Query Optimization
Exercise Session 10

Bernhard Radke

January 30, 2017
Consider the following sequence of relations $R_1$, $R_2$, $R_3$, $R_4$ and their join graph:

![Join Graph](image)

Give a fully-parenthesized, optimal join-expression that abides by this order. Use $C_{out}$ as a cost function.
Order Preserving Joins: Baseline

Let's start off with a cost analysis of the left-deep tree:

\[
\begin{align*}
C_{\text{out}} &= 100 + 100 + 40 = 240
\end{align*}
\]
Order Preserving Joins: Baseline

Let’s start off with a cost analysis of the left-deep tree:

\[
C_{out} = 100
\]
Let's start off with a cost analysis of the left-deep tree:

\[ C_{out} = 100 + 100 \]
Order Preserving Joins: Baseline

Let's start off with a cost analysis of the left-deep tree:

\[
\begin{align*}
C_{out} &= 100 + 100 + 40 = 240
\end{align*}
\]
Order Preserving Joins: Initialization

OrderPreservingJoins($R = \{ R_1, \ldots, R_n \}, P$)

**Input:** a set of relations to be joined and a set of predicates

**Output:** fills $p, s, c, t$

for each $1 \leq i \leq n$

- $p[i, i] =$ predicates from $P$ applicable to $R_i$
- $P = P \setminus p[i, i]$
- $s[i, i] =$ statistics for $\sigma_{p[i,i]}(R_i)$
- $c[i, i] =$ costs for $\sigma_{p[i,i]}(R_i)$

<table>
<thead>
<tr>
<th>predicates $p$</th>
<th>statistics $s$</th>
<th>costs $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>
Order Preserving Joins: Constructing the Bushy Tree

01 for each $2 \leq l \leq 4$ ascending  
02 for each $1 \leq i \leq 5 - l$  
03 $j = i + l - 1$  
04 $p[i, j]$ = predicates from $P$ applicable to $R_i, \ldots, R_j$  
05 $P = P \setminus p[i, j]$  
06 $s[i, j]$ = statistics derived from $s[i, j - 1]$ and $s[j, j]$ including $p[i, j]$  
07 $c[i, j] = \infty$  
08 for each $i \leq k < j$  
10 $q = c[i, k] + c[k + 1, j]$ + costs for $s[i, k]$ and $s[k + 1, j]$ and $p[i, j]$  
11 if $q < c[i, j]$  
12 $c[i, j] = q$  
13 $t[i, j] = k$

<table>
<thead>
<tr>
<th>predicates $p$</th>
<th>statistics $s$</th>
<th>costs $c$</th>
<th>split points $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>\emptyset</td>
<td>200</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>\emptyset</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>\emptyset</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\emptyset</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>line = l = i = j = k = q =</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Order Preserving Joins: Constructing the Bushy Tree

01 for each $2 \leq l \leq 4$ ascending (in text: $2 \leq l \leq n$)
02 for each $1 \leq i \leq 5 - l$ (in text: $1 \leq i \leq n - l + 1$)
03 \hspace{1em} j = i + l - 1
04 \hspace{1em} p[i, j] = \text{predicates from } P \text{ applicable to } R_i, \ldots, R_j
05 \hspace{1em} P = P \setminus p[i, j]
06 \hspace{1em} s[i, j] = \text{statistics derived from } s[i, j - 1] \text{ and } s[j, j] \text{ including } p[i, j]
07 \hspace{1em} c[i, j] = \infty
08 for each $i \leq k < j$
10 \hspace{1em} q = c[i, k] + c[k + 1, j] + \text{costs for } s[i, k] \text{ and } s[k + 1, j] \text{ and } p[i, j]
11 if $q < c[i, j]$
12 \hspace{1em} c[i, j] = q
13 \hspace{1em} t[i, j] = k

<table>
<thead>
<tr>
<th>predicates p</th>
<th>statistics s</th>
<th>costs c</th>
<th>split points t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>${p_{1,2}}$</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>100</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$\emptyset$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

line = 08
\hspace{1em} $l = 2$
\hspace{1em} $i = 1$
\hspace{1em} $j = 2$
\hspace{1em} $k =$
\hspace{1em} $q =$
Order Preserving Joins: Constructing the Bushy Tree

