Exercise 7

Matthias Gollwitzer, Jan Schalkamp

June 12, 2013

1 EXERCISE 1

Make as usual via *make* command. TPCH data is again expected in *data/tpch/tpch* with semicolons only as delimiter and no other occurrences of them.

1.1 DPSUB

- Execute via **./bin/homework7-1**
- **src/compiler/strategies/DPsubStrategy:** Implements the DPsub algorithm. We are using the algorithm that was introduced in the lecture, which enumerates subsets efficiently. The whole algorithm is essentially a 1:1 conversion from the lecture slides. Only really different part is checking the estimated sizes of operators to be joined, as tinyDB only comes with a Hash Join - so we put the smaller operation left, and the bigger right. There is a crossproduct flag in the header file. It's currently set to true, as it should be beneficial to the query we are running.
- **src/compiler/querygraph/QueryGraph:** Added the method *convertBitmapToSet()* in order to operate on sets represented each by a bitmap. Other methods added to check if a connection between two sets exists in the query graph.

1.2 DPSIZE

We did it anyway.

- Execute via **./bin/homework7-1b**
- **src/compiler/strategies/DPsubStrategy:** Implements the DPsize algorithm as shown in the lecture. Here we combine the efficient subset enumeration with efficient subset generation (using Gosper's hack[1]).

[1] `http://en.wikipedia.org/wiki/Combinatorial_number_system#Applications`

2 EXERCISE 2

- Execute via **./bin/homework7-2** (generates a pseudo-random join tree with 4 inner nodes)
- **src/compiler/randomized/DyckGenerator:** Essentially a 1:1 conversion from the lecture slides. We are not sure if it works for more than 33 inner nodes. We tend to say no, as C(34) is bigger than a unsigned long long can store.
- **src/compiler/randomized/Binomial:** Class for calculating binomial coefficients utilising memoization.