COMP 3311: Database Management Systems

Tutorial 8
Query Processing/Optimization

Exercise 1: Consider the relations R_1(A, B, *C*), R_2(C, D, *E*) and R_3(E, F). Primary keys are underlined, and all foreign keys are in italics and not null. Assume that:

R_1 has 1,000 tuples in 100 pages.
R_2 has 10,000 tuples in 1,000 pages.
R_3 has 100,000 tuples in 10,000 pages.

a) What is the query result size of $R_1 \bowtie R_2 \bowtie R_3$?

b) Give an efficient pipelining strategy to compute $R_1 \bowtie R_2 \bowtie R_3$.

Do not upload this exercise sheet to Canvas.

Exercise 2: Given the following relations and the information about them, process the relational algebra tree for the query on the right using a pipelined plan and answer the following questions.

Student(studentId, sName, gender)
1,000 tuples; 100 pages; index on studentId

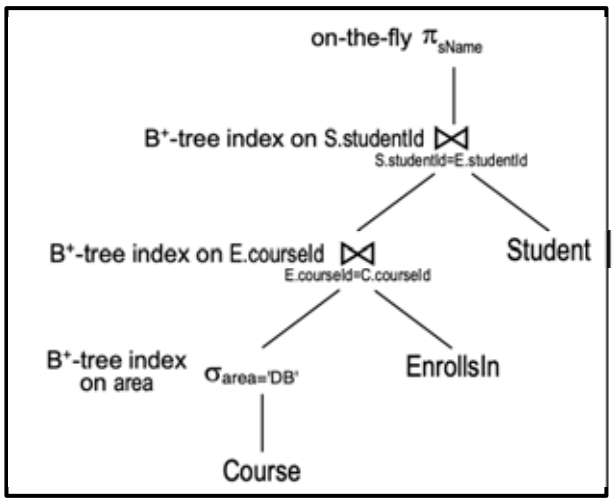
Enroll(*studentId*, *courseId*, year)
6,000 tuples; 600 pages; index on courseId

Course(courseId, cName, area, credit)
200 tuples; 40 pages; index on area;
10 different areas, with 20 tuples per area

All foreign keys are not null.

All indexes are B+-tree clustering indexes with 4 levels.

The EnrollsIn tuples are *uniformly distributed* among students and courses.




```
select sName
from Student S, EnrollsIn E, Course C
where S.studentId=E.studentId
      and E.courseId=C.courseId
      and area= 'DB';
```

a) Estimate the query result size.

b) Estimate the query page I/O cost using a pipelined plan (i.e., do not store intermediate results).

Step 1: $\sigma_{area='DB'}$Course ⟹ result A

Strategy:

Strategy explanation and cost calculation

Step 1 page I/O cost:

Step 2: result A ⋈ EnrollsIn ⟹ result B

Strategy:

Strategy explanation and cost calculation

Step 2 page I/O cost:

Step 3: result B ⋈ Student

Strategy:

Strategy explanation and cost calculation

Step 3 page I/O cost:

Query processing page I/O cost:

Do not upload this exercise sheet to Canvas.

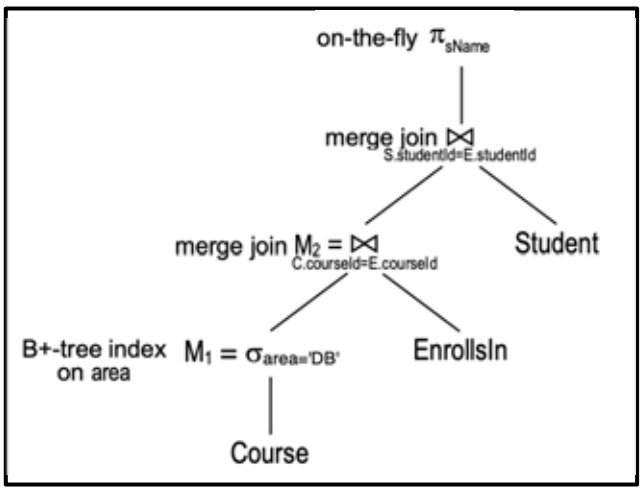
Name: (1) _____ **Student#:** (1) _____ **Date:** _____

Name: (2) _____ **Student#:** (2) _____

NOTE: You are highly encouraged to do this exercise with a partner.

COMP 3311: Database Management Systems

Tutorial 8
Query Processing/Optimization



Exercise 3: Suppose the intermediate results $M_1 = \sigma_{area='DB'}$ and $M_2 = M_1 \bowtie_{courseId}$ EnrollsIn are materialized for the relational algebra tree shown on the right. What is the query processing page I/O cost for this relational algebra tree assuming that merge join is used for all joins?

Assume there are 22 buffer pages and *attributes in the same relation all have the same size*.

Step 1: Cost to materialize $M_1 = \sigma_{area='DB'}$

Strategy: index lookup using B^+-tree on area

Strategy explanation and cost calculation

Step 1 page I/O cost to materialize M_1:

Step 2: Cost to materialize $M_2 = M_1 \bowtie_{courseId}$ EnrollsIn (using merge join)

Strategy: merge join

Strategy explanation and cost calculation

Step 2 page I/O cost to materialize M_2:

Step 3: Cost to compute $M_2 \bowtie_{studentId}$ Student (using merge join)

Strategy: merge join

Strategy explanation and cost calculation

Step 3 page I/O cost:

Query processing page I/O cost:

You must upload this completed exercise sheet to Canvas by **11 p.m. today.**