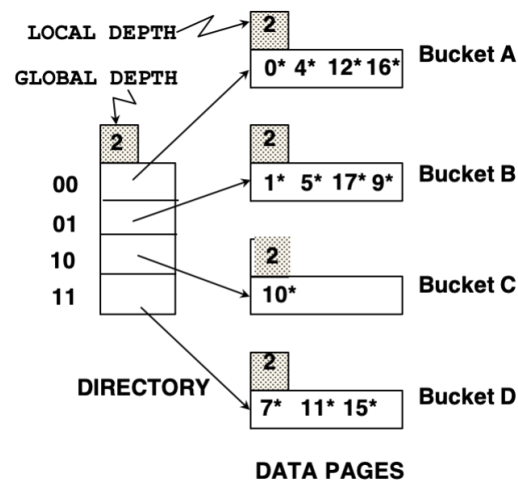# COMP 7640
# Written Assignment #3

Due: 11:59 PM April 24 (Friday), 2025

**Question 1**. Given an extendible hashing index as shown in the figure below, where each bucket can contain 4 data entries,

(1) If we are given a query with the condition key=11, how many I/Os are needed for this query? (**6 marks**)

(2) Draw the resulting index after inserting 2* and then 8*. (**16 marks**)



**Solution**:

(1) 2 I/Os. 1 I/O for reading Bucket D (**3 marks**). 1 I/O for reading the data page. (**3 marks**)

(2) Below see the updated index
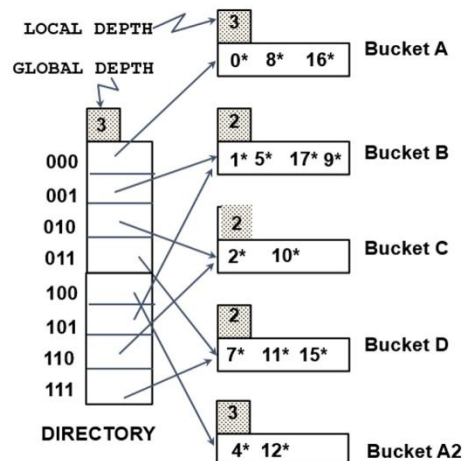
**Question 2**. We are given two relations S and E, which contain 10,000 records and 1,000 records, respectively. Each disk page can hold 100 S records and 5 E records, respectively. Consider the relational operation S $\bowtie_{S.sid=E.sid}$ E and estimate the I/O costs of the following access paths.
   (a) What is the lowest cost of using a page-oriented nested loop join?  (**10 marks**)
   (b) What is the lowest cost of using a simple-nested loop join? (**12 marks**)

   [**Remark**: There is no I/O cost for writing the result back to the disk.]

**Solution:**
#Pages for S: 10,000/100 = 100 pages (**4 marks**)
#Pages for E: 1,000/5 = 200 pages   (**4 marks**)

(a) 100 + 100 * 200 = 20,100 I/Os  (**4 marks**)
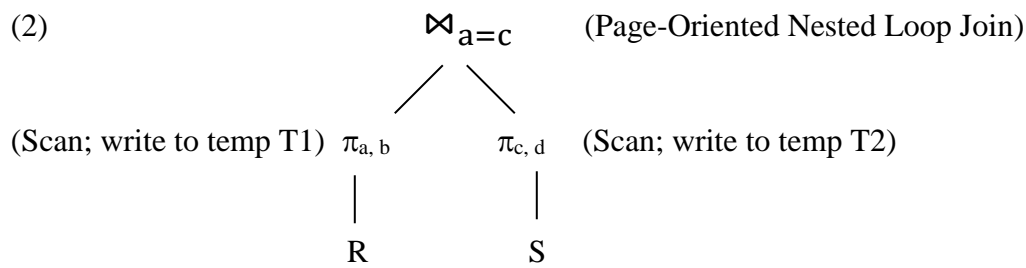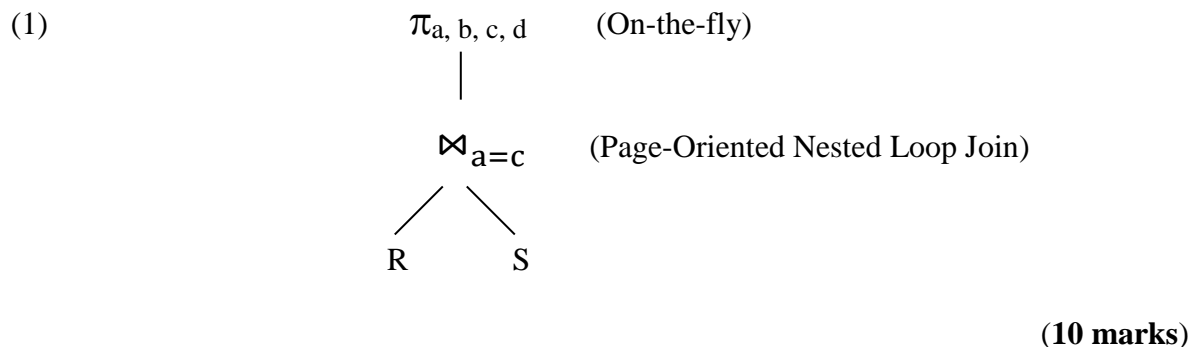(b) When S is the outer relation: 100 + 10,000 * 200 = 2,000,100 I/Os  (**4 marks**)
When E is the outer relation: 200 + 1,000 * 100 = 100,200 I/Os  (**4 marks**)
So the lowest I/O cost is 100,200 I/Os. (**2 marks**)

