

Implementation of DBMS
Exercise Sheet 15, Solutions
Klingemann, WS 2024 / 2025

1) Consider a clustered relation $R(A, B, C, D)$ that has a clustering index on A and a non-clustering index on each of the other attributes. The relevant parameters are: $B(R) = 1000$, $T(R) = 5000$, $V(R, A) = 20$, $V(R, B) = 1000$, $V(R, C) = 5000$, and $V(R, D) = 500$. Give the best query plan for the following selection and the corresponding number of disk I/O's. You can ignore the cost for accessing the index.

$\sigma_{A=10 \text{ AND } C=2 \text{ AND } D=3}(R)$

Solution:

We have the following options:

- Scan the complete relation. As we have a clustered relation the number of I/O's is $B(R) = 1000$
- Use the index on A . As it is a clustering index the number of I/O's is $B(R) / V(R, A) = 1000 / 20 = 50$
- Use the index on C . As it is a non-clustering index the number of I/O's is $T(R) / V(R, C) = 5000 / 5000 = 1$
- Use the index on D . As it is a non-clustering index the number of I/O's is $T(R) / V(R, D) = 5000 / 500 = 10$

Therefore, in the best query plan we use the index on C to read those tuples that fulfil the condition $C=2$ and check in main memory whether they also fulfil the rest of the condition. This plan requires 1 I/O.