01 for each \(2 \leq l \leq 4\) ascending (in text: \(2 \leq l \leq n\))
02 for each \(1 \leq i \leq 5 - l\) (in text: \(1 \leq i \leq n - l + 1\))
03 \(j = i + l - 1\)
04 \(p[i, j] = \text{predicates from } P \text{ applicable to } R_i, \ldots, R_j\)
05 \(P = P \setminus p[i, j]\)
06 \(s[i, j] = \text{statistics derived from } s[i, j - 1] \text{ and } s[j, j] \text{ including } p[i, j]\)
07 \(c[i, j] = \infty\)
08 for each \(1 \leq i \leq k < j\)
10 \(q = c[i, k] + c[k + 1, j] + \text{costs for } s[i, k] \text{ and } s[k + 1, j] \text{ and } p[i, j]\)
11 if \(q < c[i, j]\)
12 \(c[i, j] = q\)
13 \(t[i, j] = k\)

<table>
<thead>
<tr>
<th>predicates (p)</th>
<th>statistics (s)</th>
<th>costs (c)</th>
<th>split points (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\emptyset)</td>
<td>200</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>{(p_1, 2)}</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>20</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

line = 13
\(l = 2\)
\(i = 1\)
\(j = 2\)
\(k = 1\)
\(q = 0 + 0 + 200 \cdot 1 \cdot \frac{1}{2} = 100\)
01 for each $2 \leq l \leq 4$ ascending (in text: $2 \leq l \leq n$)
02 for each $1 \leq i \leq 5 - l$ (in text: $1 \leq i \leq n - l + 1$)
03 $j = i + l - 1$
04 $p[i, j] =$ predicates from $P$ applicable to $R_i, \ldots, R_j$
05 $P = P \setminus p[i, j]$
06 $s[i, j] =$ statistics derived from $s[i, j - 1]$ and $s[j, j]$ including $p[i, j]$
07 $c[i, j] = \infty$
08 for each $i \leq k < j$
10 $q = c[i, k] + c[k + 1, j] + \text{costs for } s[i, k] \text{ and } s[k + 1, j] \text{ and } p[i, j]$
11 if $q < c[i, j]$
12 $c[i, j] = q$
13 $t[i, j] = k$

<table>
<thead>
<tr>
<th>predicates $p$</th>
<th>statistics $s$</th>
<th>costs $c$</th>
<th>split points $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>200 100</td>
<td>0 100</td>
<td>1</td>
</tr>
<tr>
<td>${p_1, 2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

line = 11

$l = 2$

$i = 2$

$j = 3$

$k = 2$

$q = 0 + 0 + 1 \cdot 1 \cdot 1 = 1$
Order Preserving Joins: Constructing the Bushy Tree

01 for each $2 \leq l \leq 4$ ascending (in text: $2 \leq l \leq n$)
02 for each $1 \leq i \leq 5 - l$ (in text: $1 \leq i \leq n - l + 1$)
03 $j = i + l - 1$
04 $p[i, j] =$ predicates from $P$ applicable to $R_i, \ldots, R_j$
05 $P = P \setminus p[i, j]$
06 $s[i, j] =$ statistics derived from $s[i, j - 1]$ and $s[j, j]$ including $p[i, j]$
07 $c[i, j] = \infty$
08 for each $i \leq k < j$
09 $q = c[i, k] + c[k + 1, j] + \text{costs for } s[i, k] \text{ and } s[k + 1, j] \text{ and } p[i, j]$
10 if $q < c[i, j]$
11 $c[i, j] = q$
12 $t[i, j] = k$

<table>
<thead>
<tr>
<th>predicates $p$</th>
<th>statistics $s$</th>
<th>costs $c$</th>
<th>split points $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>200</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>${p_{1,2}}$</td>
<td>100</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\emptyset$</td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

line = 13
$l = 2$
$i = 2$
$j = 3$
$k = 2$
$q = 1$
Order Preserving Joins: Constructing the Bushy Tree

