Universität Freiburg   Georges-Köhler Allee, Geb. 51
Institut für Informatik   D-79110 Freiburg
Prof. Dr. G. Lausen   lausen@informatik.uni-freiburg.de
Alexander Schätzle   schaetzl@informatik.uni-freiburg.de
Martin Przyjaciel-Zablocki   zablocki@informatik.uni-freiburg.de

Advanced Information Systems
Summerterm 2011
05.07.2011

4. Exercise Sheet: Distributed Databases

Discussion: 07.07.2011

Submission Guidlines: We will discuss the solutions to the exercise sheet on 07.07.2011. If you want to have comments on your solutions you can submit them after the lesson.

Exercise 1 (Distributed Join Processing)
Consider relations EMP(E,N), PROJ(P,B) and ASG(P,E) for employees, projects and assignments of employees to projects. Assume further that EMP is stored at site 1, PROJ at site 2 and ASG at site 3.

a) Consider the query PROJ ⋈ ASG ⋈ EMP and describe at least 5 different ways to compute the result at any site.

b) For each of your versions in (a) give an example for the size of the relations and intermediate results such that this version is most efficient in overall communication costs.

Exercise 2 (Semijoin & Bloomjoin)
Assume a relation R(A,B) at site 1 and a relation S(B,C) at site 2 as follows:

R	A	B		S	B	C
	1	2			0	0
	3	4			1	1
	5	6			2	2
	7	8			3	3
	9	10			4	4

a) Apply the Semijoin-Algorithm to compute R ⋈ S and describe the necessary steps.

b) Compute R ⋈ S using Bloomjoin with $h(t_R[B]) = t_R[B] \bmod 4$ and describe the necessary steps.

Exercise 3 (Semijoin-Program)
A Semijoin-Program for relations $R_1, ..., R_n$ is a sequence of instructions:

$$R_{i_1} := R_{i_1} \ltimes R_{j_1} ;$$
$$R_{i_2} := R_{i_2} \ltimes R_{j_2} ;$$
$$...$$
$$R_{i_p} := R_{i_p} \ltimes R_{j_p} ;$$

a) Describe how a Semijoin-Program can be used to achieve an efficient processing of the query PROJ $\bowtie$ ASG $\bowtie$ EMP of Exercise 1.

b) Give a Semijoin-Program with the smallest number of instructions that avoids dangling tuples when computing the query PROJ $\bowtie$ ASG $\bowtie$ EMP of Exercise 1.

Exercise 4 (Semijoin-Program)
Consider relations EMP(E,N,C), PROJ(P,B,C), ASG(P,E).

a) Explain, why your Semijoin-Program of Exercise 3 will not be able to remove all dangling tuples before processing the query PROJ $\bowtie$ ASG $\bowtie$ EMP.

b) Show that there are instances of EMP, PROJ and ASG such that a Semijoin-Program will not be able to remove any dangling tuples.

c) Give an instance of EMP, PROJ and ASG to demonstrate that to remove all dangling tuples we may need a Semijoin-Program of length proportional to the number of tuples in the relations.