# HOMEWORK 9

**Problem 1.1**

| No. | Question | Blocking or Non-Blocking | Because/Reason |
|---|---|---|---|
| a | Duplicate elimination operator over unsorted relation R | Non-Blocking | Without consuming all input, this can produce tuples. |
| b | Grouping operator (group by column X) over a sorted relation R on column X | Non-Blocking | Without consuming all input, this can produce tuples. |
| c | Grouping operator (group by column X) over unsorted relation R | Blocking | Cannot produce any tuples to the output until it. processes all its inputs |
| d | Sorting operator (sort by column X) over unsorted relation R | Blocking | Cannot produce any tuples to the output until it. processes all its inputs |
| e | Sorting operator (sort by column X) and assume the operator can use a B-tree index that exists on R.X to read the tuples. | Non-Blocking | Without consuming all input, this can produce tuples with the use of an index. |
| f | Join of two relations R and S | Non-Blocking | Without consuming all input, this can produce tuples. |
| g | Bag Union of relations R and S | Non-Blocking | Without consuming all input, this can produce tuples. |

**Problem 1.2**

A.

Can be done in One pass: YES.

**Constraint**: the tuples should fit in M -1 blocks (199)

**I/O cost:** B(R)- 1000

Can it be done in two passes: YES, if it satisfies the

**constraint:** B(R)<= M^2

B(R)- 1000, M^2 = 200*200 i.e., 1000< 40000

**I/O cost:** 3B(R)-> 3000

B.

Can be done in One pass: **YES**.

**I/O cost:** B(R)- 1000

There are no 2 passes required.

C.

Can be done in One pass: **YES**.

**Constraint:** The tuples should fit in M -1 blocks (199)

**I/O cost:** B(R)- 1000

Can it be done in two passes: **YES**, if it satisfies the

**constraint:** B(R)<= M^2

B(R)- 1000, M^2 = 200*200 i.e., 1000< 40000

**I/O cost:** 3B(R)-> 3000

D.

Can be done in One pass: No, because the size of R is too big.

Can it be done in two passes: **Yes**.

E.

Can be done in One pass: **Yes**.

**Constraint:** The group must fit in M-1 buffer

**I/O cost:** B (R.X) = 70

There are no 2 passes required.

F.

Can be done in One pass: **Yes**.

Size of S < Memory buffer i.e., 150<200

I/O cost = B(R) + B(S) = 1000 + 150 = 1150

The size of S is small, and it easily fits into main memory, while R is much larger.

than the memory buffer, hence there is no need for two passes.

G.

Can be done in One pass: **Yes.**

**I/O cost:** B(R) + B(S) = 1000 + 150 = 1150

Can it be done in two passes: **No**

**Problem 2**

1.a opts for R2 as the outer relation.

1.b.

Total I/O cost = B(R2) + (B(R2)/ M-1) * B(R1)

= 200 + (200/(52-1)) * 1000

= 4122

1.c.

If R1 is chosen as the outer relation, then

Total I/O cost = B(R1) + (B(R1)/ M-1) * B(R2)

= 1000 + (1000/51)*200

= 4921

2.a

Total cost = 3(B(R1) + B(R2))

= 3(1000 + 200) = 3600

2.b

B(R1) + B(R2) <= M^2

The min. no of buffers needed for the cost to remain unchanged i.e., 35.

3.a

Total cost = 3(B(R1) + B(R2))

= 3(1000 + 200) = 3600

3.b

min(B(R1), B(R2)) <= M^2

B(R2) <=M^2

The min. no of buffers needed for the cost to remain unchanged i.e., 14.

4.

**#R2 is un- clustered**

Cost = T(R2) + T(R2) * (2 +B(R1)/ V (R2, Y))

= 2000 + 2000 * (2 + (1000/5))

= 2000 + 404000

= 406000

**#R2 is clustered**

Cost = B(R2) + T(R2) * (B(R1)/ V (R1, Y))

= 200 + 2000 * (1000/5)

= 200 + 400000

= 400200