Database Systems (CPH) – Fall 2025
Teacher: Tiantian Liu
Exercise Sheet: Lecture 9: Query Processing and Optimization

AALBORG UNIVERSITY
DENMARK

1. Review the following terms. (Review the slides or the textbook: Chapters 15 and 16)

External sort-merge
Runs
Simple nested-loop join
Page-Oriented nested Loops Join
Block nested-loop join
Sort-merge join
Hash join
Equivalence rules
Statistics estimation
Size estimation
Histograms

2. Let relations r_1 and r_2 have the following properties: r_1 has 20,000 tuples, r_2 has 45,000 tuples, 25 tuples of r_1 fit on one page, and 30 tuples of r_2 fit on one page.

(Hint: r_1 needs 800 pages, and r_2 needs 1500 pages. Let us assume M pages of memory and $M \leq 800$.)

1) Estimate the I/O cost required using each of the following join strategies for $r_1 \bowtie r_2$:

a. Simple nested-loop join.
b. Page-oriented nested-loop join.
c. Sort-merge join. (Assume that one buffer page is needed to hold the evolving output page.)
d. Hash join. (Assume that there is no need for recursive partitioning.)

2) Say we have M = 100 + 2 memory buffers, estimate the I/O cost required using the following join strategy for $r_1 \bowtie r_2$:

e. Block nested-loop join.

Answer:

Assume we do not concider the I/O cost of the final writing.

a. Simple nested-loop join:
If r_1 is the outer relation, the cost is 20000 * 1500 + 800.
If r_2 is the outer relation, the cost is 45000 * 800 + 1500.

b. Page-oriented nested-loop join:
If r_1 is the outer relation, the cost is 800 * 1500 + 800.
If r_2 is the outer relation, the cost is 1500 * 800 + 1500.

c. Sort-merge join:
If 1500/M + 800/M < M, the cost is 3 * (1500 + 800).
Otherwise, the sorting cost is

$B_s = 2 * 1500(\lceil \log_{M-1} \lceil 1500/M \rceil \rceil + 1) + 2 * 800(\lceil \log_{M-1} \lceil 800/M \rceil \rceil + 1)$.
The total cost is $B_s + 1500 + 800$.

d. Hash join:
The cost is 3 * (1500 + 800).

e. Block nested-loop join:
If r_1 is the outer relation, the cost is (800/100) * 1500 + 800.
If r_2 is the outer relation, the cost is (1500/100) * 800 + 1500.

3. Given a relation $r(A, B, C)$ with $n_r = 10000$ and $V(A, r) = 500$.
($V(A, r)$ means the number of distinct values that appear in the relation r for attribute A.)

a. Estimate the size of the selection operation $\sigma_{A=10}(r)$.

b. Assume the range of values for an attribute C is [7, 59] and the values are uniformly distributed. Estimate the size of the selection operation $\sigma_{c<10}(r)$.

Answer:

a. The estimated size of the selection is $n_r / V(A, r) = 10000 / 500 = 20$.

b. The rate of $C < 10$ is $(10 - 7) / (59 - 7) = 3 / 52$.
The estimated size of the selection is
$n_r * (3 / 52) = 10000 * (3 / 52)$.