

EEL 4852 Database Systems

Fall 2008

Final Exam

December 11, 2008

Time: 120 minutes

Maximum points: 90

Your Name: \_\_\_\_\_

USF ID:   U  \_\_\_\_\_

**Problem I.** True or false (9 pts, 1 pt each). Mark your choice clearly - ambiguous answers will get zero points. You do NOT need to explain your answer.

1. SQL is an implementation of the domain relational calculus language.  
T ( ) F ( X )  
-----

2. The result of an algebraic operation is always a relation.  
T ( X ) F ( )  
-----

3. Relational algebra is a procedural language.  
T ( X ) F ( )  
-----

4. In SQL, duplicate elimination is enforced by default.  
T ( ) F ( X )  
-----

5. In doing natural joins between table A and table B, if we have an index on the joining attribute(s) of A, we should use A as the outer table in the nested-loop join algorithms.  
T ( ) F ( X )  
-----

6. Given a relational instance (i.e., a real table with many tuples in it), there can be more than one unclustered index.  
T ( X ) F ( )  
-----

7. In practice, the goal of the query optimizer is to find a good query plan instead of THE optimal query plan.  
T ( X ) F ( )  
-----

8. A hash index on a selection attribute is helpful in processing selections with equality conditions (e.g., ... WHERE salary = 50K) and range conditions (e.g., ... WHERE salary > 50K).  
T ( ) F ( X )  
-----

9. Jim Gray won a Turing Award for his contributions to database systems research.  
T ( X ) F ( )

**Problem II. Query languages (21pts, 3 pts each)**

Consider the following schemas of a company library database,

*employee* (*empno*, *name*, *office*, *age*)  
*books* (*isbn*, *title*, *authors*, *publisher*)  
*loan* (*empno*, *isbn*, *date*)

where the underlined attributes are primary keys.

**Write the following queries in SQL:**

1. Print the names of employees who have borrowed any book published by McGraw-Hill.

```
select name  
from employee e, books b, loan l  
where e.empno = l.empno  
and l.isbn = b.isbn and  
b.publisher = 'McGrawHill'
```

2. How many copies of the book titled "Casablanca Express" are checked out?

```
select count(empno)  
from books b, loan l  
where l.isbn = b.isbn and  
b.title = 'Casablanca Express'
```

3. Print the names of employees who have borrowed more than five books

```
select e.name  
from employee e, loan l  
where e.empno = l.empno  
group by e.empno, e.name  
having count(isbn) > 5;
```

4. Print the names of employees who have borrowed ALL books authored by Hemingway.

```
select distinct e.name
from employee e
where not exists
((select isbn
from books
where author = 'Hemingway')
except
(select isbn
from loan l
where l.empno = e.empno))
```

**Write the following queries in relational algebra:**

I use '∞' as the join symbol.

5. Print the names of employees who checked out a book titled "Casablanca Express"

$$\prod_{\text{name}} [\sigma_{\text{title} = \text{'Casablanca Express'}} (\text{employee} \infty \text{loan} \infty \text{books})]$$

6. Print the names of employees who are more than 35 years old but have NOT checked out a book authored by Hemingway.

$$A \leftarrow \prod_{\text{name}} [\sigma_{\text{age} > 35} (\text{employee})]$$
$$B \leftarrow \prod_{\text{name}} [\sigma_{\text{author} = \text{'Hemingway'}} (\text{employee} \infty \text{loan} \infty \text{books})]$$

Result : A - B

7. Print the names of employees who have borrowed all books authored by Hemingway.

$$\Pi_{\text{name, isbn}} (\text{employee} \bowtie \text{loan} \bowtie \text{books}) \div$$
$$\Pi_{\text{isbn}} [\sigma_{\text{author}='Hemingway'} (\text{books})]$$

**Problem III.** Concurrency control (20 pts)

1. (10pts) Explain what the ACID features are for transaction management in databases. Briefly explain each item.

