CS554

Homework 7.

Question 1

Dynamic Programming:

Relations	Size of Join (T(.))	Cost (Size intern. result)	Best join order
R	1000	0	R
S	100	0	S
T	100	0	T
U	1000	0	U
R, S	1000	0	S ⋈ R
R, T	100,000	0	T ⋈ R
R, U	10,000	0	R ⋈ U
S, T	500	0	S ⋈ T
S, U	100,000	0	S ⋈ U
T, U.	666.67	0	T ⋈ U.
R, S, T	5000	~~[illegible]~~ 500	(S ⋈ T) ⋈ R
R, S, U	10000	1000	(S ⋈ U) ⋈ R
S, T, U	3333.3	500	(S ⋈ T) ⋈ U.
R, T, U.	6666.67	666.67	(T ⋈ U) ⋈ R
R, S, T, U.	—	3833.33	((S ⋈ T) ⋈ U) ⋈ R

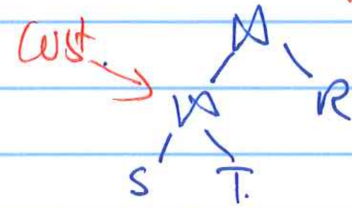
Best join order: ((S ⋈ T) ⋈ U) ⋈ R

Worked out details

R, S, T:

(1) $\{S,T\} \bowtie R$:

$T((S\bowtie T)\bowtie R)$ for next iteration.



Best: $(S \bowtie T) \bowtie R$

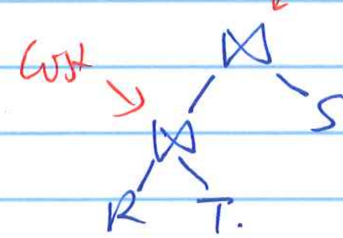
$$\text{Cost} = T(S \bowtie T) = \underline{500}.$$

$$T((S\bowtie T)\bowtie R) = \frac{T(S\bowtie T)\cdot T(R)}{\max(V(S\bowtie T, b), V(R,b))}.$$

$$= \frac{500 \cdot 1000}{\max(100, 100)} = 5000.$$

(2) $\{R,T\} \bowtie S$

$T((T\bowtie R)\bowtie S)$.



Best: $(T \bowtie R) \bowtie S$

$$\text{Cost} = T(T\bowtie R) = \underline{100000}$$

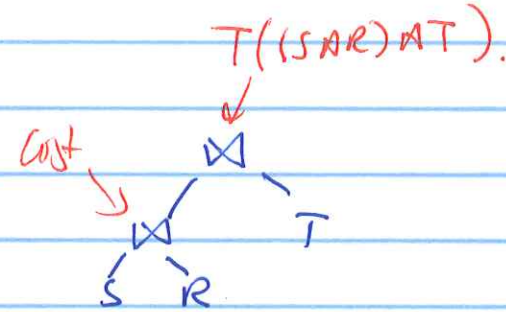
$$T((T\bowtie R)\bowtie S) = \frac{T(T\bowtie R)\cdot T(S)}{\max(V(T\bowtie R, b), V(S,b)) \cdot \max(V(T\bowtie R, c), V(S,c))}$$

$$= \frac{100000 \cdot 100}{100 \cdot 20} = 5000$$

must be same!!!

(3) $\{R, S\} \bowtie T$

Best: $(S \bowtie R) \bowtie T$



$Cost = T(S \bowtie R) = \underline{1000}$

$T((S \bowtie R) \bowtie T) = 5000$ (must be equal to prev. result)

Overall best: ~~cost = 1000~~.

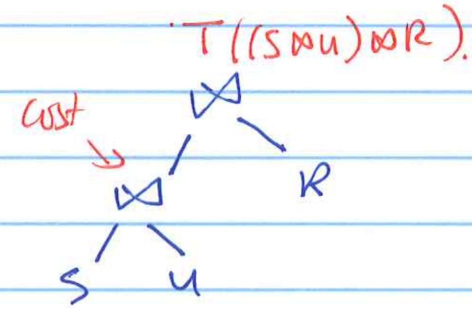
$\cancel{(S \bowtie R) \bowtie T}$.

