COMP 3311: Database Management Systems

Tutorial 8
Query Processing/Optimization

Exercise 1: Consider the relations R_1(<u>A</u>, B, *C*), R_2(<u>C</u>, D, *E*) and R_3(<u>E</u>, F).
Primary keys are underlined and foreign keys are in italics.
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 size of the final result of $R_1 \bowtie R_2 \bowtie R_3$?

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

Exercise 2: Consider the following relations and information about them

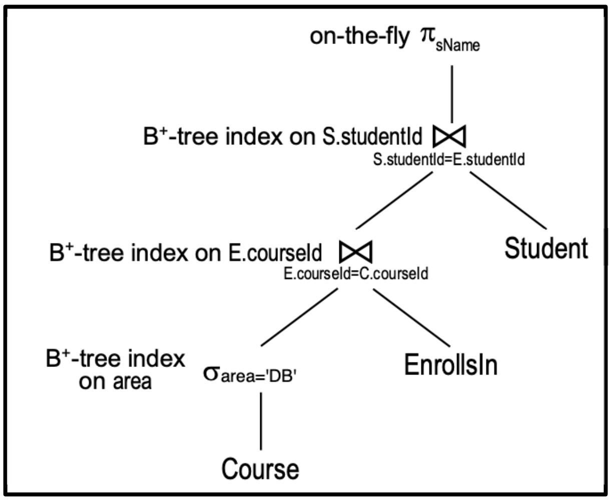
Student(studentId, sName, gender)
1,000 tuples; 100 pages; index on studentId

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

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

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

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


on-the-fly π_{sName}
B+-tree index on S.studentId $\bowtie_{S.studentId=E.studentId}$
B+-tree index on E.courseId $\bowtie_{E.courseId=C.courseId}$
Student
B+-tree index on area
$\sigma_{area='DB'}$
EnrollsIn
Course

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

Using a pipelined plan to process the relational algebra tree for the query on the right, answer the following questions.

a) Estimate the result size of the query.

b) Estimate the cost for this query using the pipelined plan.

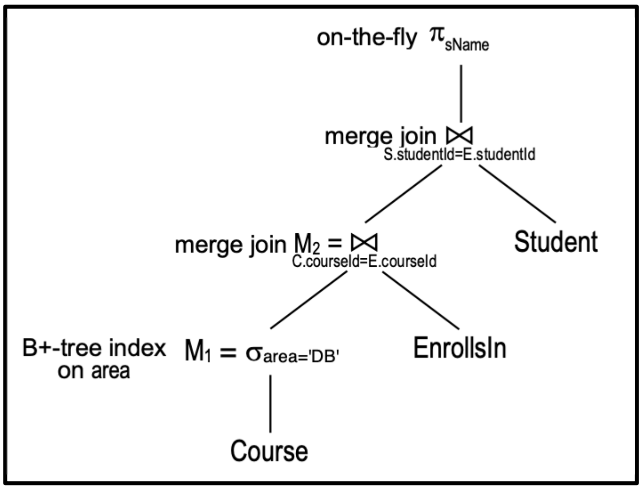
Name: (1) ________________________________ **Student#:** (1) ________________________ **Date:** __________

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


on-the-fly π_{sName}
merge join $\bowtie_{S.studentId=E.studentId}$
merge join $M_2 = \bowtie_{C.courseId=E.courseId}$
Student
B+-tree index on area
$M_1 = \sigma_{area='DB'}$
EnrollsIn
Course

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

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

Briefly explain the strategy for each step.

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

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

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

Total cost:

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