COMP 3311: Database Management Systems

Lecture 18 Exercises
Query Optimization

Sailor(<u>sailorId</u>, sName, rating, age) Reserves(<u>sailorId, boatId, rDate</u>) Boat(<u>boatId</u>, bName, color)

- There are 10,000 Sailor tuples, 100,000 Reserves tuples and 1,000 Boat tuples.
- Assume that for all files there are 10 tuples per page.
  - ☞ B_{Sailor} = 1,000 pages; $B_{Reserves}$ = 10,000 pages; B_{Boat} = 100 pages.
- Assume a main memory buffer of M = 100 pages.
- Assume that 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-2019'
        and color='red';
```

<u>Some useful statistics:</u>

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

Exercise 1: Estimate the query processing cost using materialization and the join order

(Sailor JOIN $\sigma_{rDate='01\text{-}JAN\text{-}2019'}$ Reserves) JOIN $\sigma_{color='red'}$Boat.

C_1: Cost of computing $Temp_1$ = (Sailor JOIN $\sigma_{rDate='01\text{-}JAN\text{-}2019'}$Reserves)

Strategy:

C_2: Cost of computing $Temp_2$ = $\sigma_{color='red'}$Boat (no index on color)

Strategy:

C_3: Cost of $Temp_1$ JOIN $Temp_2$

Strategy:

<u>Total cost</u>:

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

Sailor(<u>sailorId</u>, sName, rating, age) Reserves(<u>sailorId, boatId, rDate</u>) Boat(<u>boatId</u>, bName, color)

- There are 10,000 Sailor tuples, 100,000 Reserves tuples and 1,000 Boat tuples.
- Assume that for all files there are 10 tuples per page.
  - ☞ B_{Sailor} = 1,000 pages; $B_{Reserves}$ = 10,000 pages; B_{Boat} = 100 pages.
- Assume a main memory buffer of M = 100 pages.
- Assume that 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-2019'
      and color='red';
```

<u>Some useful statistics:</u>
- Reserves has 1,000 unique rDates.
- 10% of boats are red.
- A sailor has on average 10 reservations.

Exercise 2: Estimate the query processing cost using materialization and the join order
Sailor JOIN ($\sigma_{rDate='01\text{-}JAN\text{-}2019'}$Reserves JOIN $\sigma_{color='red'}$Boat).

C_1: Cost of computing $Temp_1 = \sigma_{rDate='01\text{-}JAN\text{-}2019'}$Reserves

Strategy:

C_2: Cost of computing $Temp_2$ = $Temp_1$ JOIN $\sigma_{color='red'}$Boat (no index on color)

Strategy:

C_3: Cost of Sailor JOIN $Temp_2$

Strategy:

<u>Total cost</u>:

You must upload this completed exercise sheet to Canvas by **2:30 p.m. today.**