$(S \bowtie T) \bowtie R$

$Cost = 500$.

R, S, U:

(1) $\{S, U\} \bowtie R$



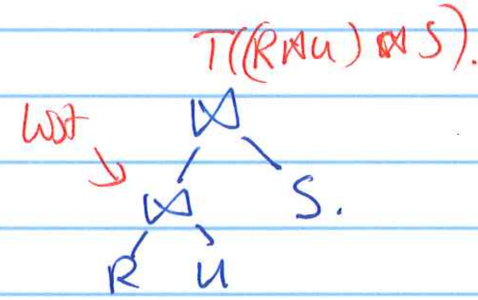
Best : $(S \bowtie U) \bowtie R$

$\text{Cost} = 100{,}000$

$$T((S \bowtie U) \bowtie R) = \frac{T(S \bowtie U) \cdot T(R)}{\max(V(\ldots, a)) \cdot \max(V(\ldots b))}$$

$$= \frac{100000 \cdot 1000}{100 \cdot 100} = 10{,}000.$$

(2) $\{R, U\} \bowtie S$



Best : $(R \bowtie U) \bowtie S$

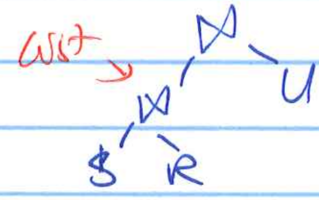
$\text{Cost} = 10{,}000.$

$$T((R \bowtie U) \bowtie S) = \frac{T(R \bowtie U) \cdot T(S)}{\max(V(R \bowtie U, b), V(S, b))}$$

$$= \frac{10000 \cdot 100}{100} = 10{,}000$$

(3) $\{R, S\} \bowtie U$

Best: $(S \bowtie R) \bowtie U$



cost = $\underline{1000}$

$$T((S \bowtie R) \bowtie U) = \frac{1000 \cdot 1000}{\max(100, 100)} =$$

$$= \frac{1000000}{100} = 10{,}000$$

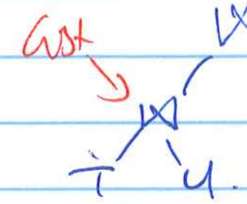
Overall best:

$(S \bowtie R) \bowtie U$

cost = 1000.

S, T, U :

$T((T \bowtie U) \bowtie S)$



(1) $\{T, U\} \bowtie S$

Best : $(T \bowtie U) \bowtie S$

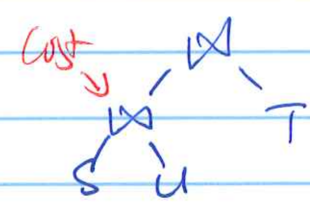
Cost = 666.67.

$$T((T \bowtie U) \bowtie S) = \frac{T(T \bowtie U) \cdot T(S)}{\max(V(T \bowtie U, c), V(S, c))}$$

$$= \frac{666.67 \cdot 100}{\max(20, 10)}$$

$$= \frac{666.67 \cdot 100}{20} = 3333.3$$

(2) $\{S, U\} \bowtie T.$



Best : $(S \bowtie U) \bowtie T$

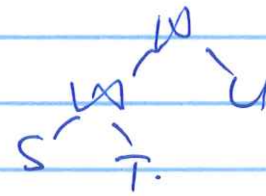
Cost = 100,000

$$T((S \bowtie U) \bowtie T) = \frac{100000 \cdot 100}{20 \cdot 150}$$

$$= 3333.3$$

3 $\{S,T\} \bowtie U$

Best: $(S \bowtie T) \bowtie U$



Cost = 500.

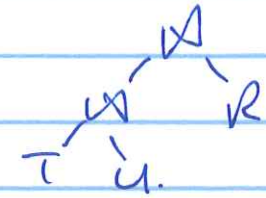
$T((S \bowtie T) \bowtie U) = 3333.3$

Overall best: $(S \bowtie T) \bowtie U$

Cost = 500.

