

**School of Information Technology
IIT Kharagpur**

Course Id: IT60107 Data Warehousing and Data Mining (Class Test 1)

Date: August 22, 2012

Total Time: 1 Hour

Max. Marks: 20

Answer all questions.

1. Consider a 3-D data array consisting of 3 dimensions A, B and C. The 3-D array is partitioned into 64 memory chunks. Dimension A is organized into 4 equi-sized partitions a₀, a₁, a₂ and a₃. Similarly dimensions B and C are organized into 4 equi-sized partitions each. Chunks are numbered as 1, 2, 3, ..., 64 corresponding to the sub cubes a₀b₀c₀, a₁b₀c₀, a₂b₀c₀, a₃b₀c₀, a₀b₁c₀, ..., a₃b₃c₃, respectively. Suppose the sizes of the dimensions A, B and C are 2400, 3600 and 1600, respectively. If we perform multi-way array aggregation in the serial order 4, 20, 36, 52, 3, 19, 35, 51, 2, 18, 34, 50, 1, 17, 33, 49, 8, 24, 40, 56, ..., 5, 21, 37, 53, ...16, 32, 48, 64, ..., 13, 29, 45, 61, then calculate the minimum memory requirement for holding all relevant 2-D partial sums in memory. **[10]**

2. Consider the following lattice of cuboids (A-H) along with a representation of the number of rows in each cuboid where A is the base cuboid. If you have to choose 3 views to materialize apart from the base cuboid, which of the views B-H would you choose using a greedy algorithm and how (show in a tabular form your choices sequentially)? Assume that the cost of running a query is the same as the number of rows in the view from which it is derived. How much is the percentage improvement in query performance with this set of materialized views with respect to no-materialization? **[8+2=10]**

