Norwegian University of Science and Technology
Department of Computer and Information Science

EXAMINATION IN TDT4150 – ADVANCED DATABASE SYSTEMS

Contact person during the examination: Kjetil Nørvåg

Tlf.: 93440

Date:	**June 8th 2012**
Time:	**09.00-13.00**
Tools:	**D: No tools allowed except approved simple calculator.**

Language: English

Grading deadline: June 29th 2012

Problem 1 – Data models and architecture – 25 %

a) Give an overview of the main components of a database system and describe briefly the function of each of them during processing of a query.
b) Describe alternative process models of a database system and discuss advantages and disadvantages of each of them.

Problem 2 – Query optimization – 10 %

a) Assume a relation R having a very non-uniform distribution of values for a particular attribute. Initially only statistics with three values were stored (lowest value, highest value, number of tuples). Then the database was changed to use histogram instead. Why will queries on R now be able to perform more efficient? Give an example on such a query.
b) How can the choice of join algorithm depend on whether the result of the query should be sorted or not?

Problem 3 – Parallel and distributed databases – 10 %

a) Describe to what extent join is suitable for pipeline parallelism and partitioned parallelism.
b) What are advantages and disadvantages of asynchronous ("lazy") group-updating ("update anywhere") of replicas?

Problem 4 – Skyline – 30 %

X	Y
2	11
7	11
1	9
2	7
3	2
2	4
6	1
7	8
6	6
4	4

a) Assume a relation with two-dimensional data. Explain what is the requirement for a tuple in this relation to be in the result of the *skyline*-operator applied on this relation. Apply the skyline operator on the relation to the right and show which tuples will belong to the result.
b) Explain the divide and conquer algorithm to process skyline queries.

Problem 5 – Various – 25 %

a) Explain the principles behind keyword search in relational databases and what you get as result of such a search.
b) Explain top-k queries. Give an example of a top-2-query and the result of the query based on the relation to the right. Make the necessary assumptions.