**Question 3**. Consider the query $\pi_{a,b,c,d}$ (R $\bowtie_{a=c}$ S) with the following information:
   a) Relation R(<u>a</u>, b, e) (a is the key) has 10 pages, and each R record needs 300 bytes;
   b) Relation S(<u>c</u>, d, f, g) (c is the key) has 100 pages, and each S record needs 500 bytes;
   c) Attributes a, b, c, and d take 100, 50, 100, and 150 bytes, respectively;
   d) The page size is 1024 bytes;
   e) Each S record joins with exactly one R record.

Estimate the total I/O cost for the following two query evaluation plans:
(1)                                    $\pi_{a, b, c, d}$      (On-the-fly)
                                         |

                                       $\bowtie_{a=c}$      (Page-Oriented Nested Loop Join)
                                        /      \
                                     R          S

                                                                    (**10 marks**)

(2)                                    $\bowtie_{a=c}$      (Page-Oriented Nested Loop Join)
                                        /      \
(Scan; write to temp T1)  $\pi_{a, b}$      $\pi_{c, d}$   (Scan; write to temp T2)
                                     |            |
                                     R            S

                                                                    (**22 marks**)

**Solution:**

(1) Cost of join =10 + 10 * 100 = 1010 page I/Os          (5 marks)
     Total cost = Cost of join = 1010 page I/Os               (5 marks)


(2)
    Cost of $\pi_{a, b}$ (R)
    = Cost to scan R + cost to write T1
    = 10 pages + size(T1)
    = 10 pages + (150 bytes/300 bytes) * 10 pages
    = 10 pages + 5 pages
    = 15 pages                                 (7 marks)

    Cost of $\pi_{c, d}$ (S)
    = Cost to scan S + cost to write T2
    = 100 pages + size(T2)
    = 100 pages + (250 bytes/500 bytes)*100 pages
    = 100 pages + 50 pages
    = 150 pages                               (7 marks)

Cost of join = 5 + (5*50) = 255 pages          (4 marks)
Total cost = Cost of projection + Cost of join = 165 + 255 = 420 I/Os          (4 marks)

**(The following is also correct :**
#records in R = 10*⌊1024/300⌋ = 30
#T1-records per page = ⌊1024/150⌋ = 6
Size(T1) = 30/6=5 pages

 Cost of $\pi_{a, b}$ (R)
    = Cost to scan R + cost to write T1
    = 10 pages + 5 pages
    = 15 pages                                 (7 marks)

#records in S = 100*⌊1024/500⌋ = 200
#T2-records per page = ⌊1024/250⌋ = 4
Size(T2) = 200/4 = 50 pages

    Cost of $\pi_{c, d}$ (S)
    = Cost to scan S + cost to write T2
    = 100 pages + 50 pages
    = 150 pages                               (7 marks)


