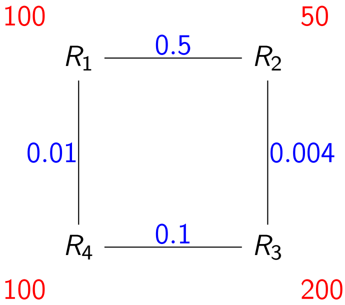
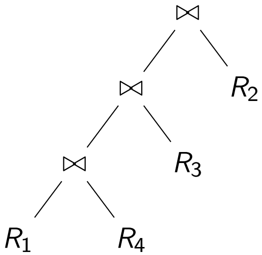
Query Optimization
Exercise Session 9

Andrey Gubichev

December 15, 2014

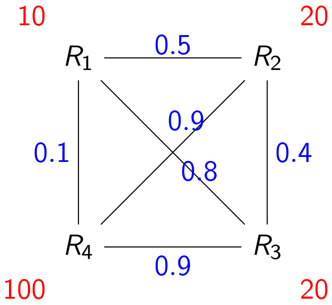
II example - 1 (Bui Nhat Nam, Nguyen Dinh Duy)





- Candidates: 3142, 4132, 3142, 4312, 3412

II example - 2



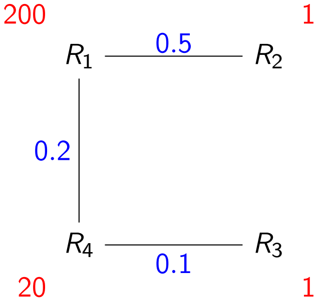
Start: $((R_1 R_2) R_3) R_4$: 6004

- assoc: $(R_1 (R_2 R_3)) R_4$: 6064
- assoc: $(R_1 R_2)(R_3 R_4)$: 7084
- l j ex: $((R_1 R_3) R_2) R_4$: 6084
- l j ex: $((R_1 R_2) R_4) R_3$: 6084

Optimal: $(R_1 R_4)(R_2 R_3)$: 5444

Order Preserving Joins: Example

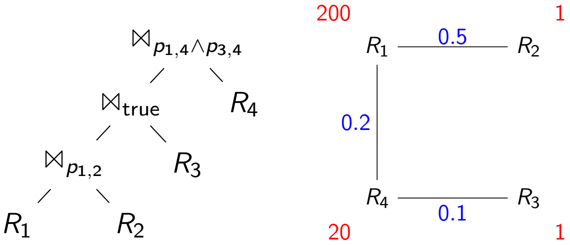
Consider the following *sequence* of relations R_1, R_2, R_3, R_4 and their join graph:



Give a fully-parenthesized, optimal join-expression that abides by this order. Use C_{out} as a cost function.

Order Preserving Joins: Baseline

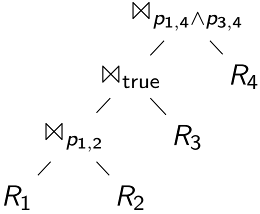
Let's start off with a cost analysis of the left-deep tree:

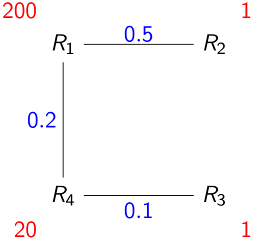


$$C_{out} =$$

Order Preserving Joins: Baseline

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

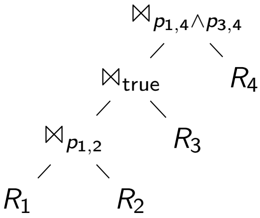


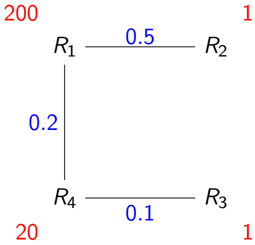


$$C_{out} = 100$$

Order Preserving Joins: Baseline

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

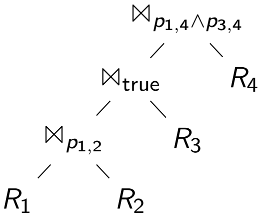


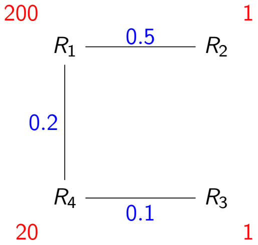


$$C_{out} = 100 + 100$$

Order Preserving Joins: Baseline

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





$$C_{out} = 100 + 100 + 40 = 240$$

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 $\sigma_{p[i,i]}(R_i)$
}

predicates p

$\emptyset$			
	$\emptyset$		
		$\emptyset$	
			$\emptyset$

statistics s

200			
	1		
		1	
			20

costs c

0			
	0		
		0	
			0

Order Preserving Joins: Constructing the Bushy Tree

```
01 for each 2 ≤ l ≤ 4 ascending (in text: 2 ≤ l ≤ n )
02   for each 1 ≤ i ≤ 5 − l (in text: 1 ≤ i ≤ n − l + 1)
03     j = i + l − 1
04     p[i, j]=predicates from P applicable to R_i, . . . , R_j
05     P = P \ 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]=∞
08     for each i ≤ 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
```

predicates p

∅			
	∅		
		∅	
			∅

statistics s

200			
	1		
		1	
			20

costs c

0			
	0		
		0	
			0

split points t

line =

l =

i =

j =

k =

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 $p[i,j]$=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 $c[i,j]$=∞
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