R, T, U:

(1) $\{T, U\} \bowtie R$



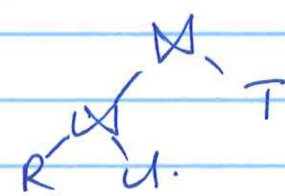
Best: $((T \bowtie U) \bowtie R)$

Cost = 666.67.

$$T((T \bowtie U) \bowtie R) = \frac{T(T \bowtie U) \cdot T(R)}{\max(V(T \bowtie U, a), V(R, a))}$$

$$= \frac{666.67 \cdot 1000}{\max(100, 100)} = 6666.67.$$

(2) $\{R, U\} \bowtie T$

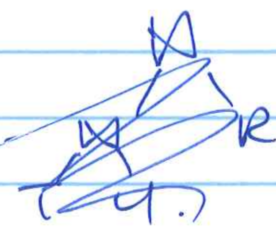


Best: $(R \bowtie U) \bowtie T$.

Cost = 10,000.

$T((R \bowtie U) \bowtie T) = 6666.67$

(3) ~~$\{T, U\} \bowtie R$~~



~~Best: $(T \bowtie U) \bowtie R$~~

Cost =

(3) $\{R, T.\} \bowtie U.$

Best: $(T \bowtie R) \bowtie U$

Cost = 100,000

$T((T \bowtie R) \bowtie u) = 6666.67.$

Overall best:

$(T \bowtie u) \bowtie R$

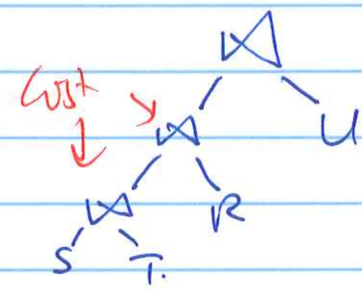
Cost = ~~33333.66666~~7.

= 666,67.

R, S, T, U :

1) $\{R, S, T\} \bowtie U$

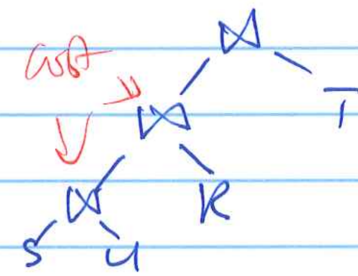
Best : $((S \bowtie T) \bowtie R) \bowtie U$



$$\text{Cost} = T(S \bowtie T) + T((S \bowtie T) \bowtie R)$$
$$= 500 + 5000$$
$$= 5500.$$

2) $\{R, S, U\} \bowtie T.$

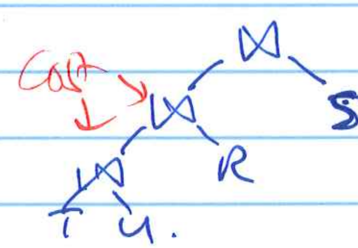
Best: $((S \bowtie U) \bowtie R) \bowtie T.$



$$\text{Cost} = T(S \bowtie U) + T((S \bowtie U) \bowtie R)$$
$$= 100{,}000 + 10000$$
$$= 110{,}000$$

3) $\{R, T, U\} \bowtie S.$

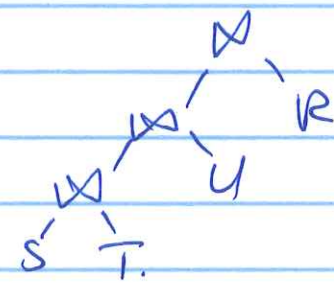
Best: $((T \bowtie U) \bowtie R) \bowtie S$



$$\text{Cost} = T(T \bowtie U) + T((T \bowtie U) \bowtie R)$$
$$= 666.67 + 6666.67 = 7333.33$$

(4) $\{S, T, U\} \bowtie R$

Best: $((S \bowtie T) \bowtie U) \bowtie R$



$$\text{Cost} = T(S \bowtie T) + T((S \bowtie T) \bowtie U)$$

$$= 500 + 3333.33$$

$$= \underline{3833.33}$$

Overall best:

$$((S \bowtie T) \bowtie U) \bowtie R$$

Cost: 3833.33.

