CS 186 Discussion 9: Query Optimization and Database Design

1 Selectivity Estimation

2 Query Optimization

Consider the relations $R(a, b)$, $S(b, c)$, and $T(c, d)$ with clustered indexes on $R.a$, $S.b$, and $T.c$ and with unclustered indexes on $S.c$ and $T.d$. All indexes have index keys in the range $[1, 100]$. We want to optimize the following query:

```
SELECT *
 FROM  R, S, T
WHERE  R.b = S.b AND S.c = T.c
  AND  R.a <= 50
  AND  (T.c <= 50 OR T.d <= 20)
```

In the first pass of the Sellinger query optimization algorithm, we compute the minimum cost access method for every (relation, interesting order) pair. Complete the following table which computes this. Let $[R]$, $[S]$, and $[T]$ be the number of *pages* in R, S, and T. Similarly, let $|R|$, $|S|$, and $|T|$ be the number of *tuples* in R, S, and T.

Relation	Access Method	Interesting Order	I/O Cost	Output Size	Retained
R	Full table scan	None	$[R]$	$0.5[R]$	No
	Index scan on a	(a)	$2 + 0.5 * [R]$		Yes
S	Full table scan	None	$[S]$	$[S]$	Yes
	Index scan on b	(b)	$2 + [S]$		Yes
	Index scan on c	(c)	$2 + \|S\|$		Yes
T	Full table scan	None	$[T]$	$0.6[T]$	No
	Index scan on c	(c)	$2 + 0.5[T]$		Yes
	Index scan on d	(d)	$2 + 0.2\|T\|$		Yes

In the second pass of the Sellinger query optimization algorithm, we consider each (relation, interesting order) pair in turn. For each, we compute the cost of joining every other relation into it. In the end, we retain the minimum cost join for each (relations, interesting order) pair. Complete the following table which computes *part* of the second pass. Let B be the number of buffer pages, and let $SC(x) = 2x\lceil 1 + \log_{B-1}(\frac{x}{B})\rceil$. Note that we do not consider joining R and T because we favor joins over cross products.

Left Relations	Left Interesting Order	Right Relation	Join	Interesting Order	I/O Cost	Output Size
R	(a)	S	PNLJ	None	$0.5[R] + \left\lceil \frac{0.5[R]}{B-2} \right\rceil [S]$	$\frac{0.5[R][S]}{100}$
			SMJ	(b)	$\text{SC}(0.5[R]) + \text{SC}([S]) + 0.5[R][S]$	
S	None	R	PNLJ	None	$[S] + \left\lceil \frac{[S]}{B-2} \right\rceil 0.5[R]$	$\frac{0.5[R][S]}{100}$
			SMJ	(b)	$\text{SC}([S]) + \text{SC}(0.5[R]) + [S]0.5[R]$	
S	None	T	PNLJ	None	$[S] + \left\lceil \frac{[S]}{B-2} \right\rceil 0.6[T]$	$\frac{0.6[S][T]}{100}$
			SMJ	(c)	$\text{SC}([S]) + \text{SC}(0.6[T]) + [S]0.6[T]$	
S	(b)	R	PNLJ	None	$[S] + \left\lceil \frac{[S]}{B-2} \right\rceil 0.5[R]$	$\frac{0.5[R][S]}{100}$
			SMJ	(b)	$SC(0.5[R]) + 0.5[R][S]$	
S	(b)	T	PNLJ	None	$[S] + \left\lceil \frac{[S]}{B-2} \right\rceil 0.6[T]$	$\frac{0.6[S][T]}{100}$
			SMJ	(c, b)	$\text{SC}([S]) + \text{SC}(0.6[T]) + [S]0.6[T]$	

3 ER Diagrams