Query Optimization: Exercise

Session 3

Bernhard Radke

November 3, 2017

Homework

Exercise 1

$$\sigma_{p_1}(R_1 \bowtie_{p_2} R_2) = \sigma_{p_1}(R_1) \bowtie_{p_2} R_2 \text{ if } \mathcal{F}(p_1) \subseteq \mathcal{A}(R_1)$$

$$\begin{aligned}
\text{Let } t \in \sigma_{p_1}(R_1 \bowtie_{p_2} R_2) \quad & \Leftrightarrow \quad t \in (R_1 \bowtie_{p_2} R_2) \text{ and } p_1 \text{ holds for } t \\
& \Leftrightarrow \quad \exists t_1 \in R_1, t_2 \in R_2 \text{ s.t. } t = t_1 \circ t_2 \wedge p_1(t) \wedge p_2(t) \\
& \overset{\mathcal{F}(p_1) \subseteq \mathcal{A}(R_1)}{\Leftrightarrow} \quad \exists t_1 \in R_1, t_2 \in R_2 \text{ s.t. } t = t_1 \circ t_2 \wedge p_1(t_1) \wedge p_2(t) \\
& \Leftrightarrow \quad \exists t_1 \in \sigma_{p_1}(R_1), t_2 \in R_2 \text{ s.t. } t = t_1 \circ t_2 \wedge p_2(t) \\
& \Leftrightarrow \quad t \in \sigma_{p_1}(R_1) \bowtie_{p_2} R_2
\end{aligned}$$

- $\sigma_{p_1}(R_1 \leftouterjoin_{p_2} R_2) = \sigma_{p_1}(R_1) \leftouterjoin_{p_2} R_2$ if $\mathcal{F}(p_1) \subseteq \mathcal{A}(R_1)$: similar
- $\sigma_{p_1}(R_1 \rightouterjoin_{p_2} R_2) \neq \sigma_{p_1}(R_1) \rightouterjoin_{p_2} R_2$ if $\mathcal{F}(p_1) \subseteq \mathcal{A}(R_1)$: Let $R_1 = \emptyset$
- $\sigma_{p_1}(R_1 \fullouterjoin_{p_2} R_2) \neq \sigma_{p_1}(R_1) \fullouterjoin_{p_2} R_2$ if $\mathcal{F}(p_1) \subseteq \mathcal{A}(R_1)$: Let $R_1 = \emptyset$

Exercise 2

We know $|R_1|$, $|R_2|$, domains of $R_1.x$, $R_2.y$, (that is, $|R_1.x|$, $|R_2.y|$), and whether x and y are keys or not.

The selectivity of $\sigma_{R_1.x=c}$ is...

- if x is the key: $\frac{1}{|R_1|}$
- if x is not the key: $\frac{1}{|R_1.x|}$

We know $|R_1|$, $|R_2|$, $|R_1.x|$, $|R_2.y|$, and whether x and y are keys or not.
First, the size of $R_1 \times R_2$ is $|R_1||R_2|$
The selectivity of $\bowtie_{R_1.x=R_2.y}$ is...

- if both x and y are the keys: $\frac{1}{max(|R_1|,|R_2|)}$
- if only x is the key: $\frac{1}{|R_1|}$
- if both x and y are not the keys: $\frac{1}{max(|R_1.x|,|R_2.y|)}$

Exercise 3

- $|R| = 1,000$ pages, $|S| = 100,000$ pages
- 1 page $=$ 50 tuples, 1 block $=$ 100 pages
- avg. access $=$ 10 ms, transfer speed $=$ 10,000 pages/sec
- Time for (blockwise) nested loops join?

Selectivity estimation

We know $|R_1|$, $max(R_1.x)$, $min(R_1.x)$, $R_1.x$ is arithmetic.

The selectivity of $\sigma_{R_1.x>c}$ is $\frac{max(R_1.x)-c}{max(R_1.x)-min(R_1.x)}$

The selectivity of $\sigma_{c_1<R_1.x<c_2}$ is $\frac{c_2-c_1}{max(R_1.x)-min(R_1.x)}$

Homework

- Give the query graphs for the two queries from Exercise 1
- Give an example query where the optimal join tree (using C_{out}) is bushy and contains a cross product
- based on the parser you built in exercise 1, implement canonical translation for tinydb

- 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 13