DATABASE DESIGN II – 1DL400 – 2022

Assignment 2 (Lab 2)
Query Processing and Optimization

Examination
This assignment gives you the opportunity to practice with query processing and optimization. The assignment must be submitted before the indicated deadline on Studium. Please notice that you are not expected to finish this assignment during the lab: it may take longer. You must submit the result of this assignment on the Studium as one single PDF file per group. Please indicate group name, group participants with ID (personnummer, if you have one) and draw readable diagrams.

Questions 1. Assume that two tuples fit in a block and the memory holds at most 3 blocks. Show the runs created on each pass of the **sort-merge algorithm**, when applied to sort the following tuples on the first attribute:

(kangaroo, 17), (wallaby, 21), (emu, 1), (wombat, 13), (platypus, 3), (lion, 8), (warthog, 4), (zebra, 11), (meerkat, 6), (hyena, 9), (hornbill, 2), (baboon, 12).

Show the algorithm by clearly indicating how memory (e.g. blocks) or disk (e.g. runs) are used.

Question 2. Let relations ***r1(A, B, C)*** and ***r2(C, D, E)*** have the following properties: ***r1*** has 20,000 tuples, ***r2*** has 45,000 tuples, 25 tuples of ***r1*** fit on one block, and 30 tuples of ***r2*** fit on one block. Estimate the number of block transfers required, using each of the following join strategies for ***r1*** and ***r2***:

a. **Nested-loop join**.
b. **Block nested-loop join**.
c. **Merge join** (assuming both relations are sorted on the join attribute).

Question 3. Suppose that a B+-tree index on ***(dept_name, building)*** is available on relation ***department***. What would be the best way to handle the following selection?

$\sigma_{(building < "Watson") \wedge (budget < 55000) \wedge (dept\_name = "Music")}$(department)

Question 4. Show how to derive the following equivalences by a sequence of transformations using the equivalence rules.

a. $\sigma_{\theta_1 \wedge \theta_2 \wedge \theta_3}(E) = \sigma_{\theta_1}(\sigma_{\theta_2}(\sigma_{\theta_3}(E)))$

b. $\sigma_{\theta_1 \wedge \theta_2}(E_1 \bowtie_{\theta_3} E_2) = \sigma_{\theta_1}(E_1 \bowtie_{\theta_3} (\sigma_{\theta_2}(E_2)))$, where θ_2 involves only attributes from E_2

Question 5. Consider the relations ***r1(A, B, C)***, ***r2(C, D, E)***, and ***r3(E, F)***, with primary keys *A*, *C*, and *E*, respectively. Assume that ***r1*** has 1000 tuples, ***r2*** has 1500 tuples, and ***r3*** has 750 tuples. Estimate the size of ***r1*** $\bowtie$ ***r2*** $\bowtie$ ***r3***, and give an efficient strategy for computing the join.