
Information Systems ***September 18, 2014***
Master of Science in Computer Engineering

Exercise 1 (9 points)

Let's consider the following relational schema:

EMPLOYEE(EName, Phone, Address, Depart, Salary)
DEPARTMENT(D_Id, DName, City, Budget)
PROJECT(Project_Id, Title, Budget, starting_date, end_date)
CONTRACT(Contract_Id, Empl, Project, from_date, to_date)

Primary keys are underlined in the relations. Moreover, Depart in EMPLOYEE is foreign key of DEPARTMENT; Empl in CONTRACT is foreign key of EMPLOYEE and Project in CONTRACT is foreign key of PROJECT.
An employee can have contracts on different projects at a given time.

Assume that:

$n_{EMPLOYEE}$ =10.000	V(D_Id, CONTRACTS) = 50
$n_{DEPARTMENT}$= 50	V(Project_Id, CONTRACT) = 150
$n_{PROJECT}$= 250	V(Empl, CONTRACT) = 10.000
$n_{CONTRACT}$ = 1.000.000	V(City, DEPARTMENT) = 5

Given the query:
"Name and Salary of employees with a contract on the project "Software reliability", working at a Department in Pisa"

1) express the query as a relational-algebra expression;
2) show the basic steps of the query optimization process in terms of relational-algebra expression transformations
3) give an efficient strategy for computing the query.

Exercise 2 (6 points)
Consider the following schedule of concurrent transactions:

S: `w2(y) w3(z) w1(y) w3(x) r2(x) w1(x) r2(z)`

1) Show if it is conflict serializable (CSR) or view serializable (VSR). Explain why. If serializable, show equivalent serial schedules.
2) Apply the rigorous two-phase locking protocol to the schedule. Is the schedule accepted?
3) Apply the timestamp-ordering protocol to the schedule, assuming that aborted transactions are immediately restarted. Is the schedule accepted?

Exercise 3 (6 points)
Consider an empty B+-tree with fanout m= 4.

1) Show the B+-tree after the insertion of the following values of the search key (assume that the tree is initially empty):
5 21 8 16 13 3 19 50 4 7

2) Show the form of the B+-tree after each operation of the sequence:
Delete 8; Delete 13

Exercise 4 (9 points)

Let r=(A,B,C), with A a key.
Assume
nr = 800.000 number of records in the relation
Lr = 50 byte size of a record (fixed length records)
LA = 6 byte size of attribute A
Lp = 4 byte size of a pointer
Lb = 1000 byte size of a block
Heap file organization on search-key A.

1. Show the number of blocks of a **B+-tree index** on search-key A, assuming that each block contains 70% of the records that can fit in a block.

2. Outline the steps in answering the following queries, showing **the best strategy** and **the cost** in terms of number of block transfers from disk:

1) select * from r where A=xxx;

2) select * from r where 100.000 <= A < 500.000;
assuming A uniformly distributed on the interval [1; 1.000.000]

3) select * from R where B=xxx;