Question 2

Greedy Algorithm:

Expression	Cost
$R \bowtie S$	$T(R \bowtie S) = \frac{T(R) \cdot T(S)}{\max(V(R,b), V(S,b))} = \frac{1000 \cdot 100}{100} = 1000$
$R \bowtie T$	$T(R \bowtie T) = T(R \times T) = 1000 \cdot 100 = 100000$
$R \bowtie U$	$T(R \bowtie U) = \frac{1000 \cdot 1000}{100} = 10{,}000$
$S \bowtie T$	$T(S \bowtie T) = \frac{100 \cdot 100}{20} = 500$
$S \bowtie U$	$T(S \bowtie U) = \frac{100 \cdot 1000}{1} = 100{,}000$ $= T(S \times U)$.
$T \bowtie U$	$T(T \bowtie U) = \frac{100 \cdot 1000}{150} = 666.$

Best: $S \bowtie T$.

$T(S \bowtie T) = 500$

$V(S \bowtie T, b) = 100$

$V(S \bowtie T, d) = 100$

Next join:

Expression	Cost
$(S \bowtie T) \bowtie R$	$T((S \bowtie T) \bowtie R) = \frac{500 \cdot 1000}{\max(100, 100)} = \frac{500 \cdot 1000}{100} = 5000$
$(S \bowtie T) \bowtie U$	$T((S \bowtie T) \bowtie U) = \frac{100 \cdot 1000}{\max(100, 150)} = \frac{100 \cdot 1000}{150} = \underline{3333}$

Best: $(S \bowtie T) \bowtie U)$

Only R left.

Solution. $((S \bowtie T) \bowtie U) \bowtie R$.

Question 3

$B(R)=1000$	$V(R,a)=20$
$T(R)=5000$	$V(R,b)=1000$
	$V(R,c)=5000$
	$V(R,d)=500$

(1) $\boxed{\sigma_{a=1 \wedge b=2 \wedge d=3}(R)}$

① choice 1: use index on a:

filter: $b=2 \wedge d=3$

index scan $a=1$.

$$\#\text{blks accessed} \approx \frac{1}{V(R,a)} \cdot B(R)$$

$$= \frac{1}{20} \cdot 1000 = \underline{\underline{500}}.$$

② Choice 2:

filter: $a=1 \wedge d=3$.

index scan $b=2$

$$\#\text{blks accessed} \approx \frac{1}{V(R,b)} \cdot T(R)$$

$$= \frac{1}{1000} \cdot 5000 = \underline{\underline{5}}$$

(3) Choice 3:

filter: $a=1 \wedge b=2$

index Scan $d=3$

$$\#\text{blks accessed} \approx \frac{1}{V(R,d)} \cdot T(R)$$

$$= \frac{1}{500} \cdot 5000 = \underline{\underline{10}}$$

<u>Best choice:</u>

filter: $a=1 \wedge d=3$

index Scan $b=2$.

R

<u>Cost: 5 blks.</u>

(2) $\boxed{\sigma_{a=1 \wedge b=2 \wedge c \geq 3}(R)}$

① choice 1:

filter: $b=2 \wedge c \geq 3$

Index Scan $a=1$

$\# \text{blks} = \underline{\underline{500}}$ (see prev. question)

② choice 2:

filter: $a=1 \wedge c \geq 3$

Index Scan $b=2$

$\# \text{blks} = \underline{\underline{5}}$ (see prev. question)

③ choice 3:

filter: $a=1 \wedge b=2$

Index Scan $c \geq 3$

estimate. (access $\frac{1}{3}$ of all tuples).

