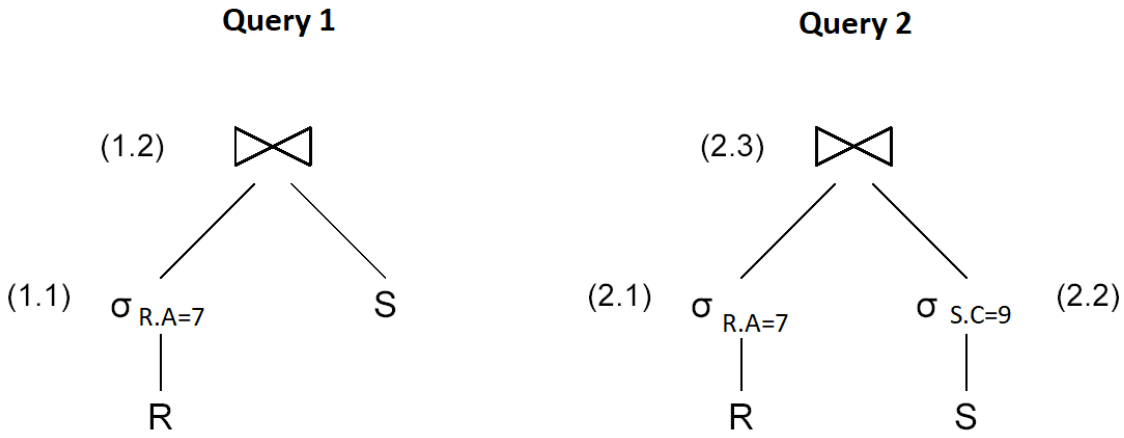
Implementation of DBMS
Exercise Sheet 14
Klingemann, WS 2024 / 2025

1) Consider the following queries on relations R(A, B) and S(B, C):
Query 1: SELECT * FROM R, S WHERE R.B = S.B AND R.A = 7
Query 2: SELECT * FROM R, S WHERE R.B = S.B AND R.A = 7 AND S.C=9
The following two plans are being considered:



Relation R has 60,000 tuples with 10 tuples per block, V(R, A) = 6 and V(R, B) = 12. Similarly, S has 30,000 tuples with 30 tuples per block, V(S, B) = 5 and V(S, C) = 20. Assume values are distributed over possible V(Relation, Attribute) values (not over possible domain values).
In this exercise we make the following additional assumptions:

- The join is implemented as a hash-join;
- The intermediate result produced by operation (1.1) is not written to disk;
- Hash buckets are stored on disk;
- The final result is kept in main-memory;
- There is enough memory to execute the hash join algorithm.

a) How many I/O's does query 1 require?
b) Also assume that the result of neither (2.1) nor (2.2) is stored to disk.
How many I/O's does query 2 require?

2) Suppose B(R) = 20000, B(S) = 50000, and the number of main memory blocks is M = 101. We want to perform the natural join of R and S using a merge join algorithm. Both relations are clustered but none of them is sorted.
a) How many passes do we need?
b) Describe how the join is executed and how many I/O's are required.

3) Relations R(a,b) an S(b,c) have r and s tuples, respectively. In both relations, attribute b takes values 1, 2, ..., 10 only, with equal probability. We also assume that both relations have duplicates, and on average a tuple that appears at all appears three times (i.e., eliminating duplicates reduces the size of R or S by a factor of 3). We wish to compute, without duplicates, those tuples in the natural join of R and S that have b = 10. Here are three possible plans for answering this query (none may be the absolute best in some circumstances):

Plan A	**Plan B**	**Plan C**
T1 := $\sigma_{b=10}$(R)	T4 := $\sigma_{b=10}$(R)	T7 := δ(R)
T2 := δ(S)	T5 := $\sigma_{b=10}$(S)	T8 := $\sigma_{b=10}$(S)
T3 := T1 JOIN T2	T6 := T4 JOIN T5	T9 := T7 JOIN T8
Ans := δ(T3)	Ans := δ(T6)	Ans := δ(T9)

Note, that δ means duplicate elimination.
We want to determine the cost of a plan by its expected sum of sizes of the intermediate relations. Provide for each plan a formula for the expected sum of sizes of the intermediate relations depending on the values of r and s.

4) Suppose B(R) = 20000, B(S) = 50000, and the number of main memory blocks is M = 101. We want to perform the natural join of R and S. Both relations are clustered but none of them is sorted. How many main memory blocks would be needed to perform a one-pass join?