COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Exercise 1: Given relation R(A, B, C)

Assume: R contains 10,000 tuples in 1,000 pages.
A has 50 distinct values in the range 1…50.
B has 100 distinct values in the range 0…100.

Estimate the size, *SC* (number of tuples), of each of the following operations *assuming uniform distribution and attribute independence*.

a) $\sigma_{A=10}R$

b) $\sigma_{A=10 \wedge 20<B}R$

c) $\sigma_{C=1}R$

d) $\sigma_{C=10 \wedge A=10}R$

e) $\sigma_{C=10 \wedge A=10 \wedge 20<B}R$

Do not upload this exercise sheet to Canvas.

COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Exercise 2: Consider the relation Sailor(<u>sailorId</u>, sName, rating, age) and the query:

```
select sName
from Sailor
where rating=7
      and age=40;
```

n_{Sailor} = 10,000 B_{Sailor} = 1,000 pages $bf_{Sailor} = \lceil 10{,}000 / 1{,}000 \rceil = 10$

V(rating, Sailor) = 10 (10 distinct rating values)

V(age, Sailor) = 100 (100 distinct age values)

SC(rating=7, Sailor) = n_{Sailor} / V(rating=7, Sailor) = 10,000 / 10 = 1,000 tuples

SC(age=40, Sailor) = n_{Sailor} / V(age=40, Sailor) = 10,000 / 100 = 100 tuples

Estimate the page I/O cost to process the query of the following alternative plans assuming uniform distribution and attribute independence. *Ignore the cost of searching any indexes.*

a) file scan

b) binary search
   cost to search on rating

   cost to search on age

c) single B^+-tree index (on either attribute)
   index on rating

   index on age

d) multiple B^+-tree indexes (on both rating and age)

<u>Do not</u> upload this exercise sheet to Canvas.

COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Exercise 3: Employee(empId: 4 bytes, name: 35 bytes, title: 2 bytes, salary: 5 bytes, deptId: 4 bytes)
Department(deptId: 4 bytes, name: 25 bytes, location: 7 bytes, projectId: 4 bytes)
DeptProject(deptId: 4 bytes, projectId: 4 bytes)
Project(projectId: 4 bytes, title: 20 bytes, budget: 6 bytes, report: 970 bytes)

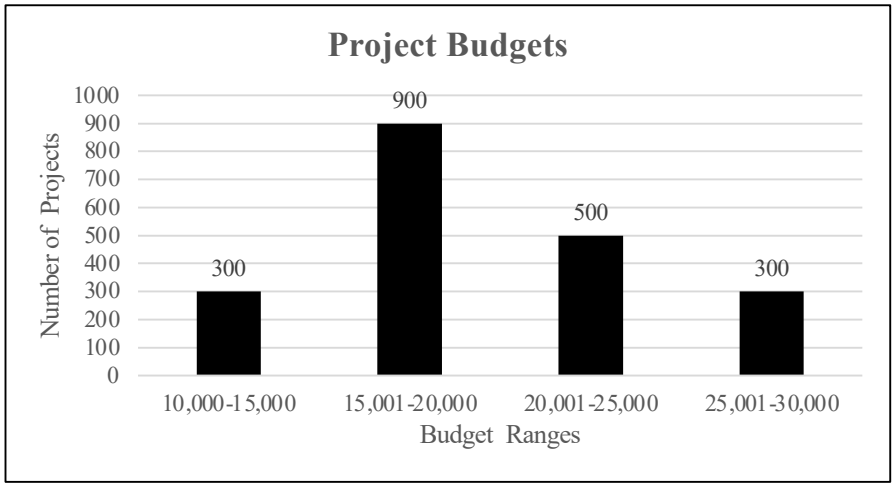
Employee:	50 bytes/tuple;	20,000 tuples	250 pages
Department:	40 bytes/tuple;	100 tuples	1 page
DeptProject	8 bytes/tuple	4,000 tuples	8 pages
Project:	1,000 bytes/tuple;	2,000 tuples	500 pages

Page size: 4,000 bytes Buffer pages: 12

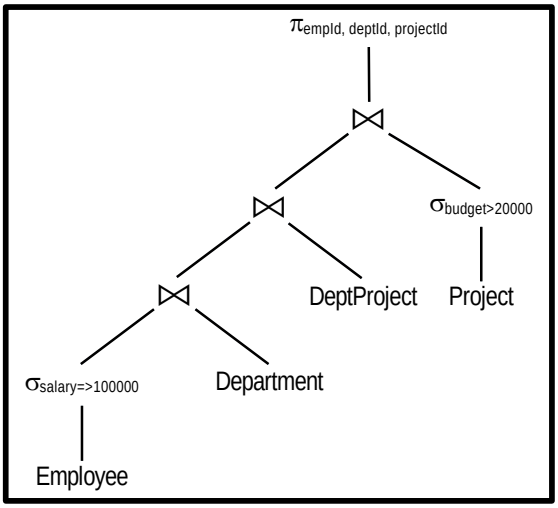
Employee salaries: uniformly distributed in the range 10,000 to 110,000.

Project budgets: distributed in the range 10,000 to 30,000 according to the histogram below.

- There is a clustering B+-tree index with 3 levels on salary for Employee.
- There is a hash index on deptId for Department; Department is ordered on deptId.
- There is a hash index on projectId for Project; Project is ordered on projectId.




```
select distinct empId, deptId, projectId
from Employee natural join Department
        natural join DeptProject
        natural join Project
where salary=>100000
        and budget>20000;
```



Do not upload this exercise sheet to Canvas.

Name: (1) _____/_____ **Student#:** (1) _____ **Date:** _____
Family/Given (PRINT) Given/First (PRINT)

Name: (2) _____/_____ **Student#:** (2) _____
Family/Given (PRINT) Given/First (PRINT)

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

COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Exercise 3

a) Use the relational algebra tree to estimate the output size of the query in tuples and in pages.

b) Evaluate the query using the relational algebra tree and the steps given below. The goal is to minimize the average number of page I/Os. Reorder operations, as necessary, to reduce the page I/O cost. Where possible, use pipelining rather than materialization (i.e., keep intermediate results in memory where possible). Assume the file organizations and indexes described above. For each step, give the strategy used, the cost calculation and the average case page I/O cost. Give the total page I/O cost to process the query.

Step 1: $\sigma_{salary=>100000}$Employee $\Longrightarrow$ result A

Strategy:

Cost calculation:

Step 1 page I/O cost:

Step 2: result A $\bowtie$ Department $\Longrightarrow$ result B

Strategy:

Cost calculation:

Step 2 page I/O cost:

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

Name: (1) _____/_____ **Student#:** (1) _____ **Date:** _____
Family/Given (PRINT) Given/First (PRINT)

Name: (2) _____/_____ **Student#:** (2) _____
Family/Given (PRINT) Given/First (PRINT)

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

COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Step 3: result B ⋈ DeptProject ⟹ result C

Strategy:

Cost calculation:

Step 3 page I/O cost:

Step 4: $\sigma_{budget>20000}$Project ⟹ result D

Strategy:

Cost calculation:

Step 4 page I/O cost:

Step 5: result C ⋈ result D

Strategy:

Cost calculation:

Step 5 page I/O cost:

Query page I/O cost:

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