$\# \text{blks accessed} \approx \frac{1}{3} \cdot T(R)$

$= \frac{1}{3}(5000) = 1667$

<u>Best choice</u>:

filter: $a=1 \wedge c \geq 3$

Index Scan $b=2$

<u>cost: 5 blks</u>

(3) $\boxed{\sigma_{a=1 \wedge b \leq 2 \wedge c \geq 3}(R)}$

① choice 1:

filter: $b \leq 2 \wedge c \geq 3$

Index Scan $a=1$

blks accessed = $\underline{\underline{500}}$ (see prev. question)

② choice 2:

filter: $a=1 \wedge c \geq 3$

index scan $b \leq 2$

estimate (non-clustered access $\frac{1}{3}$ # tuples).

blks accessed $\approx \frac{1}{3} \cdot T(R)$

$= \frac{1}{3}(5000) = 1667.$

③ choice 3:

filte: $a=1 \wedge b \leq 2.$

Index Scan $c \geq 3$

blks accessed $\approx 1667.$

Best:

$b \leq 2 \wedge c \geq 3.$

Index Scan $a=1$

Cost: $\underline{\underline{500 \text{ blks}}}$

Question 4.

(1) $$\frac{T(R)}{V(R,x)} < \frac{T(R)}{V(R,y)}.$$

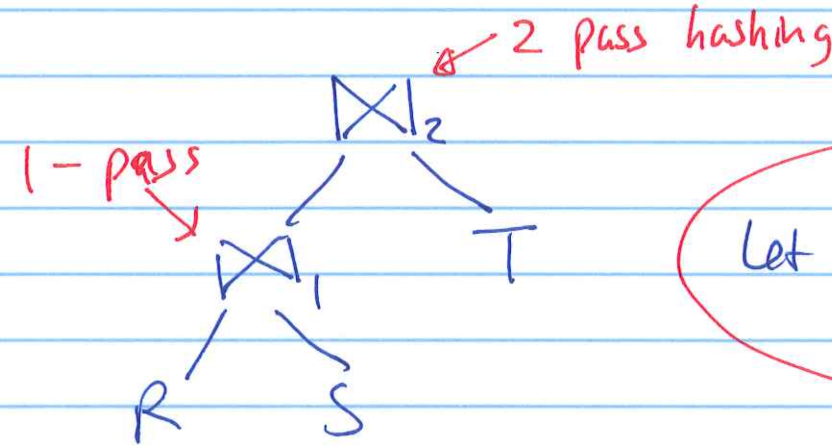
clusterng.

(2) $$\frac{T(R)}{V(R,x)} < \frac{B(R)}{V(R,y)}.$$

(3) $$\frac{T(R)}{V(R,x)} < \frac{1}{3} B(R)$$

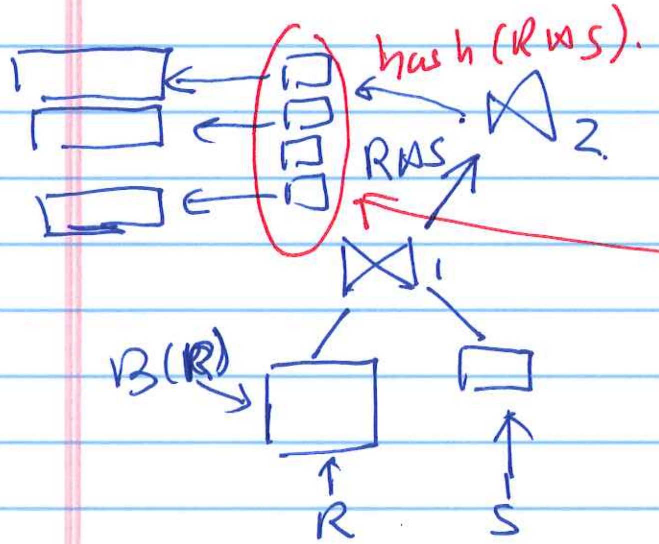
estimate # tuple that satisfy $y > c$.

Question 5



Let M = # available buffers.

