

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



EXAMINATION IN DT8801 – ADVANCED DATABASE SYSTEMS

Contact person during the examination: Kjetil Nørvåg

Tlf.: 93440

Date: May 30th 2011

Time: 09.00-13.00

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

Language: English

Grading deadline: June 20th 2011

Problem 1 – Query optimization – 15 %

- Why does the System R optimizer only consider "left-deep" join-trees?
- Is it always possible to change the order of selection and projection? What can be the reason for changing the order of such operators?

Problem 2 – Parallel and distributed databases – 15 % (10% on a and 5% on b)

- What is partitioned parallelism in the context of execution of queries in distributed databases? Consider how easy it is to employ partitioned parallelism for selection, join, and aggregation.
- Discuss briefly to what extent *lazy update anywhere* replication supports the ACID properties.

Problem 3 –Data mining – 15 %

Assume the market basket to the right. Use the apriori-algorithm to find association rules, given minimum support of 50 % (i.e., *minimum support count* is 2) and confidence of 90 %.

| TransactionID | Items |
|---------------|---------|
| T100 | A,C,D |
| T200 | B,C,E |
| T300 | A,B,C,E |
| T400 | B,E |

Problem 4 – Data streams – 15 %

- Explain the concept of continuous queries and how they differ from traditional queries. Give three examples of continuous queries.
- Explain the concept of sliding windows and their role in a Data Stream Management System.

Problem 5 – Skyline and ranking – 25 % (5% on a and b and 15% on c)

- Is it possible to write a SQL query that gives the same result as a skyline query (without using the SKYLINE operator)? If not, describe why it is not possible. If yes, describe the disadvantages of this approach.
- Discuss the advantages and disadvantages of the Block Nested Loop (BNL) algorithm for processing skyline queries. Especially explain in which scenarios it performs well and in which scenarios it should not be used.
- Explain the Rank Join algorithm. What are its advantages compared to a naive algorithm?

Problem 6– Various – 15 % (10% on a and 5% on b)

- Discuss similarities between the network model and XML, and the potential problems with the network model.
- Explain the problems related to under- and over-provisioning and how cloud computing can be used to minimize these problems.