Cost of join = 5 + (5*50) = 255 pages          (4 marks)
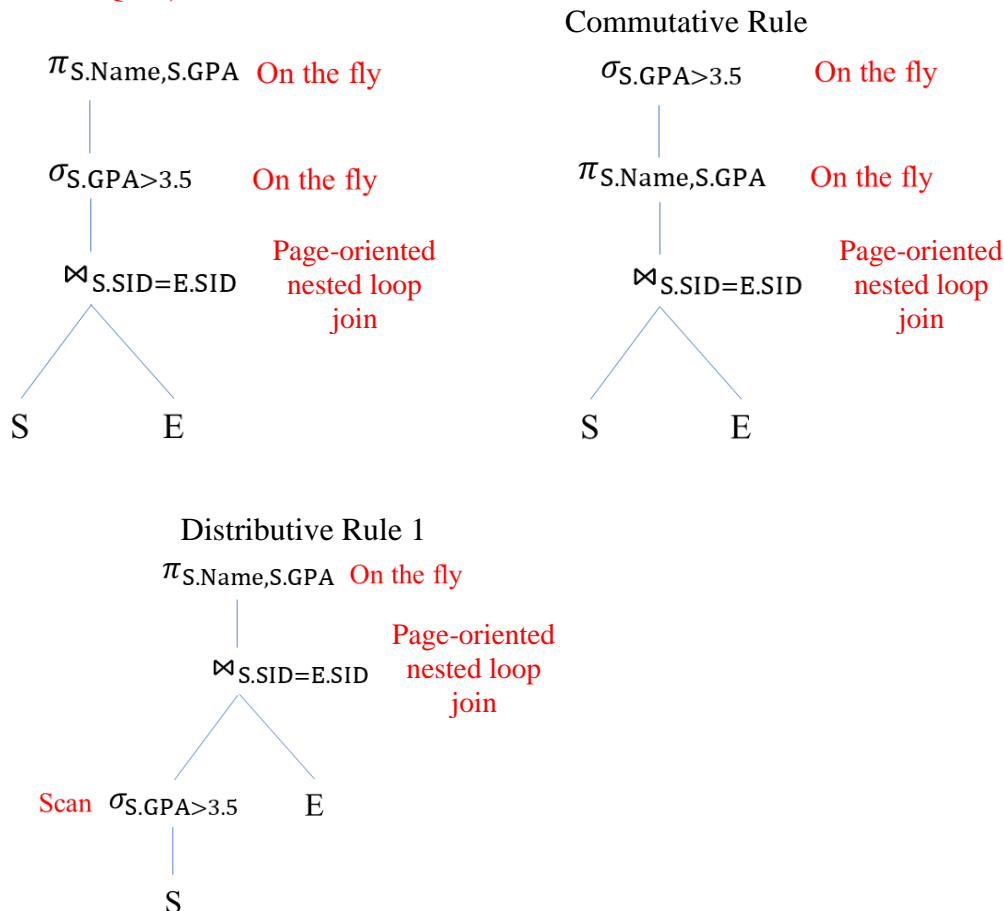Total cost = Cost of projection + Cost of join = 165 + 255 = 420 I/Os          (4 marks)

**Question 4**. Given two relations S(SID, Name, GPA, age) and E(CID, SID, Date), and the following SQL query:

```
SELECT S.Name, S.GPA
FROM S, E
WHERE S.GPA>3.5 AND S.SID=E.SID
```

Draw three query evaluation plans for this SQL query. (**24 marks**)

**Solution:** Other appropriate access paths are also correct. (No access paths. Reduce 3 marks for each QEP.)

Commutative Rule

$\pi_{S.Name,S.GPA}$  On the fly          $\sigma_{S.GPA>3.5}$      On the fly

$\sigma_{S.GPA>3.5}$     On the fly          $\pi_{S.Name,S.GPA}$     On the fly

$\bowtie_{S.SID=E.SID}$     Page-oriented nested loop join          $\bowtie_{S.SID=E.SID}$     Page-oriented nested loop join

S          E          S          E

Distributive Rule 1

$\pi_{S.Name,S.GPA}$  On the fly

$\bowtie_{S.SID=E.SID}$     Page-oriented nested loop join

Scan  $\sigma_{S.GPA>3.5}$     E

S

———————————————————— ❖ ————————————————————