

COMP 3311: Database Management Systems

Lecture 17 Exercises Query Optimization

Sailor(sailorId, sName, rating, age) Reserves(sailorId, boatId, rDate) Boat(boatId, bName, color)

- There are 10,000 Sailor tuples, 100,000 Reserves tuples and 1,000 Boat tuples.
- For all files, $bf = 10$ tuples per page.
- $B_{\text{Sailor}} = 1,000$ pages; $B_{\text{Reserves}} = 10,000$ pages; $B_{\text{Boat}} = 100$ pages.
- For a join result $bf = 5$ tuples per page.
- The buffer $M = 100$ pages.
- There are the following indexes:

- hash index on sailorId for Sailor (no overflow buckets).
- clustered B+-tree on rDate for Reserves (2 levels).
- hash index on boatId for Boat (no overflow buckets).

Our goal is to process the query:

```
select *
from Sailor natural join Reserves natural join Boat
where rDate='01-JAN-2020'
and color='red';
```

Some useful statistics:

- Reserves has 1,000 unique rDates.
- 10% of boats are red.
- A sailor has on average 10 reservations.

Exercise 1: Estimate the minimum page I/O cost to process the query using materialization and the join order

(Sailor JOIN $\sigma_{rDate='01-JAN-2020'}$ Reserves) JOIN $\sigma_{color='red'}$ Boat

C₁: Cost of computing Temp₁ = (Sailor JOIN $\sigma_{rDate='01-JAN-2020'}$ Reserves)

Strategy:

Cost calculation for Sailor JOIN $\sigma_{rDate='01-JAN-2020'}$ Reserves:

Page I/O cost to write Temp₁:

Total page I/O cost for C₁:

C₂: Cost of computing Temp₂ = $\sigma_{color='red'}$ Boat (no index on color)

Strategy:

Cost calculation for $\sigma_{color='red'}$ Boat:

Page I/O cost to write Temp₂:

Total page I/O cost for C₂:

C₃: Cost of Temp₁ JOIN Temp₂

Strategy:

Cost calculation for Temp₁ JOIN Temp₂:

Total page I/O cost for C₃:

Query processing page I/O cost:

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 17 Exercises Query Optimization

Sailor(sailorId, sName, rating, age) Reserves(sailorId, boatId, rDate) Boat(boatId, bName, color)

- There are 10,000 Sailor tuples, 100,000 Reserves tuples and 1,000 Boat tuples.

- For all files, $bf = 10$ tuples per page.

☞ $B_{\text{Sailor}} = 1,000$ pages; $B_{\text{Reserves}} = 10,000$ pages; $B_{\text{Boat}} = 100$ pages.

- For a join result $bf = 5$ tuples per page.
- The buffer $M = 100$ pages.
- There are the following indexes:
 - hash index on sailorId for Sailor (no overflow buckets).
 - clustered B+-tree on rDate for Reserves (2 levels).
 - hash index on boatId for Boat (no overflow buckets).

Our goal is to process the query:

```
select *
from Sailor natural join Reserves natural join Boat
where rDate='01-JAN-2020'
and color='red';
```

Exercise 2: Estimate the minimum page I/O cost to process the query using materialization and the join order

Sailor JOIN ($\sigma_{\text{rDate}='01-JAN-2020'}$ Reserves JOIN $\sigma_{\text{color}='red'}$ Boat)

Some useful statistics:

- Reserves has 1,000 unique rDates.
- 10% of boats are red.
- A sailor has on average 10 reservations.

C₁: Cost of computing Temp₁ = $\sigma_{\text{rDate}='01-JAN-2020'}$ Reserves

Strategy:

Cost calculation for $\sigma_{\text{rDate}='01-JAN-2020'}$ Reserves:

Page I/O cost to write Temp₁:

Total page I/O cost for C₁:

C₂: Cost of computing Temp₂ = Temp₁ JOIN $\sigma_{\text{color}='red'}$ Boat (no index on color)

Strategy:

Cost calculation for Temp₁ JOIN $\sigma_{\text{color}='red'}$ Boat:

Page I/O cost to write Temp₂:

Total page I/O cost for C₂:

C₃: Cost of Sailor JOIN Temp₂

Strategy:

Cost calculation for Sailor JOIN Temp₂:

Total page I/O cost for C₃:

Query processing page I/O cost:

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