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$		
		$\emptyset$	
			$\emptyset$

statistics s

200	100		
	1		
		1	
			20

costs c

0	∞		
	0		
		0	
			0

split points t

line = 08
$l =$ 2
$i =$ 1
$j =$ 2
$k =$
$q =$

Order Preserving Joins: Constructing the Bushy Tree

```
01for each 2 ≤ l ≤ 4 ascending (in text: 2 ≤ l ≤ n )
02   for each 1 ≤ i ≤ 5 − l (in text: 1 ≤ i ≤ n − l + 1)
03     j = i + l − 1
04     p[i, j]=predicates from P applicable to R_i, . . . , R_j
05     P = P \ 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]=∞
08     for each i ≤ 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
```

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$		
		$\emptyset$	
			$\emptyset$

statistics s

200	100		
	1		
		1	
			20

costs c

0	100		
	0		
		0	
			0

split points t

	1		

line = 13

$l =$ 2

$i =$ 1

$j =$ 2

$k =$ 1

$q = 0 + 0 + 200 \cdot 1 \cdot \frac{1}{2} = 100$

Order Preserving Joins: Constructing the Bushy Tree

```
01 for each 2 ≤ l ≤ 4 ascending (in text: 2 ≤ l ≤ n )
02   for each 1 ≤ i ≤ 5 − l (in text: 1 ≤ i ≤ n − l + 1)
03     j = i + l − 1
04     p[i, j]=predicates from P applicable to R_i, . . . , R_j
05     P = P \ 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]=∞
08     for each i ≤ 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
```

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$	$\emptyset$	
		$\emptyset$	
			$\emptyset$

statistics s

200	100		
	1	1	
		1	
			20

costs c

0	100		
	0	∞	
		0	
			0

split points t

	1		

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     $c[i,j]$=∞
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

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$	$\emptyset$	
		$\emptyset$	
			$\emptyset$

statistics s

200	100		
	1	1	
		1	
			20

costs c

0	100		
	0	1	
		0	
			0

split points t

	1		
		2	

line = 13
l = 2
i = 2
j = 3
k = 2
q = 1

Order Preserving Joins: Constructing the Bushy Tree

01 **for each** $2 \le l \le 4$ **ascending** (in text: $2 \le l \le n$)
02   **for each** $1 \le i \le 5 - l$ (in text: $1 \le i \le n - l + 1$)
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     $c[i,j]$=∞
08   for each $i \le k < j$
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

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$	$\emptyset$	
		$\emptyset$	$\{p_{3,4}\}$
			$\emptyset$

statistics s

200	100		
	1	1	
		1	2
			20

costs c

0	100		
	0	1	
		0	∞
			0

split points t

	1		
		2	

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 ≤ l ≤ 4 ascending (in text: 2 ≤ l ≤ n )
02   for each 1 ≤ i ≤ 5 − l (in text: 1 ≤ i ≤ n − l + 1)
03     j = i + l − 1
04     p[i, j]=predicates from P applicable to R_i, . . . , R_j
05     P = P \ 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]=∞
08     for each i ≤ 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
```

predicates p

$\emptyset$	$\{p_{1,2}\}$		
	$\emptyset$	$\emptyset$	
		$\emptyset$	$\{p_{3,4}\}$
			$\emptyset$

statistics s

200	100		
	1	1	
		1	2
			20

costs c

0	100		
	0	1	
		0	2
			0

split points t

	1		
		2	
			3

line = 13

$l =$ 2

$i =$ 3

$j =$ 4

$k =$ 3

$q =$ 2

Order Preserving Joins: Calling `extract-plan`

The values of t are:

i/j	1	2	3	4
1		1	1	1
2			2	3
3				3
4				

ExtractPlan($R = \{R_1, \ldots, R_n\}$,t,p)
Input: a set of relations, arrays t and p
Output: a bushy join tree
return ExtractPlanRec(R,t,p,1,n)

ExtractPlanRec($R = \{R_1, \ldots, R_n\}$,t,p,i,j)
if $i < j$
$T_1 =$ExtractPlanRec(R,t,p,i,$t[i,j]$)
$T_2 =$ExtractPlanRec(R,t,p,$t[i,j]+1,j$)
return $T_1 \bowtie^L_{p[i,j]} T_2$
else
return $\sigma_{p[i,j]} R_i$

Order Preserving Joins: `extract-plan` callstack

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

The total cost of this plan is $c[1,4] = 43$.

Info

- Submit exercises to Andrey.Gubichev@in.tum.de
- Due December 22, 2014.