

**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 7th 2010

Time: 09.00-13.00

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

Language: English

Grading deadline: June 28th 2010

Problem 1 – Parallel and distributed databases – 20 %

- Explain *speedup* and *scaleup*.
- Is it always sensible to divide a relation into equal-sized partitions?
Why / Why not?
- Has the choice between synchronous and asynchronous update of replicas any consequences for the probability of deadlock? State the reasons for your answer.

Problem 2 – Peer-to-peer – 20 %

Describe Pastry (including contents of routing tables and their structure, how nodes become part of a Pastry network, and how data is inserted and retrieved).

Problem 3 – Data warehouse and data mining– 30 %

- Describe differences between OLTP (*on-line transaction processing*) and OLAP (*on-line analytical processing*) which make it desirable to have separate systems for data warehouses.
- Assume the market basket to the right. Use the apriori-algorithm to find association rules, given minimum support of 22% (i.e., *minimum support count* is 2) and confidence of 70%.

TransactionID	Items
T100	A,B,E
T200	B,D
T300	B,C
T400	A,B,D
T500	A,C
T600	B,C
T700	A,C
T800	A,B,C,E
T900	A,B,C

Problem 4 – Various– 30 %

- Does it matter which relation is the outer and inner relation during execution of a join?
Why / Why not ?
- Is a column-based database system like C-Store optimized for reading or writing? State the reasons for your answer.
- Assume that in a database system with support for top-*k* join queries we want these queries to be easy to express in SQL. Give an example of how such a query could have been expressed in SQL.
- Describe what we mean when we say that a point is dominated by another point.