November 17, 2016
Arnab Chakrabarti, Rihan Hai, Sandra Geisler

Implementation of Databases (WS 16/17)

Exercise 3

Due until November 29, 2016, 2pm.
Please submit your solution *in a single PDF file* before the deadline to the L²P system!
Please submit solutions in groups of three students.

Exercise 3.1 (Tableau) **(8 pts)**

1. Let R be a relation of attributes A, B and C. Given is a query in relational algebra (RA):

$$\pi_{A,B}(R) \bowtie \pi_{B,C}(\pi_{A,C}(\sigma_{B=1}(R)) \bowtie \pi_{A,B}(R))$$

   (a) Translate the RA query into an equivalent query in Domain Relational Calculus (DRC).

   (b) Construct an equivalent tableau of the query.

2. Given the following two tableau queries T_1 and T_2, decide whether they are equivalent, subsumed by each other ($T_1 \subseteq T_2$ and/or $T_2 \subseteq T_1$), or not related at all.

T_1	a	b		
	a	b_1	c_1	R
	a_1	b	c_1	R
	a_1	b_2	c_2	R
	a	b_3	c_3	R

T_2	a	b		
	a_5	b	c_4	R
	a	b_5	c_4	R

Exercise 3.2 (Quant graph) (9 pts)

Given is the following relational database schema (keys are underlined):

$$
\begin{aligned}
&Chef(\underline{CID}, ChefName, Nationality, CityOfBirth)\\
&Brand(\underline{BID}, BrandName)\\
&Shop(\underline{SHID}, ShopName, Address, City, BID)\\
&Sweet(\underline{SID}, SweetName, Country)\\[6pt]
&Create(\underline{CID}, \underline{BID}, YearOfCreation)\\
&Purchase(\underline{SID}, \underline{SHID}, DeliveryMode, Price)\\[6pt]
&Create[CID] \subseteq Chef[CID]\\
&Create[BID] \subseteq Brand[BID]\\
&Shop[BID] \subseteq Brand[BID]\\
&Purchase[SID] \subseteq Sweet[SID]\\
&Purchase[SHID] \subseteq Shop[SHID]
\end{aligned}
$$

1. **TRC:**
   Specify the following query in the tuple relational calculus (TRC) and draw the corresponding quant graph:
   Find the names of the chefs, who created the brand, and the shops of his brand are located in the same city where the chef was born.
2. Does the graph contain a cycle? What does the result mean for optimization?

Exercise 3.3 (Cost Estimation) (13 pts)

Consider the following relational schema and SQL query. The schema captures information about employees, departments, and projects.

$$Emp(\underline{eid}, sal, age, did)$$
$$DeptProj(\underline{did, pid}, budget)$$
$$Proj(\underline{pid}, code, report)$$

$$Emp[did] \subseteq Dept[did]$$
$$Dept[pid] \subseteq Proj[pid]$$

Assume that each Emp record is 20 bytes long, each DeptProj record is 40 bytes long, and each Proj record is 2000 bytes long. There are 20,000 tuples in Emp, 5000 tuples in DeptProj, and 1000 tuples in Proj. Each department, identified by *did*, has 10 projects on average. The file system supports 4000 byte pages, and 12 buffer pages are available. All following questions are based on this information. You can assume uniform distribution of values. State any additional assumptions which you do to answer the questions. The cost metric to use is the number of page I/Os. The costs for an index access is 3 in all cases. Ignore the cost of writing out the final result.

1. Compute the number of pages for each relation.
2. Consider the following query: "Find all employees with age more than 30." Assume that there is an unclustered index on *age*. Let the number of qualifying tuples be N. For what values of N is a sequential scan cheaper than using the index?
3. Consider the following query:

   ```
   SELECT *
   FROM Emp E, DeptProj D
   WHERE E.did=D.did
   ```

   (a) Compute the costs for the query using a block-nested loop join.

   (b) Suppose that there is a clustered hash index on *did* on *Emp*. Compute the costs for the query using an index-nested loop join.

   (c) Assume that both relations are sorted on the join column. Which join method should be applied and what are the costs?

   (d) Suppose that there is a clustered B+ tree index on *did* on *Emp* and *DeptProj* is sorted on *did*. Which join method should be applied and what are the costs?