01 for each \(2 \leq l \leq 4\) ascending (in text: \(2 \leq l \leq n\))  
02 for each \(1 \leq i \leq 5 - l\) (in text: \(1 \leq i \leq n - l + 1\))  
03 \(j = i + l - 1\)  
04 \(p[i, j] = \) predicates from \(P\) applicable to \(R_i, \ldots, R_j\)  
05 \(P = P \setminus p[i, j]\)  
06 \(s[i, j] = \) statistics derived from \(s[i, j - 1]\) and \(s[j, j]\) including \(p[i, j]\)  
07 \(c[i, j] = 0\)  
08 for each \(i \leq k < j\)  
09 \(q = c[i, k] + c[k + 1, j] + \) costs for \(s[i, k]\) and \(s[k + 1, j]\) and \(p[i, j]\)  
10 if \(q < c[i, j]\)  
11 \(c[i, j] = q\)  
12 \(t[i, j] = k\)

<table>
<thead>
<tr>
<th>predicates (p)</th>
<th>statistics (s)</th>
<th>costs (c)</th>
<th>split points (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\emptyset)</td>
<td>{(p_{1,2})}</td>
<td>(200)</td>
<td>(0)</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>{(p_{3,4})}</td>
<td>(1)</td>
<td>(0)</td>
</tr>
<tr>
<td>(\emptyset)</td>
<td>(\emptyset)</td>
<td>(2)</td>
<td>(\infty)</td>
</tr>
</tbody>
</table>

\(c[i, j] = q\)  
\(t[i, j] = k\)

\(\text{line} = 11\)  
\(l = 2\)  
\(i = 3\)  
\(j = 4\)  
\(k = 3\)  
\(q = 0 + 0 + 1 \cdot 20 \cdot \frac{1}{10} = 2\)
Order Preserving Joins: Constructing the Bushy Tree

01 for each $2 \leq l \leq 4$ ascending (in text: $2 \leq l \leq n$)  
02 for each $1 \leq i \leq 5 - l$ (in text: $1 \leq i \leq n - l + 1$)  
03 $j = i + l - 1$  
04 $p[i, j] =$ predicates from $P$ applicable to $R_i, \ldots, R_j$  
05 $P = P \setminus p[i, j]$  
06 $s[i, j] =$ statistics derived from $s[i, j - 1]$ and $s[j, j]$ including $p[i, j]$  
07 $c[i, j] = \infty$  
08 for each $i \leq k < j$  
10 $q = c[i, k] + c[k + 1, j] + $ costs for $s[i, k]$ and $s[k + 1, j]$ and $p[i, j]$  
11 if $q < c[i, j]$  
12 $c[i, j] = q$  
13 $t[i, j] = k$

<table>
<thead>
<tr>
<th>predicates $p$</th>
<th>statistics $s$</th>
<th>costs $c$</th>
<th>split points $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>200 100</td>
<td>0 100</td>
<td>1</td>
</tr>
<tr>
<td>${p_{1, 2}}$</td>
<td>1 1</td>
<td>0 1</td>
<td>2</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>1 2</td>
<td>0 2</td>
<td>3</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

line = 13  
$l = 2$  
i = 3  
j = 4  
k = 3  
q = 2
Order Preserving Joins: Calling extract-plan

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The values of $t$ are:

$T_1 = \text{ExtractPlanRec}(R, t, p, i, t[i,j])$

$T_2 = \text{ExtractPlanRec}(R, t, p, t[i,j] + 1, j)$

return $T_1 \Join_{p[i,j]} T_2$

else

return $\sigma_{p[i,j]} R_i$
Order Preserving Joins: extract-plan callstack

extract-subplan(\ldots, i=1, j=4)
  extract-subplan(\ldots, i=1, j=1)
  extract-subplan(\ldots, i=2, j=4)
    extract-subplan(\ldots, i=2, j=3)
      extract-subplan(\ldots, i=2, j=2)
      extract-subplan(\ldots, i=3, j=3)
    \textbf{return} (R_2 \bowtie_{\text{true}} R_3)
    extract-subplan(\ldots, i=4, j=4)
  \textbf{return} ((R_2 \bowtie_{\text{true}} R_3) \bowtie_{p_{3,4}} R_4)
\textbf{return} (R_1 \bowtie_{p_{1,2} \wedge p_{1,4}} ((R_2 \bowtie_{\text{true}} R_3) \bowtie_{p_{3,4}} R_4))

The total cost of this plan is $c[1, 4] = 43$. 
Submit exercises to radke@in.tum.de
Due February 6, 2017.