

**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.: 41440433

Date: December 15th 2012

Time: 09.00-13.00

Tools: D: No tools allowed except approved simple calculator.

Language: English

Grading deadline: January 12th 2013

Problem 1 – Data models – 10 %

Explain logical and physical data independence, and why these can be important features of a data model.

Problem 2 – Query optimization – 20 %

a) Explain why query optimization is an important component in a database system (DBMS).

b) Assume two queries:

Q1: SELECT * FROM r1 NATURAL JOIN r2;

Q2: SELECT * FROM r1 NATURAL JOIN r2 ORDER BY r2.c1;

Describe some ways the execution plan for Q2 may differ from the execution plan for Q1. State the assumptions you make.

Problem 3 – Parallel and distributed databases – 20 %

a) Is horizontal or vertical partitioning preferable?

b) Suppose that two transactions are performed simultaneously from two different nodes and that they simultaneously want to write to the same object. This object is replicated on both nodes. What happens:

1) In the context of eager primary copy replication?

2) In the context of eager update anywhere replication?

3) In the context of lazy update anywhere replication?

Problem 4 – Rank join – 20 %

A tourist would like to visit Trondheim and she also wants to rent a car. She desires the car rental shop to be near the hotel. She has the options shown in the table below. Run the rank-join algorithm for $k=3$ on the two relations below and state clearly the step that the algorithm will finish and the threshold value at this step (join on the Area attribute). The scoring function is $f=0.5*PriceA+0.5*PriceB$.

ID	Rent a Car	Area	PriceA
a1	Wreck	Sentrum	100
a2	Lux	Flatåsen	110
a3	Bil	Sentrum	120
a4	RentBil	Flatåsen	130
a5	SuperBil	Flatåsen	140

ID	Hotel	Area	PriceB
b1	Cozy	Flatåsen	90
b2	Cheap	Sentrum	100
b3	Good	Sentrum	120
b4	Better	Flatåsen	120
b5	Best	Sentrum	150

Problem 5 – Various – 30 % (12% on a, 6% on b/c/d)

- a) Describe the differences between a traditional database system and a data stream management system.
- b) Explain the main principles behind MapReduce (hint: describe using the figure in the curriculum article as starting point).
- c) Explain the main principles behind the Google File System (hint: describe using the architecture figure in the curriculum article as starting point).
- d) Explain the motivation behind Hive, and give a brief overview of the functionality Hive offers.