Will be executed as follows:

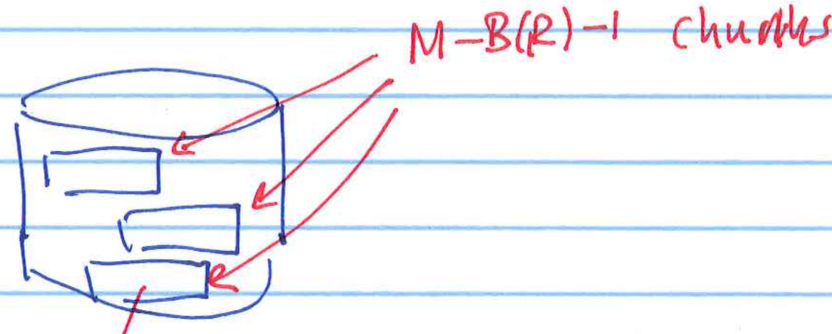


buckets $= M - (B(R)+1)$.

We will have: $M - B(R) - 1$ chunks.

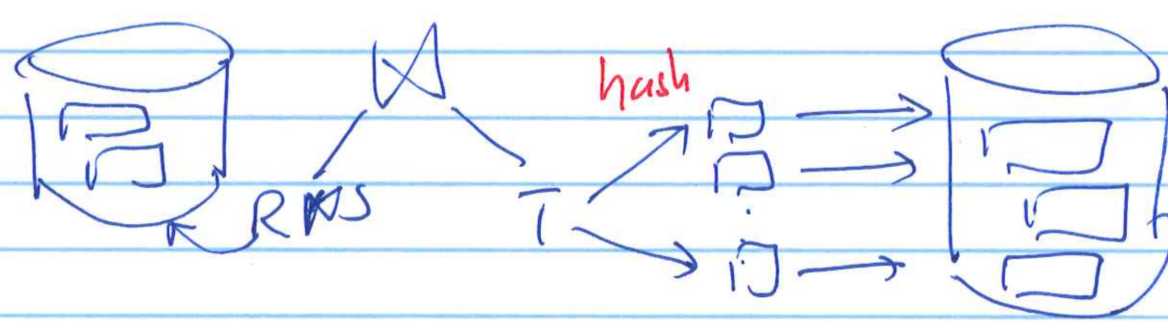
So each chunk is approx. $\dfrac{B(R \bowtie S)}{M - B(R) - 1}$ blks

- After pass 1 of the 2-pass hashing is complete:



Size of each chunk $\simeq \dfrac{B(R \bowtie S)}{M-B(R)-1}$.

- We hash T using M−B(R)−1 buckets:

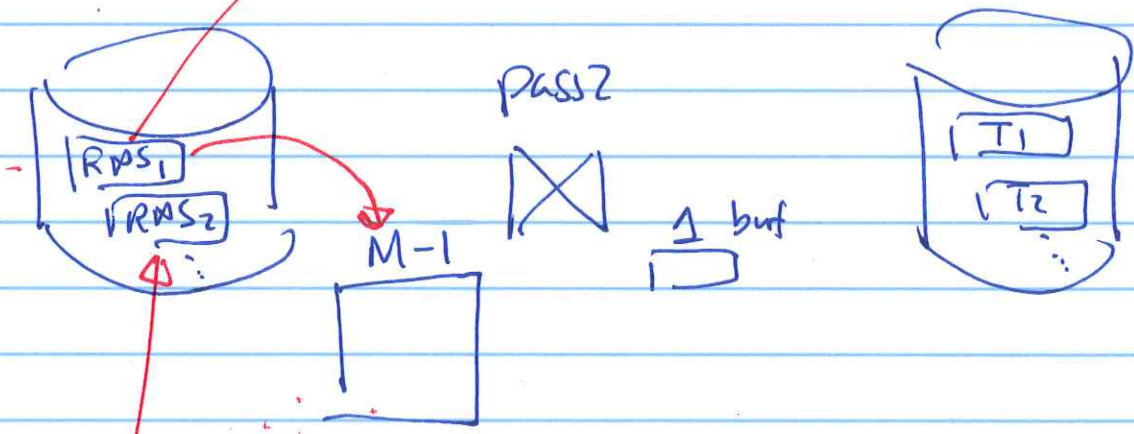


- Then join each (pair) of chunk of

$$(R \bowtie S)_i \bowtie T_i$$

Using all M buffers.

