

Naam: _____

Collegekaart: _____

Tentamen Database Techniek

Dinsdag, 27 Maart 2007, 14.00 – 17.00 uur

Motiveer Uw antwoorden.

*Antwoorden zonder motivatie zijn **FOUT!***

U mag antwoorden in het Nederlands of in het Engels.

*Bij dit tentamen mogen **geen** boeken en dictaten worden gebruikt.*

Motivate your answers.

*Answers without motivation are considered **WRONG!***

You may answer in English or Dutch.

*Books or scripts must **not** be used during the test/exam.*

Storage, Indexing & Hashing

1. (5 pt) When is it preferable to use a dense index rather than a sparse index? Explain your answer.
2. (5 pt) What is the difference between a clustering index and a non-clustering index?
3. (5 pt) Why is a hash structure not the best choice for a search key on which range queries are likely?
4. (10 pt) Construct a B+-tree for the following list of values. Assume that each node can contain at most 3 pointers. Start with an empty B+-tree and insert the values in the given order.
12,18,20,22,44,66,32,36,38,40.
5. Describe for RAID-0, RAID-1 and RAID-5 volumes consisting of N magnetic disks:
 - (a) (1 pt) what is the usable capacity (size) of the entire RAID disk?
 - (b) (1 pt) how many disks must fail before data loss can occur?
 - (c) (2 pt) for large file reads, what bandwidth can maximally be achieved?
 - (d) (2 pt) for large file writes, what bandwidth can maximally be achieved?
 - (e) (2 pt) what is the maximum amount of single-block reads that can occur at the same time?
 - (f) (2 pt) what is the maximum amount of single-block writes that can occur at the same time?(with respect to the size and bandwidth of a single magnetic disk).

Query Processing & Optimization.

6. (7 pt) Name and describe three different join algorithms. Discuss and compare their complexity.
7. (7 pt) Consider two tables T1(a,b), T2(x,y) and the following two SQL queries (C is an arbitrary constant):

```
select * from T1,T2 where T1.a between T2.y-C and T2.y+C;
select T1.a,T2.y from T1,T2 where T1.b = T2.x;
```

For both queries, say which of the above join algorithms could be applied to evaluate the query and which ones cannot. Explain your answers!

8. Consider the following database schema:
Students(SID, Name), Lecturers(LID, Name), Courses(CID, Title, Lecturer), Exams(EID, Course, Student, Mark)
with xID the primary key of each relation and Lecturer, Course, Student the obvious foreign keys.

Consider further the following SQL query that selects all students that failed to pass a course given by professor Kersten:

```
select s.Name, s.SID
from Courses c, Exams e, Lecturers l, Students s
where l.LID = c.Lecturer and l.Name = "Kersten" and
      c.CID = e.Course and e.Mark >= 6 and
      s.SID = e.Student;
```

- (a) (7 pt) Translate this SQL query into a *canonical* relational algebra expression. Depict the algebra expression as a graph (tree).
- (b) (7 pt) Name two simple heuristics to transform the canonical tree into an improved tree, trying to minimize intermediate result sizes. Depict a possible result tree as a graph.
- (c) (7 pt) Describe an alternative optimization technique to applying heuristics. Could you use this technique here? If so, how? If not, why not?

Transaction Processing, Concurrency Control, & Recovery.

9. (5 pt) What benefit does strict two-phase locking provide? What disadvantages result?
10. (5 pt) Most implementations of database systems use strict two-phase locking. Suggest three reasons for the popularity of this protocol.
11. Serializability tests:
- (a) (5 pt) Describe a simple technique to test, whether a given concurrent schedule is conflict serializable.
- (b) (5 pt) Use this technique to test the following sample schedule for conflict serializability. Explain each step.

	T_1	T_2	T_3	T_4	T_5
1				read(X)	
2					read(Y)
3					read(Z)
4			read(V)		
5			read(W)		
6			read(W)		
7				read(Y)	
8				write(Y)	
9		write(Z)			
10					read(U)
11	read(Y)				
12	write(Y)				
13	read(Z)				
14	write(Z)				
15					read(U)
16					write(U)

- (c) (5 pt) Is there a similarly *simple* technique to test for view serializability? If yes, name it. If not, why not?
12. (5 pt) Explain the difference between a system crash and a “disaster”.

(Total: 100 pt)

Success!