November 03, 2016

Arnab Chakrabarti, Rihan Hai, Sandra Geisler

Implementation of Databases (WS 16/17)

Exercise 2

Due until November 15, 2016, 2pm.
Please submit your solution *in a single PDF file* before the deadline to the L^{2}P system!
Please submit solutions in groups of three students.

Exercise 2.1 (Query Languages) (5 pts)

1. Variants of relational algebra are used internally in DBMS to represent queries and query evaluation plans. Please explain, why relational algebra is suitable for this?

2. What does "relational completeness" mean (in your own words, please)? Show that SQL is relationally complete by enumerating SQL constructs corresponding to selection, projection, cartesian product, union, and difference.

3. Explain how the intersection operator is used in relational algebra. What is important for its usage? Is it omittable (i.e., could you express it by other operators)? If so, please give the corresponding relational algebra.

4. Explain the difference between DRC and TRC.

Exercise 2.2 (Domain Relational Calculus) (10 pts)

The exercise is based on the Mondial database, whose schema can be accessed at `http://www.dbis.informatik.uni-goettingen.de/Mondial/` This database schema will be used throughout the exercises. The database dump for the corresponding PostgreSQL database as well as an installation description can be downloaded from the L2P. There you will also find a pdf file describing postgreSQL and its usage.

For the set of tasks as stated below, formulate the **Domain Relational Calculus** expressions

1. The names of the seas that are deeper than 3000 mt and into which at least one river finally flows.

2. The names of the groups of islands that consist only of islands that are neither 'volcanic' nor 'coral'.

3. The name of the mountain, the name of the mountains where it belongs to and the name of the country where the mountain is located for each mountain that is located on an island that is of type 'coral'.

4. Give the name and country name of all cities with a population higher than 500,000.

Exercise 2.3 (Sorting) **(5 pts)**

Suppose you have a file of 15,000 pages and eight buffer pages and you are sorting it using general (external) merge-sort. Please answer the following questions:

1. How many runs will you produce? Remark: When a file is sorted, in intermediate steps subfiles are created. Each sorted subfile is called a *run*. See also slide 35 in Chapter 2.
2. How many passes will it take to sort the file completely?
3. How many buffer pages do you need at least to sort the file in two passes?
4. How many runs and passes would a Two-Way-Sort algorithm take?

Exercise 2.4 (Join Implementations) **(4 pts)**

Given are the two relations Album and Tracks, with the following specifics:

- Album has a size of 10.000 pages, 40 bytes record size and 100 tuples/page
- Track has a size of 200.000 pages, 30 bytes record size and 80 tuples/page

You have 16 buffer pages available.

1. Calculate the I/O requirements of a simple nested loop join
2. Calculate the I/O requirements of a block nested loop join
3. Explain the differences between the two algorithms. What are the similarities and differences? How does the block nested loop join reduce I/O costs?

Exercise 2.5 (Selection) **(6 pts)**

Given is a relation with 50.000 records. Each page for a node in a B^+-tree can hold 20 pointers to records or pages. A data page can store 20 records.

1. Assume that each node is 70 % full. What is the height of the B^+-tree?
2. What are the I/O costs for an equality selection on a non-key attribute for the following cases?
   (a) with a clustered B^+-tree of height 3 (matching records are located in one page);
   (b) without any index, nor is the file sorted on the attribute occurring in selection;
   (c) with an unclustered B^+-tree index of height 4, and there are 2 matching records;
   (d) with an unclustered B^+-tree of height 5 and three tenth of the records match the selection.