Query Optimization: Exercise

Session 2

Bernhard Radke

October 30, 2017
Please start subject of emails regarding this exercise with `qo17`

Please attach code as tarball to your submission email (hint: `git archive`)
Homework

Exercise 1
Find all students that attended the lectures together with 'Schopenhauer', excluding Schopenhauer himself.

- SQL

```sql
select s2.name
from studenten s1, hoeren h1, hoeren h2, studenten s2
where s1.name='Schopenhauer' and s1.matrnr=h1.matrnr
    and h1.vorlnr=h2.vorlnr and h2.matrnr=s2.matrnr
    and h1.matrnr<>h2.matrnr
```

- tuple calculus

```tuple calculus
\{ s_1 | s_1 \in \text{Studenten} \land \exists h_1 \in \text{hoeren}( s_1.\text{MatrNr} = h_1.\text{MatrNr} \\
\land \exists h_2 \in \text{hoeren}( h_1.\text{VorlNr} = h_2.\text{VorlNr} \land h_1.\text{MatrNr} \neq h_2.\text{MatrNr} \\
\land \exists s_2 \in \text{Studenten}( h_2.\text{MatrNr} = s_2.\text{MatrNr} \land h_2.\text{Name} = 'Schopenhauer') \\
\) \}
```
Find all students that attended the lectures together with 'Schopenhauer', excluding Schopenhauer himself.

\[
\left\{ [n_1] \exists m_1, s_1 ([m_1, n_1, s_1] \in \text{Studenten} \land \exists v ([m_1, v] \in \text{hoeren} \land \exists m_2 ([m_2, v] \in \text{hoeren} \land m_2 \neq m_1 \land \exists n_2, s_2 ([m_2, n_2, s_2] \in \text{Studenten} \land n_2 = 'Schopenhauer') \}}\right\}
\]
Find all professor whose lectures attended at least two students

No group by in TinyDB
Textbook Optimization
Selectivity $f_R$ of a selection $\sigma(R)$

$$f_R = \frac{|\sigma(R)|}{|R|}$$

Selectivity $f_{1,2}$ of a join $R_1 \bowtie R_2$

$$f_{1,2} = \frac{|R_1 \bowtie R_2|}{|R_1 \times R_2|} = \frac{|R_1 \bowtie R_2|}{|R_1| \cdot |R_2|}$$
Basic cost function

\[ C_{\text{out}}(T) = \begin{cases} 
0 & \text{if } T \text{ is a leaf } R_i \\
|T| + C_{\text{out}}(T_1) + C_{\text{out}}(T_2) & \text{if } T = T_1 \bowtie T_2 
\end{cases} \]

Find the cheapest execution plan
Physical Optimization
Choose the actual implementation of an operator

- choosing index or table scan
  - index vs. table scan: 10% selectivity threshold
  - clustered vs. non-clustered index

- choosing types of joins
  - nested loops join
  - blockwise nested loops join
  - index nested loop join
  - merge join
  - hash join
Courses(ID, Title, Room, Time)
Exercises(ID, CID, TID, Room)
Tutors(ID, Name)

select C.Name, T.Name, E.Room
from Courses C, Tutors T, Exercises E
where C.ID = E.CID and T.ID = E.TID
    and C.Room like '02.11.\%'
    and E.Room not like '02.11.\%

clustered indexes on Exercises.TID, Tutors.ID
only clustered index on Tutors.ID
Homework
Homework

Prove an equivalence
Derive formulae to estimate selectivities
Join costs: nested loops vs. blockwise nested loops
Slides and exercises: db.in.tum.de/teaching/ws1718/queryopt
Send any questions, comments, solutions to exercises etc. to radke@in.tum.de
Exercise due: 9 AM, November 6