Query Optimization: Exercise

Session 1

Bernhard Radke

October 23, 2017
Exercise sessions are here to illustrate the material of the course with examples, special cases, etc.

Homework every week: programming assignment and a few problems

Do 75% or better to get a bonus of 0.3 on your exam grade

Written exam at the end of the semester

Slides on the website (db.in.tum.de/teaching/ws1718/queryopt)

Email subject should start with [qo17]
Algebra Revised
uni schema:

- Studenten : \{[MatrNr: integer, Name: string, Semester: integer]\}
- Vorlesungen : \{[VorlNr: integer, Titel: string, SWS: integer, gelesenVon: integer]\}
- Professoren : \{[PersNr: integer, Name: string, Rang: string, Raum: integer]\}
- Assistenten : \{[PersNr: integer, Name: string, Fachgebiet: string, Boss: integer]\}
- hoeren : \{[MatrNr: integer, VorlNr: integer]\}
- voraussetzen : \{[Vorgaenger: integer, Nachfolger: integer]\}
- pruefen : \{[MatrNr: integer, VorlNr: integer, PersNr: integer, Note: decimal]\}
Relational Calculus
what the result looks like (declarative)

tuple calculus: \( \{ t \mid P(t) \} \)

- \( \{ p \mid p \in \text{Professoren} \land p.\text{Rang} = 'C4' \} \)
- \( \{ s \mid s \in \text{Studenten} \land \exists h \in \text{hoeren}(s.\text{MatrNr} = h.\text{MatrNr}) \land \exists v \in \text{Vorlesungen}(h.\text{VorlNr} = v.\text{VorlNr}) \land \exists p \in \text{Professoren}(p.\text{PersNr} = v.\text{gelesenVon} \land p.\text{Name} = 'Curie')) \}

domain calculus: \( \{ [v_1, \ldots, v_n] \mid P(v_1, \ldots, v_n) \} \)

- \( \{ [p, n, r, o] \mid [p, n, r, o] \in \text{Professoren} \land r = 'C4' \} \)
- \( \{ [m, n, s] \mid \exists m([m, n, s] \in \text{Studenten} \land \exists v([m, v] \in \text{hoenen} \land \exists p([v, t, d, p] \in \text{Vorlesungen} \land \exists a([p, a, r, o] \in \text{Professoren} \land a = 'Curie')))) \} \)
compare that to SQL

- SELECT * FROM Professoren p WHERE p.Rang='C4'
- SELECT s.MatrNr, s.Name, s.Semester
  FROM Studenten s, hoeren h, Vorlesungen v, Professoren p
  WHERE s.MatrNr=h.MatrNr AND h.VorlNr=v.VorlNr AND
  v.gelesenVon=p.PersNr AND p.Name='Curie'

what the result looks like (declarative)
Relational Algebra
how the result is built (procedural)

- $\sigma_{\text{Rang} = 'C4'}(\text{Professoren})$
- $\sigma_{S.\text{MatrNr}=H.\text{MatrNr}}(S \times \sigma_{H.\text{VorlNr}=V.\text{VorlNr}}(H \times \sigma_{V.\text{gelesenVon}=P.\text{PersNr}}(V \times \sigma_{P.\text{Name}=\text{Curie}}(P))))$
- $S \bowtie (H \bowtie (V \bowtie_{V.\text{gelesenVon}=P.\text{PersNr}} \sigma_{P.\text{Name}=\text{Curie}}(P))))$
Textbook Optimization
Translate SQL into an executable plan
Many equivalent plans
Large differences in resource consumption
Minimize cost function

\[ C_{\text{out}}(T) = \begin{cases} 
0 & \text{if } T \text{ is a leaf } R_i \\
|T| + C_{\text{out}}(T_1) + C_{\text{out}}(T_2) & \text{if } T = T_1 \Join T_2 
\end{cases} \]
Find all Students that attend the course 'Ethik'

- SQL query
- canonical translation
- break up conjunctive selections
- push down selections
- introduce joins
- determine join order
- introduce and push down projections
Programming Assignments

TinyDB
TinyDB

- very simple database system
- storage layer and runtime system already implemented
- you will build a compile time system step by step
- initial code base at
  db.in.tum.de/teaching/ws1718/queryopt/tinydb.tar.gz
Homework Guidelines
Homework Guidelines

General

▶ You can work in groups of up to two students
▶ Handwritten (and/or scanned) solutions will not be accepted! Use LaTeX (preferable) or Word.
▶ Submit as PDF

Programming

▶ Target: GNU/Linux
▶ Language: c++ (great opportunity to learn it)
▶ Build: gcc, GNU make
▶ Submissions:
  ▶ Submit the whole project directory (tarball, man git archive, no binaries!)
  ▶ You can work within the TinyDB directory, changing its structure if needed
  ▶ (Briefly) comment the code: every class, field, method, design choice
  ▶ Include a Makefile and instructions on how to build/run it
  ▶ Give examples of the input queries for which you tested. How about unit tests (e.g. github.com/google/googletest)?
INFO

- 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, Oktober 30