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 2nd 2008**

Time: **09.00-13.00**

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

Language: English

Grading deadline: June 23rd 2008

Problem 1 – Query processing– 15 %

a) For which purpose are histograms used in the context of query-optimization?
b) What is the difference between *equiwidth* og *equidepth* histograms?
c) What are the necessary conditions for relational algebra expressions Q1 and Q2 to be equivalent?
   Q1: π_a (R $\bowtie_c$ S)
   Q2: $(\pi_{a1}$ (R))$\bowtie_c$ $(\pi_{a2}$ (S))

Problem 2 – Parallel and distributed databases – 25 %

a) Describe 3 different techniques for partitioning and describe briefly what are good and bad properties of each of them.
b) Describe *synchronous* and *asynchronous* updating in the context of replication, and discuss what are good and bad properties of each of them.
c) Draw a sketch with states and messages during execution of two-phase-commit (2PC).

Problem 3 – Data warehouse and data mining– 30 % (5% on a, 10% on b, 15% on e)

a) Name 3 common data-warehouse operations/operators that can be applied to multidimensional data and describe the function of each of them.
b) What is the aim of clustering? Show (using pseudo-code) the algorithm of one common method for clustering.
c) Assume the market basket to the right. Use the apriori-algorithm to find association rules, given minimum support of 50% and confidence of 80%.

TransactionID	Items
1000	A,B,C
2000	A,C
4000	A,D
5000	B,E,F

Problem 4 – Various– 30 % (15% on a and 5% on each of the following)

X	Y
1	9
2	11
3	2
4	4
5	9
6	6
7	11
7	8
8	5
9	1

a) Assume a relation with two-dimensional data. Describe what is the condition for a point in this relation to be in the result of the *skyline*-operator applied on this relation. Assuming the relation to the left, what tuples will appear in the result of the *skyline*-operator applied to that relation?

b) What are the principles behind column-based database systems, and for which access patterns are such database systems most suitable?
c) What is the difference between *valid-time* and *transaction-time* in the context of temporal databases?
d) Draw a sketch with the data-structures of a POSTGRES-page, and describe briefly the contents/purpose of each of them.