<u>CS386D Database Systems</u>
<u>HW7 Solutions</u>

<u>Part A</u> (20 pts)

15.4.2

a) 3(B(R) + (B(S) = 3(10,000 + 10,000) = 60,000 disk I/O
b) b)5(B(R) + (B(S) = 5(10,000 + 10,000) = 100,000 disk I/O
c) c)3(B(R) + (B(S) = 3(10,000 + 10,000) = 60,000 disk I/O

15.4.3

The initial buffers of the last sorted sublist in the first pass can be kept in the extra buffers and thus, save on disk write I/Os.

<u>Part B</u> (30 pts)

16.4.1

a.

$T(W \bowtie X \bowtie Y \bowtie Z) = \frac{T(W) * T(X) * T(Y) * T(Z)}{max(V(W,b),V(X,b)) * max(V(X,c),V(Y,c)) * max(V(Y,d),V(Z,d))} = \frac{100*200*300*400}{60*100*50} = 8000$

d.

$T(\sigma_{c=20}(Y) \bowtie Z) = \frac{T(\sigma c{=}20(Y)) * T(Z)}{max(V(Y,d),V(Z,d))} = \frac{6 * T(Z)}{max(V(Y,d),V(Z,d))} = \frac{6 * 400}{50} = 48$

Note: Some students noted that the denominator term is 6 since there can be at most 6 distinct d's after the selection is applied. This has been marked wrong, but points weren't deducted. The answer is wrong as the textbook cost model is clearly an artificial set of rules for the purpose of creating a simple cost model for use in this class. That cost Model stipulates, pg 797 and on the course slides:
"**Preservation of Value Sets**. If we join a relation R with another relation, then an attribute A that is not a join attribute (i.e., not present in both relations) does not lose values from its set of possible values. More precisely, if A is an attribute of R but not of S, then V(R ix 5, A) = V(R, A).

And the subsequent discussion of preservation of value sets does not suggest any flexibility on that. That said, the reasoning to choose 6 as the maximum number of possible values is sound.

i.

T (X $\bowtie_{X.c<Y.c}$ Y) = $\frac{T(X) * T(Y)}{3}$ = $\frac{200*300}{3}$ = 20,000

<u>Part C</u> (40 pts)

1.

DP matrix:

W⋈X⋈Y⋈Z					
W⋈X⋈Y	W⋈X⋈Z	W⋈Y⋈Z	X⋈Y⋈Z		
W⋈X	W⋈Y	W⋈Z	X⋈Y	X⋈Z	Y⋈Z
W	X	Y	Z		

Singletons:

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

Pairs:

	W⋈X	W⋈Y	W⋈Z	X⋈Y	X⋈Z	Y⋈Z
Size	$\frac{100*200}{60}$ = 333	100*300 = 30k	100*400 = 40k	$\frac{200*300}{100}$ = 600	200×400 = 80k	$\frac{300*400}{50}$ = 2400
Cost	0	0	0	0	0	0
Best Plan	W⋈X	W⋈Y	W⋈Z	X⋈Y	X⋈Z	Y⋈Z

Triples:

	W⋈X⋈Y	W⋈X⋈Z	W⋈Y⋈Z	X⋈Y⋈Z
Size	$\frac{(1000/3)*300}{100}$ = 1000	$\frac{1000*400}{3}$ = 133,333	2400*100 = 240k	$\frac{600*400}{50}$ = 4800
Cost	333	333	2400	600
Best Plan	(W⋈X)⋈Y	(W⋈X)⋈Z	W⋈(Y⋈Z)	(X⋈Y)⋈Z

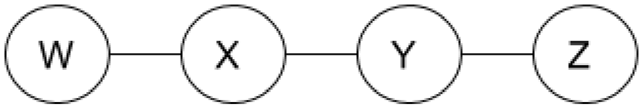
All 4 relations:

	W⋈X⋈Y⋈Z
Size	8000
Cost	1333
Best Plan	((W⋈X)⋈Y)⋈Z

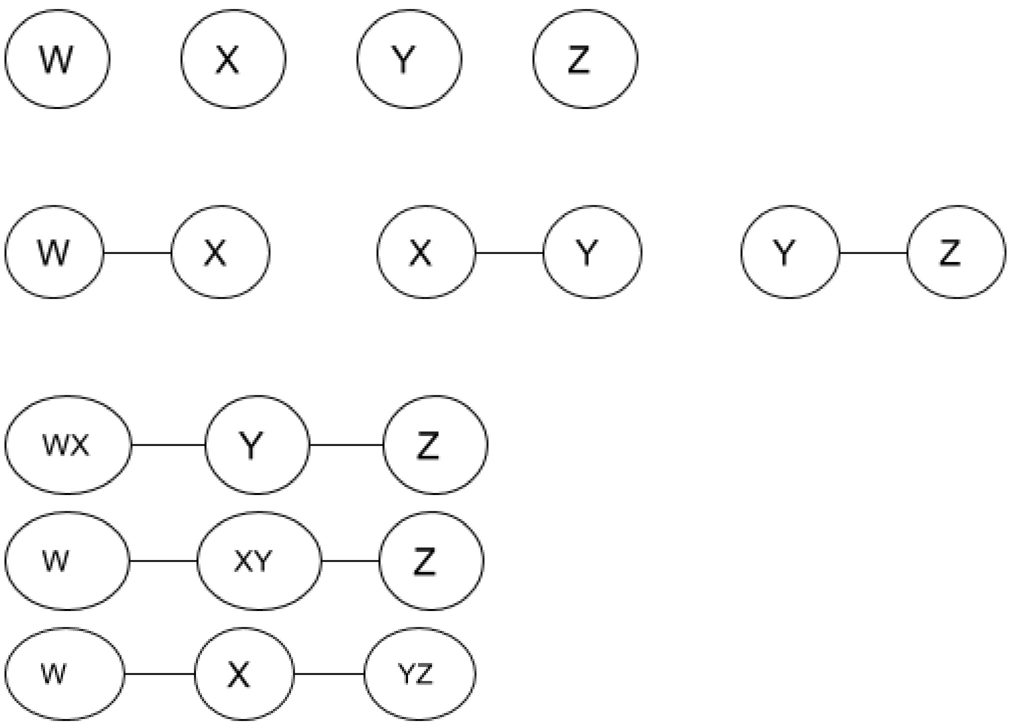
((W⋈X)⋈Y)⋈Z is the best plan.

2.

a) The query graph is as follows:



b)



So entries of the DP matrix reduce to:

W⋈X⋈Y⋈Z					
W⋈X⋈Y	~~W⋈X⋈Z~~	~~W⋈Y⋈Z~~	X⋈Y⋈Z		
W⋈X	~~W⋈Y~~	~~W⋈Z~~	X⋈Y	~~X⋈Z~~	Y⋈Z
W	X	Y	Z		

The size and cost of remaining plans are the same as before.