Set up: size of each chunk = $\frac{B(R \bowtie S)}{M - B(R) - 1}$.



$R \bowtie S_i$ must FIT in $M-1$ buffers.

Therefore:

$$\frac{B(R \bowtie S)}{M - B(R) - 1} \leq M - 1.$$

$$\Leftrightarrow \quad B(R \bowtie S) \leq (M-1)(M-1-B(R)).$$

$$\Leftrightarrow \quad (M-1)(M-1-B(R)) \geq B(R \bowtie S).$$

$$\Leftrightarrow \quad (M-1)^2 - B(R)(M-1) - B(R \bowtie S) \geq 0.$$

Solve this quadratic equation.

a, b, c formula: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

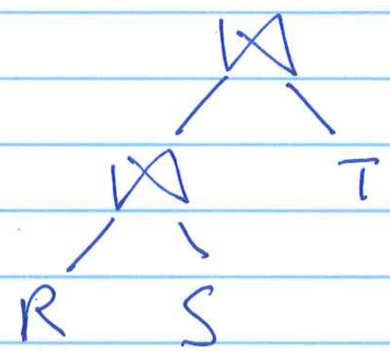
From Calculus:

$$M - 1 \geq \frac{+B(R) + \sqrt{B(R)^2 + 4 \cdot B(R \bowtie S)}}{2}$$

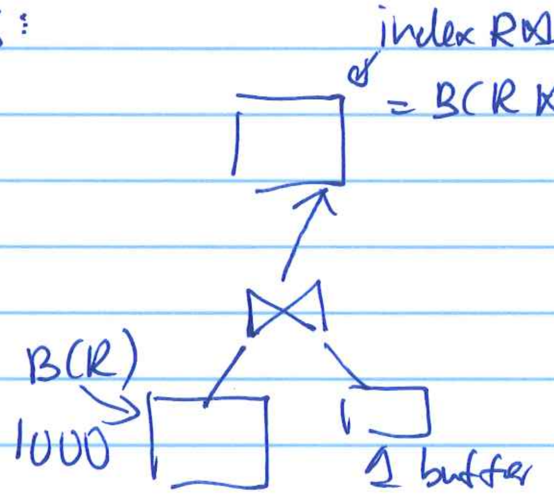
$$\Leftrightarrow \quad M \geq \frac{B(R) + \sqrt{B(R)^2 + 4B(R \bowtie S)}}{2} + 1.$$

Question 6

Min. # buffer to perform:

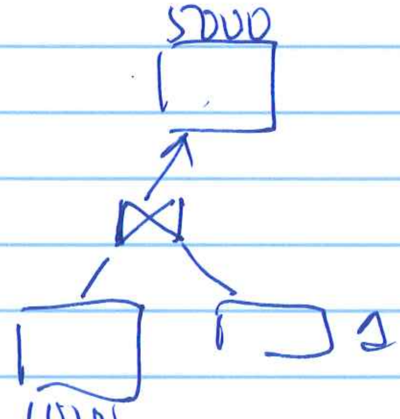


IS:



$= B(R \bowtie S) = \frac{1000 \cdot 1000}{200} = 5000.$

or:

