Query Optimization ’16
Exercise Session 1

Bernhard Radke

November 7
Organizational Matters

- Exercise sessions are here to illustrate the material of the course with examples, special cases, etc.
- Homework every week: programming assignment and 2-3 problems
- Do 75% or better and get the bonus for the final grade
- Written exam at the end
- Slides on the website
- Email subject should start with [qo16]
Textbook Optimization

Find the students that attend the course 'Ethik'
Textbook Optimization

Find the students that attend the course 'Ethik'

- SQL query
Find the students that attend the course 'Ethik'

- SQL query
- canonical translation
Textbook Optimization

Find the students that attend the course ’Ethik’

- SQL query
- canonical translation
- break up conjunctive selections
Find the students that attend the course ’Ethik’

- SQL query
- canonical translation
- break up conjunctive selections
- push down selections
Find the students that attend the course 'Ethik'

- SQL query
- canonical translation
- break up conjunctive selections
- push down selections
- introduce joins
The goal of optimization is to minimize the cost function.

Reminder: $C_{out}$

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

The goal of optimization is to minimize the cost function

Reminder: $C_{out}$

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

The goal of optimization is to minimize the cost function

Reminder: $C_{out}$

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

That's why we need join ordering!

- $|R_1| = 100$
- $|R_2| = 200$
- $|R_3| = 100$
- $f_{1,2} = 0.1$
- $f_{2,3} = 0.0001$
Cost Estimation

The goal of optimization is to minimize the cost function
Reminder: $C_{out}$

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

That's why we need join ordering!

- $|R_1| = 100$
- $|R_2| = 200$
- $|R_3| = 100$
- $f_{1,2} = 0.1$
- $f_{2,3} = 0.0001$
Cost Estimation

The goal of optimization is to minimize the cost function

Reminder: $C_{out}$

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

That’s why we need join ordering!
Info for Homework

- You can work in groups with up to two students
- Handwritten (and/or scanned) solutions will not be accepted. Use LaTeX (preferable) or Word.
- Programming assignment:
  - Implement your own query optimizer step by step
  - Initial code base (very simple database system) is available on the website
  - Language: C++11 (great opportunity to learn it btw)
  - Solutions should come with a Makefile and instructions on how to build/run it
  - Future assignments will build upon the current
Homework - Guidelines

- Submit the whole project directory, not just separate source files (no binaries!)
- You can work within the TinyDB directory, changing its structured if needed
- (Briefly) comment the code: every class, field, method, design choice
- Give examples of the input queries for which you tested. How about unit tests (e.g. googletest)
Info

- Slides and exercises:
  http://db.in.tum.de/teaching/ws1617/queryopt/
- Send any questions, comments, solutions to exercises etc. to radke@in.tum.de
- Exercises due: 9 AM, November 14