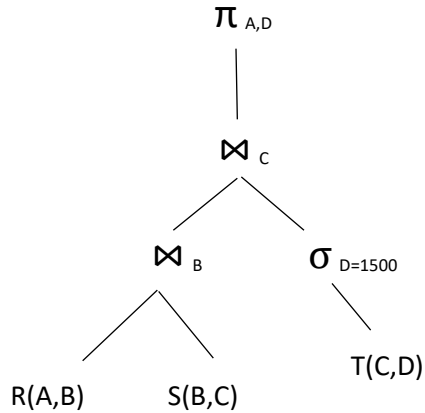


**CENG 352**  
**Database Management Systems**  
**Spring 2024**  
**Written Assignment 2**

**Q1.** Consider three relations  $R(A, B)$ ,  $S(B, C)$ ,  $T(C, D)$  in the query plan shown below.



Assume that

- all attributes have integer values.
- **every intermediate result is materialized** (i.e. written to disk).
- 42 memory pages are available.

Consider the following statistics:

$T(R) = 2000$	$B(R) = 200$
$T(S) = 10000$	$B(S) = 1000$
$T(T) = 200,000$	$B(T) = 20,000$

$V(R, B) = 200$	$\min = 1, \max = 2000$
$V(S, B) = 2000$	$\min = 1, \max = 4000$
$V(S, C) = 10000$	$\min = 1, \max = 10,000$
$V(T, C) = 2000$	$\min = 1, \max = 10,000$
$V(T, D) = 2000$	$\min = 1, \max = 500,000$

1. Estimate the number of tuples and blocks returned by  $\sigma_{D=1500}(T)$ .
2. Estimate the number of tuples and blocks returned by the join  $R \bowtie_B S$ . (Note that the number of attributes will increase by 50% in the joined tuples).
3. What is the estimated I/O cost of  $R \bowtie_B S$  using **block nested loop join**?
4. What is the estimated I/O cost of  $R \bowtie_B S$  using **sort-merge join**?
5. What is the cost of joining  $R$  and  $S$  using a **partitioned hash join**? You may assume the hash function works perfectly, creating partitions of equal size. Verify that the condition discussed

in class, on the buffer size and size of the input relations, holds in this case. Then state the cost of the join.

6. What is the estimated I/O cost of  $R \bowtie_B S$  if implemented by an **index nested loop join**? Assume that there is an unclustered index on  $S(B)$ .
7. What is the estimated total I/O cost of the given query plan if all three relations are accessed by file scan and both joins are implemented by **block nested loop join** algorithm? Remember that intermediate results are written and read from disk.

**Q2.** Consider three relations  $R(A, B)$ ,  $S(B, C)$ ,  $T(C, D)$  and the following statistics:

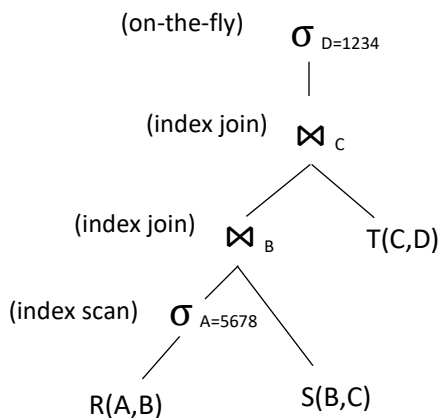
$B(R) = 2,000$	$B(S) = 1,000,000$	$B(T) = 1,000$
$T(R) = 200,000$	$T(S) = 10,000,000$	$T(T) = 10,000$
$V(R,A) = 10,000$	$V(S,B) = 200,000$	$V(T,D) = 100$
	$V(S,C) = 5,000$	

Assume the size of the main memory is  $M = 2000$  pages. Also assume that all indexes (except the  $S.B$  index) are unclustered and are stored in main memory (hence accessing the index nodes requires zero disk I/O's). The  $S.B$  is a clustered B+tree index.

a) Estimate the size of the answer of the following SQL query:

```
select *
from R, S, T
where R.B = S.B and S.C = T.C
and A = 5678 and D = 1234;
```

b) Assuming pipelined execution, estimate the total I/O cost of the following physical plan.



c) Assuming pipelined execution, estimate the total I/O cost of the following physical plan.

