Databases Spring 2005 - Exercise 3

Due: May 13, 17:00

1 Query Evaluation

Let *Sailors (sid:int, sname:string, saddress:string, sage:int, srating: int), Reserves(sid:int, bid:int, day:date), Boats(bid:int, color: string),* denote a database schema. Some additional information concerning this database is given:

1. The number of boats is always about 50. Ten boat tuples fit in one page.
2. There are only eight different colors for boats.
3. There are 100 to 800 reservations in the database, depending on the season. Ten reservation tuples fit in one page.
4. There are 10,000 sailors. Ten sailor tuples fit in one page.
5. Only 15% of sailors are younger than 25 years.

Consider the following SQL query:

```
SELECT sname, sid
FROM Sailors S, Reserves R, Boats B
WHERE R.sid = S.sid AND R.bid = B.bid
AND B.color = red AND S.sage < 20
```

- Translate the query in a straightforward manner to *some* RA-tree!
- Derive an optimized RA tree structure by pushing selections and writing the RA tree in a form that supports on the fly evaluations (pipelining)!
- Derive an evaluation plan for each node of the tree! Consider (only) the introduction of B+ and/or hash indexes and use (page-based, tuple, index) nested loops for the evaluation of joins! Provide an estimation for the cost of your plan!

2 Embedded SQL

In this part of the exercise you are required to supply an upper-language executable which runs the following embedded SQL specs. Notice that you are free to implement this task with ProC, JDBC, PHP3 or any other platform, as long as it connects to your LIACS Oracle account and performs all the specified operations on that account.

- Let the following **relation definitions** describe the world of Sports, as was given in Ex-2:

  Sportsman(sid, sname, sage, scountry)
  Coach(cid, cname, cage, ccountry)
  Coaches(cid, sid)

- Your program should assume that the tables corresponding to the above information already exist in the database and contain data. In order to test your program, you are advised to generate those tables and feed them with data.
- Your program should ask the user for login and password, and connect right-after to the LIACS Oracle DB, based on the given information.

- Once the connection to the database has been established, the user will be given a menu with the possible queries to be executed.
  Implement the following 5 queries in embedded SQL:

  1. Find the names of sportsmen who are from Norway, older than 25, and have a coach (at least one).
  2. Find the names of all coaches who coach a sportsman from their own country and they are not the youngest coaches.
  3. Find the names of all sportsmen whose names begin with the letter 'K' (capital), ends with the letter 'e' (not capital), and the letter 't' (not capital) appears in their name.
  4. Find the pairs of names of (sportsman, coach) where the sportsman is coached by the coach, and the age difference between them is less than 4 years.
  5. Find the names of coaches from Kazakhstan who coach more than 3 sportsmen.

- Error handling should be done as was in the sample file, given to you in a previous werkcollege.

Submission Guidelines for the Embedded-SQL Part

- Your executable must connect to the Oracle system of LIACS and act on any account on that system!
- Produce a README file, containing all your comments concerning your implementation.
- You are required to submit electronically a file **ex3.tar** containing the following: source code, Makefile, the executable and the README file.
- Submit hard copies of your code and your README file.

3 Transaction management

Consider the schedules of interleaved transactions displayed above:

Schedule A:

T1			R(A)	R(B)	W(A)		commit
T2	R(A)	W(A)				W(B)	commit

Schedule B:

T1	R(A)	W(A)			R(B)	W(B)		commit
T2			R(C)	R(A)			W(C)	commit

Schedule C:

T1				W(A)		W(B)	commit
T2	R(A)	W(A)	R(B)		W(B)		commit

Schedule D:

T1	R(A)	W(A)			R(B)	W(B)			commit
T2			R(B)	W(B)			R(A)	W(B)	commit

- Which of these schedules are *serializable* and/or *conflict serializable* ? For those schedules that are not conflict serializable classify the anomaly (WW, RW, WR)!
- Assume strict 2PL would be used to schedule transactions. Describe the resulting schedules for the transactions given above ! Use X(.) in order to indicate that an exclusive lock is set and S(.) in order to indicate that an shared lock is set.
- For one of the schedules the application of the strict 2PL leads to an deadlock.

– Demonstrate by means of this example, how a deadlock can be recognized by a Waits-for graph.

– Instead of strict 2PL, apply conservative 2PL to the schedule. How does it affect the schedule ?

Another schedule is given by (cf. Sect. 6.3.4, in Ramakrishnan, Gehrke):

Schedule 5:

T1	R(A)	W(A)						abort
T2			R(A)	W(A)	R(B)	W(B)	commit	

- Demonstrate the difference of 2PL and strict 2PL with respect to recoverability by means of this schedule!

4 Schema Refinement

A new mobile media player product X-Bod shall be launched by a company called Macrosoft. The player allows to watch films and listen to audio, all of which can be downloaded as media files from the Macrosoft site.
Right now, data on the music files is stored in a simply structured file. An exerpt of this file is given in the following table:

Title	Artist	Genre	Format	Duration	Publisher
Heli-Cops	Rotorhead	Hardrock, Soundtrack	Music	4:02	EMI
Romance	Vanessa Mae	Classic, Pop	Music	20:00	EMI
Maximum BPM	Strobos-kop	Elektro, Hardtrance	Music	1:34	CEC
Solaris	Sonderbergh	Sci-Fi	Film	110:56	Universal
Data Base Pusher	Strobos-kop	Elektro, Hardtrance	Music	23:34	CEC
Solaris	Takorsky	Sci-Fi	Film	87:00	KGB
Divided	Blockkids	Pop	Music	2:34	Virgin
Reunited	Blockkids	Pop	Music	4:23	Virgin
Easy Rider	Bob Harley	Hardrock, Reggae	Music	3:34	CIA
Easy Rider	Hopper	Road Movie	Film	91:00	Warner Boss
Titanic	Cameron	Drama, Romance	Film	94:43	Paramount
Deadlock	Stones	Action, Sci-Fi	Film	120:0	W.Disney
The 6th Normal form	Ramakrishna	Sci-Fi	Film	120:0	McGraw
⋮					⋮

Mrs. Gill Bates, the CTO of Macrosoft, decided to establish a professional database for the management of media files. The following information from the requirement analysis should be taken into account:

1. Artists have unique names
2. The genres of films and music titles have disjoint names (e.g. Sci-Fi is exclusively a film-genre and Rock is exclusively a music genre)
3. An artist name together with the title uniquely identifies a media file
4. Artists are always related to exactly the same list of genres
5. Artists are featured exclusively by one publisher

Your task is to design a database with a refined schema:

- Derive a schema in 1st normal form! Identify keys for all tables!
- Identify all non-trivial functional dependencies of this schema!

- Derive a lossless-join decomposition of the schema in 3NF!
- Discuss differences of BCNF and 3NF! Does this difference matter in the given example ?

Use the abbreviation T for Title, A for Artist, G for Genre, F for format and D for duration.

General

The homepage of the course is:

`http://www.liacs.nl/databases`

Assistants:

Ruben van Bodegom (S.A.): `rvbodego@liacs.nl`
Ofer Shir (AIO): `oshir@liacs.nl`

Electronic submission: **ex3.tar** should be sent to

`csdb@liacs.nl`

Exercises have to be submitted on a hard copy (including the source code). Please leave a copy in the mailbox labeled "O. Shir" in room 156.

SUXES!