Bill Yang Due: 3/30/20

Homework 7

CS386D Database Systems Instructor: Daniel Miranker

1 Part A

15.3.2
$B(S) + (B(S)B(R))/(M-1) = 10000 + \lceil(10000/(1000-1)\rceil \times 10000) =$
120000 disk I/O.

15.3.3 a

$$\begin{aligned} 100000 &= 10000 + \lceil 10000/(M-1) \rceil (10000) \\ M &= 1112.11 \end{aligned}$$

$M \geq 1113$

15.4.2 a, b, c
a) $3(B(S) + B(R)) = 3(10000 + 10000) = 60000$ I/Os.
b) $5(B(S) + B(R)) = 5(10000 + 10000) = 100000$ I/Os.
c) $3(B(S) + B(R)) = 3(10000 + 10000) = 60000$ I/Os.

15.4.3
For 2PPMS, we can decrease the size of the sublists such that we do have M sublists, eg. we only use pM buffers to sort sublists and sublists are pM long where $0 < p < 1$. Then when we sort a sublist, we can place the first block of the sublist in our unused blocks, the $(1-p)M$ blocks not used for sorting, until these are full. Then the algorithm would be more or less the same, where we must write out entire lists and reread them to merge sort. Thus we save $2((1-p)M)$ disk I/Os.

2 Part B

16.4.1 a, d, i
a)
$\frac{100\times200\times300\times400}{60\times100\times50} = 8000$ tuples

d) Selection gives us $T(\sigma_{c=20}(Y)) = 6$.
$\frac{6\times400}{50} = 48$ tuples.

i) $\frac{200\times300}{3} = 20000$ tuples.

3 Part C

1. Text problem 16.6.1a, (which concerns precisely part a of 16.4.1). In addition to answering the text question, show the entries in the dynamic programming matrix.

W	X	Y	Z

	W	X	Y	Z
Size	100	200	300	400
Cost	0	0	0	0
Best Plan	W	X	Y	Z

WX	WY	WZ	XY	XZ	YZ
W	X	Y	Z		

	(W, X)	(W, Y)	(W, Z)	(X, Y)	(X, Z)	(Y, Z)
Size	333	30000	40000	600	80000	2400
Cost	0	0	0	0	0	0
Best Plan	$W \bowtie X$	$W \bowtie Y$	$W \bowtie Z$	$X \bowtie Y$	$X \bowtie Z$	$Y \bowtie Z$

WXY	WXZ	WYZ	XYZ		
WX	WY	WZ	XY	XZ	YZ
W	X	Y	Z		

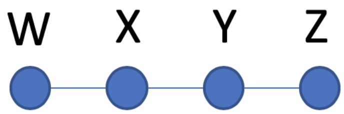
	(W, X, Y)	(W, X, Z)	(W, Y, Z)	(X, Y, Z)
Size	1000	133333	240000	4800
Cost	333	333	2400	600
Best Plan	$Y \bowtie (W \bowtie X)$	$(W \bowtie X) \bowtie Z$	$W \bowtie (Y \bowtie Z)$	$Z \bowtie (X \bowtie Y)$

WXYZ					
WXY	WXZ	WYZ	XYZ		
WX	WY	WZ	XY	XZ	YZ
W	X	Y	Z		

Plan	Cost
$Z \bowtie (Y \bowtie (W \bowtie X))$	1333
$Y \bowtie ((W \bowtie X) \bowtie Z)$	133666
$X \bowtie (W \bowtie (Y \bowtie Z))$	242400
$W \bowtie (Z \bowtie (X \bowtie Y))$	5400
$(W \bowtie X) \bowtie (Y \bowtie Z)$	2733
$(W \bowtie Y) \bowtie (X \bowtie Z)$	110000
$(X \bowtie Y) \bowtie (W \bowtie Z)$	40600

$Z \bowtie (Y \bowtie (W \bowtie X))$ is the best plan

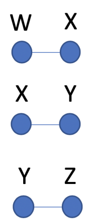
2. Repeat 16.6.1a, this time.
a) draw the query graph.



b) leveraging the query graph to avoid Cartesian products, limit the amount of work and number of entries you create in the dynamic programming matrix.

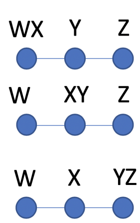
W	X	Y	Z

	W	X	Y	Z
Size	100	200	300	400
Cost	0	0	0	0
Best Plan	W	X	Y	Z



WX	~~WY~~	~~WZ~~	XY	~~XZ~~	YZ
W	X	Y	Z		

	(W, X)	(X, Y)	(Y, Z)
Size	333	600	2400
Cost	0	0	0
Best Plan	W $\bowtie X$	X $\bowtie Y$	Y $\bowtie Z$



WXY	~~WXZ~~	~~WYZ~~	XYZ		
WX	~~WY~~	~~WZ~~	XY	~~XZ~~	YZ
W	X	Y	Z		

	(W, X, Y)	(X, Y, Z)
Size	1000	4800
Cost	333	600
Best Plan	Y $\bowtie (W \bowtie X)$	Z $\bowtie (X \bowtie Y)$

WXYZ					
WXY	~~WXZ~~	~~WYZ~~	XYZ		
WX	~~WY~~	~~WZ~~	XY	~~XZ~~	YZ
W	X	Y	Z		

Plan	Cost
Z $\bowtie (Y \bowtie (W \bowtie X))$	1333
W $\bowtie (Z \bowtie (X \bowtie Y))$	5400
(W$\bowtie X) \bowtie (Y \bowtie Z)$	2733