Carsten Schürmann Date: December 16, 2009

# FINAL EXAMINATION DBS - Database Systems Fall 2009

## Instructions

- Written examination with all written and printed aids (A2).
- There are six pages.
- This examination consists of four problems.
- Each problem is worth 25 points.
- Each problem contains multiple questions.
- The point value of each quesition is given explicitly.
- Read each question completely before attempting to solve any part.

#### Problem 1: SQL [30 points]

There are six tables describing a company, describing employees, departments, buildings, which department(s) an employee works in (and a percentage of the time for each), department managers (possibly more than one per department), and in which building an employee works (an employee may have more than one office). The primary key of each table is the attribute(s) in capitals. Other attributes are not necessarily unique. Consider the following relational schemas

```
EMP (EID, EName, Salary, Start_Date, End_Date)
BUILDING (BID, BName, Address)
DEPT (DID, DName, Annual_Budget)
IN_DEPT (EID, DID, Percent_Time)
IN_BUILDING (EID, BID)
MANAGES_DEP (EID, DID)
```

*Hint:* In the following you will be asked if a query works or not. To explain that it does not work, you may need to define sample content of relations and show that the query computes the wrong result.

- 1. Which of the following queries finds the names of buildings where more than 50 employees work? (If you think it does write "Yes", if not write "No" and explain why not.) [10 points]
  - (a) SELECT Bname FROM IN\_BUILDING GROUP\_BY BID WHERE Count(\*) > 50
  - (b) SELECT Bname FROM BUILDING WHERE BID IN (SELECT BID FROM In\_Building GROUP BY BID HAVING Count(\*) > 50)
  - (c) SELECT Bname FROM Building B, In\_Building I WHERE B.BID = I.BID GROUP BY B.BID HAVING Count(\*) > 50
  - (d) SELECT Bname FROM Building B WHERE 50 < (SELECT Count(\*) FROM In\_Building I WHERE I.BID = B.BID)
- 2. Which of the following queries finds the name of Departments where no employees work? (If you think it does write "Yes", if not write "No" and explain why not.) [10 points]

```
(a) SELECT Dname
FROM Dept
WHERE DID IN (SELECT I.DID
FROM In_Dept I
GROUP BY I.DID
HAVING COUNT(*) = 0)
(b) SELECT Dname
FROM Dept D, In_Dept I, Emp E
WHERE I.EID = E.EID AND D.DID = I.DID AND Count(E.EID) = 0
```

```
(c) SELECT Dname
FROM Dept
WHERE DID NOT IN (SELECT DISTINCT DID FROM In_Dept I)
(d) SELECT Dname
FROM Dept D
WHERE Not Exists (SELECT *
FROM In_Dept I, EMP
WHERE I.EID = EMP.EID AND I.DID = D.DID)
```

3. Which of the following queries finds the name of the Department(s) where the highest paid employee works? (If you think it does write "Yes", if not write "No" and explain why not.) [10 points]

```
(a) SELECT D.Dname
   FROM Dept D
   WHERE D.DID IN (SELECT T.DID, MAX(Salary)
                   FROM Dept T, In_Dept I, Emp E
                   WHERE T.DID = I.DID AND E.EID = I.EID)
(b) SELECT Dname
   FROM Dept D, In_Dept I, Emp E
   WHERE D.DID = I.DID AND E.EID = I.EID AND E.Salary = MAX(Salary)
(c) SELECT DName
   FROM Dept D, In_Dept I, Emp E
   WHERE D.DID = I.DID AND E.EID = I.EID
         AND E.Salary >= ALL (SELECT Salary FROM EMP)
(d) SELECT DName
   FROM Dept
   WHERE DID IN (SELECT I.DID
                 FROM In_Dept I, Emp E
                 WHERE I.EID = E.EID
                       AND E.Salary = (SELECT MAX(Salary) FROM Emp))
```

## Problem 2: Query Optimization [20 points]

Reconsider the schema from Problem 1 with the following statistical information about number of tuples/pages for each relation.

Relation	Tuples	Pages
EMP	100 000	1 000
BUILDING	2  000	10
DEPT	1 000	5
IN_DEPT	110000	550
IN_BUILDING	110000	550
MANAGES_DEPT	800	4

Furthermore, consider the following query:

SELECT Bname FROM EMP E, Building B, In\_Building I WHERE E.EID < 500 AND E.EID = I.EID AND B.BID = I.BID

- 1. Write this query in relational algebra. (5 points)
- 2. If the database has and B\*-Tree index on EMP.EID, what is the best plan you can find to execute this query? Show the query plan here, including the costs for each step and the total cost. (15 points)

#### Problem 3: Functional Dependencies, Normalization [25 points]<sup>5</sup>

Consider the relation schema R = (A, B, C, D, E, F) and the set of functional dependencies  $F : A \to B, A \to C, BC \to E, BC \to D, E \to F, BC \to F$ . Note that in all the following, you will never have to compute the closure of F!

- 1. List the minimal candidate key(s) for R. Write "none" if you think there are no candidate keys. [5 points]
- 2. List the Functional Dependencies in F that violate BCNF. (Hint: There are four) [5 points]
- 3. Is R in 3NF (yes or no)? [5 points]
- 4. Is F a canonical cover? [5 points]
- 5. Suppose we decompose R into the following tables:

$$\begin{array}{rcl} R1 &=& (B,C,E) \\ R2 &=& (B,C,F) \\ R3 &=& (B,C,D) \\ R4 &=& (A,B,C). \end{array}$$

This decomposed schema is indeed in BCNF (you can trust us on this!) Unfortunately, this decomposition is not dependency-preserving; in particular, the dependency  $E \to F$  cannot be checked on a single table. A CHECK ASSERTION can be used to enforce  $E \to F$ . Complete the following SQL statement for this particular CHECK ASSERTION needed to guarantee  $E \to F$ . [5 points]

```
CREATE ASSERTION checkDep
CHECK ( NOT EXISTS
( SELECT * FROM R1, R2
WHERE ______
GROUP BY ______
HAVING COUNT (_____) _____
)
```

## Problem 4: Concurrency Control [25 points]

1. Consider the following modified definition of serializability: A schedule S is serializable iff it produces the same database state as a serial schedule T, where the transactions in T are exactly those in S, and are ordered in T according to their first appearance in S. Is this definition: [5 points]

6

- (a) Correct?
- (b) Overly restrictive: i.e. there are some serializable schedules not covered by this definition?
- (c) Overly permissive: i.e. there are some unserializable schedules that are covered by this definition?
- (d) (ii) and (iii)

Choose one and justify your choice.

2. Bob and Anne share a bank account for their business. Today they went to the bank at the same time. Is the following schedule conflict serializable? If so, give an equivalent serial schedule. If not, enumerate all the serial schedules, and explain how Bob and Anne's experience would be changed in each. [10 points]

Transaction for An	ne	Transaction for Bo	b
Looks at checking balance	R(C)		
		Looks at savings balance	 R(S)
		Looks at checking balance	R(C)
		Transfer \$200 from checking	 W(C)
		to savings	 W(S)
Withdraw \$100 from checking	 W(C)		
			Commit
	Commit		

3. Suppose that the database system at the bank implemented strict 2-phase locking as we studied in class. Assume that Bob and Anne's requests for actions arrive in the same order as in 2., but if either of them is blocked while waiting for a lock, their actions stop arriving until they acquire the lock (after which time they continue as fast as they can). Describe what happens in that scenario: a few words should suffice. [10 points]