dname, P.Pname, P.Plocation
FROM PROJECT P, DEPARTMENT D, EMPLOYEE E
WHERE P.Dnum = D.Dnumber AND D.Mgr_ssn = E.Ssn AND
      E.Ssn = '111-11-1111';
  
```

- (a) Produce a naive query plan for evaluating this query as discussed in lecture.

- (b) Show the result of applying the first two steps of the top-down optimization algorithm described in lecture (“Break up and push down selects”, and “Convert cross products to joins”).

- (c) List the remaining steps that you would have to take to complete optimizing this query plan according to the top-down algorithm described in lecture.

6. (8 points) Name and describe the two phases of 2-phase locking.

State the goal of 2-phase locking.

Describe how it achieves this goal.

7. (8 points) Assume that you are designing a data management solution for the manufacturing company ManuCo. You have decided to create a table storing a tuple about each employee with the following fields:

- An employee ID number
- The employee's salary
- The employee's department ID

ManuCo has specified that your solution must support the following operations:

- Print out the ID and salary of every employee (ordered by ID and performed yearly).
- Compute the average salary for each dept (performed monthly).

Describe the physical layout of this table that you think would lead to the best performance. Be sure to mention any file organizations, block layouts, and indices that you think would be beneficial. State any assumptions you make and be sure to justify your response.