
Information Systems January 29, 2016
Master of Science in Computer Engineering

Exercise 1 (9 points)
Let's consider the following relational schema for a group of insurance companies located in different cities:

CUSTOMER(Id_cust, Name, Gender, YearOfBirth)
INSURANCE_COMPANY(Id_company, _Director_name, Budget, City)
POLICY(Policy_number, Id_cust, Id_company, Policy_type, expiry_date)

Primary keys are underlined in the relations. Moreover, Id_cust in POLICY is foreign key of CUSTOMER and Id_company in POLICY is foreign key of INSURANCE_COMPANY.
A customer can have more than one policy in the same company or in different companies.

Assume that:

$n_{CUSTOMER}$ =5000
$n_{INSURANCE\_COMPANY}$ = 28
n_{POLICY} = 800.000

V(Id_cust, POLICY) = 5000
V(Id_company, POLICY) = 28
V(City, INSURANCE_COMPANY) = 7
V(YearOfBirth, CUSTOMER) = 40

Given the query:
Name of customers born in 1960 with policies stipulated with companies located in Florence or 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: w3(y) r2(z) r2(y) r1(y) w2(x) r3(z) w3(x) w1(z) w1(x)

1) Show if S 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.
3) Apply the timestamp-ordering protocol to the schedule, assuming that aborted transactions are immediately restarted.

Exercise 3 (6 points)
Given the Log file,

<start, T1> <start, T2> <T2, O1, V1, V1'> <T1, O2, V2, V2'<start, T3> <start, T4>
<T3, O3, V3, V3'> checkpoint (L) <T4, O4, V4, V4'> <commit, T4> <start, T5> <T3, O5, V5, V5'> <T5, O6, V6, V6'> <T3, O7, V7, V7'> <commit, T3> <commit, T5> <T2, O8, V8, V8'> CRASH

apply the Log-based recovery algorithm, and show
1) the list L of transactions active at the checkpoint.
2) the undo-list and the redo-list.
3) the actions executed by the system in the correct order.

<T, Oi, V, V'> : transaction T updates object Oi;
V is the state of Oi before the update; V' is the state of Oi after the update.

Exercise 4 (9 points)

Let r=(A,B,C), with primary key A uniformly distributed on the interval [1; 1.000.000].

Assume
nr = 100.000 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 = 1024 byte size of a block
Sequential file organization on A.

1. Show (a) the minimum and (b) the maximum height of a B+tree tree index on search-key A.

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

1) select * from r where A=xxx;
2) select * from r where 200.000 <= A < 700.000;
3) select * from r where B=xxx