

Implementation of Databases Exercise 4

Ilya Kulikov 351063, Alina Shigabutdinova 351017, Oleg Chernikov 351016

December 12, 2015

Exercise 4.1 (Query Optimization)

Here we have relations $R(A, B)$ and $S(B, C)$ such that $|R| = |S| = 1000000$ tuples. One page can hold 100 R tuples and 10 S tuples. Number of join possibilities $|R \bowtie S| = 10000$. Also we have 102 buffer pages B . Hence, 10000 pages for R and 100000 for S and 1000 for R index and 1000 for S index.

P1

Here we have BNLJ with R as the outer relation and S as the inner relation.

$$\begin{aligned} \text{IO} &= M + \frac{M}{B-2}N = \\ &= P_R \frac{x}{100} + \frac{P_R}{B-2} \frac{x}{100} P_S = \\ &= 100x + 100000x = 100100x; \end{aligned}$$

P2

Here we have BNLJ with R as the outer relation and the index on $S.B$ as the inner relation.

$$\begin{aligned} \text{IO} &= 10000 \frac{x}{100} + \frac{10000}{100} \frac{x}{100} 1000 = \\ &= 100x + 1000x = 1100x; \end{aligned}$$

The best use case is $x \leq 10$.

P3

Here we have BNLJ with the index on $S.B$ as the outer relation and R as the inner relation.

$$\text{IO} = 1000 + \frac{1000}{100} 10000 \frac{x}{100} = 1000x + 1000;$$

The best use case when $x = 10; x = 11$.

P4

Here we have Merge join with the index on $S.B$ and the index on $R.B$. As the indexes are already sorted according to the join attribute B , no sorting is necessary. Each access to a tuple of R to retrieve the value for $R.A$ costs one additional IO operation, but this needs to be done only for

matching tuples of the join (i.e., the selection condition $R.A < x$ is checked after the join).

$$IO = 1000 + 1000 + |R \bowtie S| = 12000;$$

The best use case for $x \geq 11$.