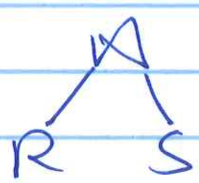


$\rightarrow MM = 5000 + 1000 + 1$

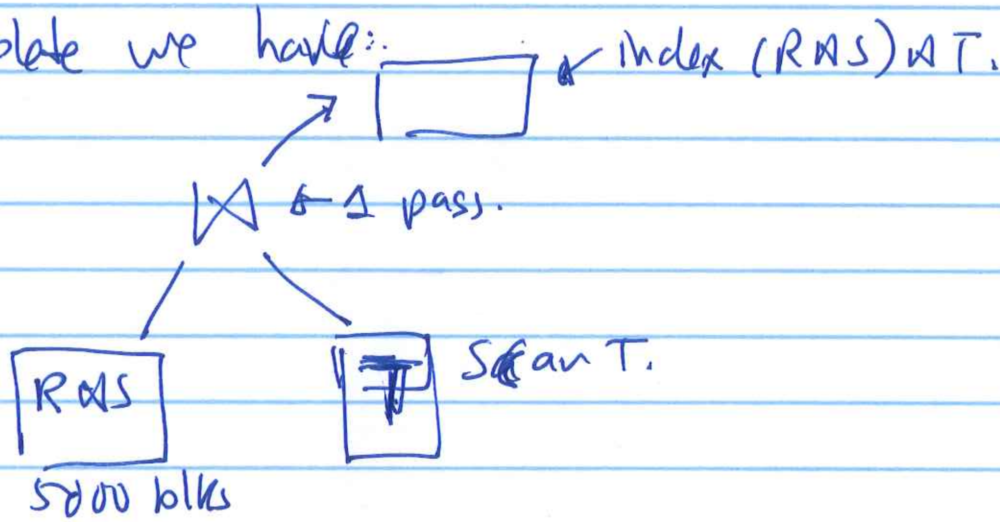
$= 6001.$

A

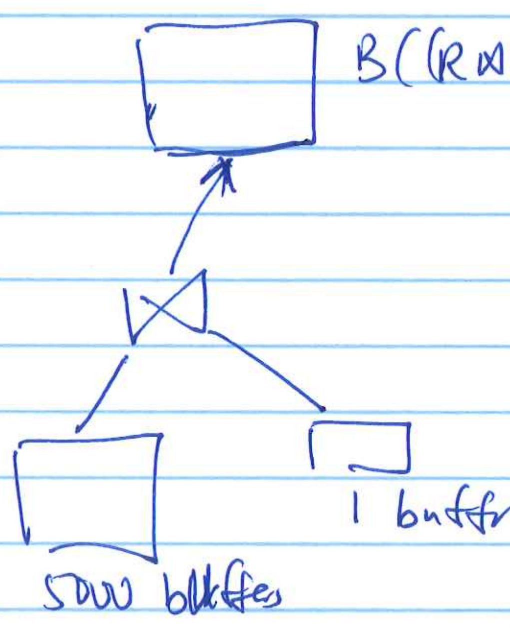
When



is complete we have:



Buffers used:

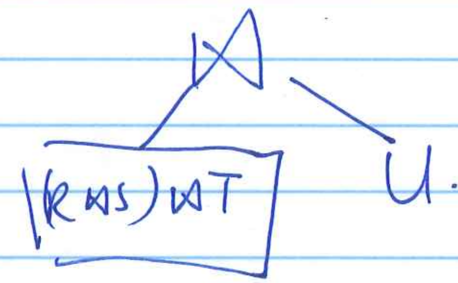


$$B((R\bowtie S)\bowtie T) = \frac{5000 \cdot 1000}{\max(200,100)} = 5000 \cdot \frac{1000}{200} = 25{,}000 \text{ blks.}$$

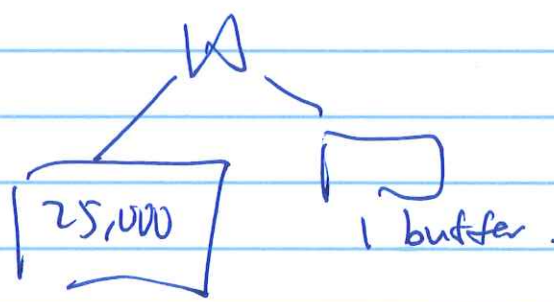
$$\# \text{ buffers} = 25000 + 5000 + 1 = 30001$$

B

When the 2^{nd} join is done, we reclaim the buffers for $R \bowtie S$ and T and use them for the last $\bowtie$:



Buffers used:



C.

buffers needed = 25001.

So the HIGHEST # buffers = 30001

(= max (6001, 30001, 25001)

= 30001