# CENG 352
# Database Management Systems
# Spring 2023
# Written Assignment 2

Q1. (20 pts.) Consider the relational schema below:

Product(pid, pname, price)
Orders(oid, pid, cid, qty, date)
Customer(cid, cname, city)

For each of the following SQL queries below, write an equivalent logical query plan. (You should draw a relational algebra tree).

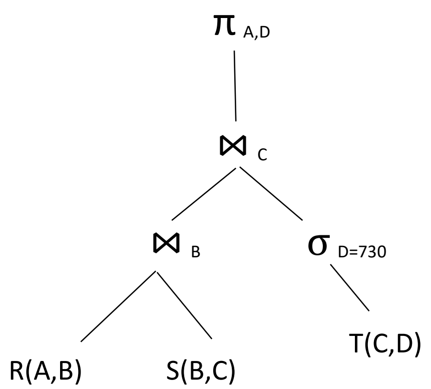
a)

```
SELECT p.pid, p.pname, SUM(p.price*o.qty) as revenue
FROM Product p, Orders o, Customer c
WHERE p.pid = o.pid and o.cid = c.cid and c.city = ’Ankara’
GROUP BY p.pid, p.pname
HAVING SUM(o.qty) > 1000;
```

b)

```
SELECT *
FROM Product p
WHERE NOT EXISTS (
      (SELECT c.cid
       FROM Customer c
       WHERE c.city = ‘Ankara’ )
      EXCEPT
      (SELECT o.cid
      FROM Order o
      WHERE o.pid = p.pid)
```

Q2. (50 pts.) 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).
- ~~12~~ 22 memory pages are available.

memory size is updated

Consider the following statistics:

| | |
|---|---|
| T(R) = 1000 | B(R) = 100 |
| T(S) = 5000 | B(S) = 200 |
| T(T) = 100,000 | B(T) = 10,000 |
| | |
| V(R,B) = 100 | min = 1, max = 1000 |
| V(S,B) = 1000 | min = 1, max = 2000 |
| V(S,C) = 5000 | min = 1, max = 2000 |
| V(T,C) = 1000 | min = 1, max = 10,000 |
| V(T,D) = 1000 | min = 1, max = 500,000 |

(7 pts. each)

a) Estimate the number of tuples and blocks returned by $\sigma_{D=730}(T)$.

b) 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).

Assume the size of joined tuples will be 50% more than the size of S tuples.

c) What is the estimated I/O cost of $R \bowtie_B S$ using **block nested**

d) What is the estimated I/O cost of $R \bowtie_B S$ using **sort-merge jo**

e) 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.

f) 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).

(8 pts.)

g) 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 **nested loop join** algorithm? Note that intermediate results are written and read from disk.

Q3. (30 pts.) Consider three relations R(A, B), S(B, C), T(C, D) and the following statistics:

| | |
|---|---|
| T(R) = 100000 | B(R) = 100 |
| T(S) = $6 \cdot 10^6$ | B(S) = 3000 |
| T(T) = $5 \cdot 10^4$ | B(T) = 40000 |

T(R)=10^5

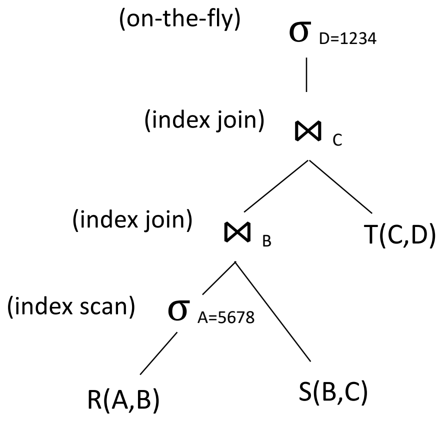
V(R,A) = $5 \cdot 10^4$
V(R,B) = V(S,B) = $3 \cdot 10^3$
V(S,C) = V(T,C) = $2 \cdot 10^4$
V(T,D) = $10^4$

Assume the following indices:

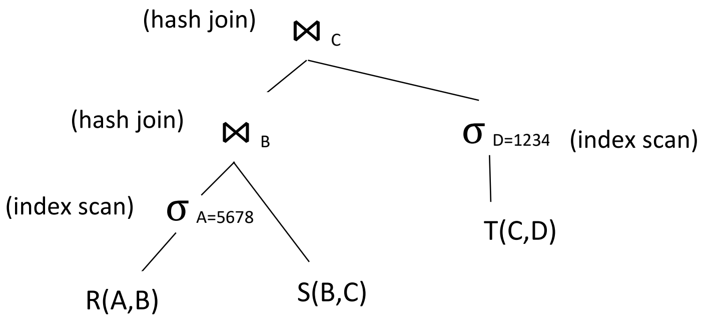
- Unclustered indexes on R.A and R.B
- Clustered index in S.B, unclustered index on S.C.
- Clustered index on T.C, unclustered index on T.D.

Estimate the I/O cost for the two physical plans given below.

Plan1



Plan2


(hash join)
⋈ C
(hash join)
⋈ B
σ D=1234
(index scan)
(index scan)
σ A=5678
T(C,D)
R(A,B)
S(B,C)