



## Query Optimization

2. Exercise

Due 21.11.2016, 9 AM

submit via email (radke@in.tum.de)

### Exercise 1

Consider the [TPC-H benchmark](#) and the query:

```
select *
  from lineitem l, orders o, customers c
 where l.l_orderkey=o.o_orderkey
        and o.o_custkey=c.c_custkey
        and c.c_name="Customer#000014993".
```

Do canonical translation and logical optimization.

### Exercise 2

Given  $|R1|$ ,  $|R2|$ , the domain of  $R1.x$  and  $R2.y$ , and the information if  $R1.x$  and/or  $R2.y$  are keys of  $R1$  and  $R2$ .

1. How can we estimate the selectivity of  $\sigma_{R1.x=c}$ , where  $c$  is a constant?
2. How can we estimate the selectivity of  $\bowtie_{R1.x=R2.y}$ ?

Note that we don't know the output size of  $\sigma_{R1.x=c}$  ( $\bowtie_{R1.x=R2.y}$ , respectively), so we can't simply use the definition of selectivity.

### Exercise 3

Given are two relations  $R$  and  $S$ , with sizes 1,000 and 100,000 pages respectively. Each page has 50 tuples. The relations are stored on a disk, the average access time for the disk is 10 ms and the transfer speed is 10,000 pages/sec. How long does it take to perform the Nested Loops Join of  $R$  and  $S$ ? How long does it take to perform the Block Nested Loops Join with a block size of 100 pages? Assume that CPU costs are negligible and ignore I/O costs for the join output.