A

C

I

D

Consider the following schedule  $S_1$ ,

T1	T2	T3
	Read(X)	
	Read(Y)	
	Write(X)	
Read(Y)		
		Write(Y)
	Write(Y)	
Write(Y)		

2. (5pts) Draw the precedence graph of schedule  $S_1$ ,

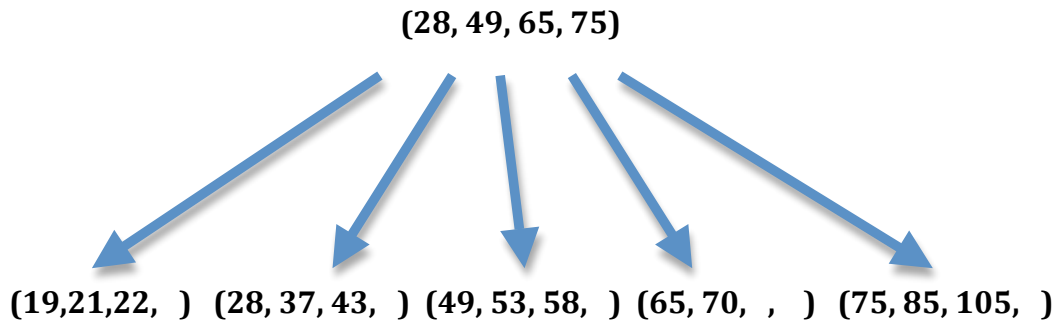
3. (5pts) Is  $S_1$  conflict serializable? Explain your answer.

**Problem IV. Indexing (25pts)**

1. (10pts) Construct a B<sup>+</sup>-tree for the following set of key values:

(37, 28, 53, 49, 75, 27, 65, 43, 21, 85, 58, 19, 105, 70)

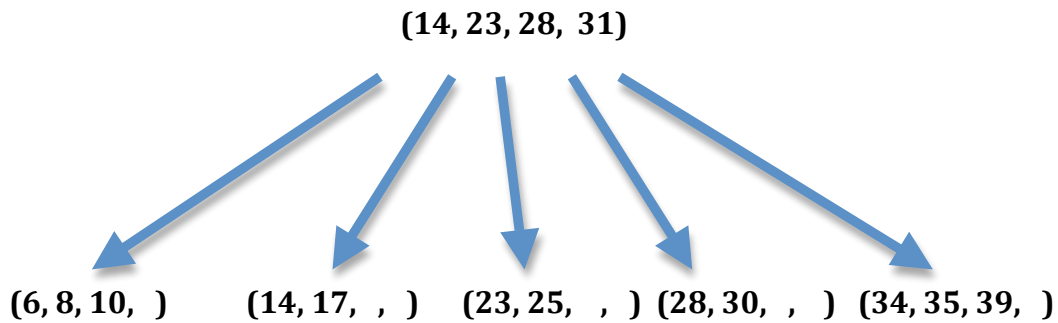
Assume the tree is initially empty and the values are added in the order specified by the above list. The number of pointers that will fit in one node is FIVE. Draw intermediate trees for partial credits.



Given the following B<sup>+</sup>-tree:



4. (5pts) Draw the tree after deleting key value 4 from the resulting tree of question 3.



**Problem V.** Relational database design (15 pts)

Given a relational schema  $R = (A, B, C, D, E)$  and the following set of functional dependencies:

- 1)  $A \rightarrow BC$     2)  $CD \rightarrow E$     3)  $B \rightarrow D$     4)  $E \rightarrow A$

Answer the following questions:

1. (4pts) List at least one candidate key of this schema, show your reasoning step by step to get partial credits.

**A, E, CD, BC are all candidate keys.**

2. (4 pts) Is R in Boyce-Codde Normal form? Explain your answer.

**No.**

**Definition of BCNF: for any  $X \rightarrow A$ , X must be a key. FD 3) violates this definition since B is not a key.**

3. (4 pts) Is R in 3rd normal form? Explain your answer.

**Yes.**

**Definition of 3NF: for any  $X \rightarrow A$ , it should satisfy either 1) X is a key; or 2) A is a prime attribute. None of the 4 FDs violates this definition.**

4. (3 pts) Is R in 2<sup>nd</sup> Normal Form? Explain your answer.

**Yes.**

**Definition of 2NF: no non-prime attribute functionally depends on a partial key. Again, all 4 FDs satisfy this.**

**Another way to reason: since R is in 3NF, it must be in 2NF.**