

## Class Test I

### Course Id: IT60107 Data Warehousing and Data Mining

**Date: September 1<sup>st</sup>, 2004.**

**Total Time: 60 minutes**

**Max. Marks: 60**

Instructions: Answer all questions. You may answer the questions in any order. However, all parts of the same question must be answered together. Clearly state any reasonable assumption you make.

1. Consider a 3-D data array consisting of 3 dimensions A, B and C. The 3-D array is partitioned into 64 memory-based chunks. Dimension A is organized into 4-equisized partitions a0, a1, a2 and a3. Similarly dimensions B and C are also organized into 4-equisized partitions each. Chunks are numbered as 1, 2, 3, ..., 64 corresponding to the sub cubes a0b0c0, a1b0c0, a2b0c0, a3b0c0, a0b1c0, ..., a3b3c3, respectively. Suppose the size of the array of the dimensions A, B and C are 20, 2,000 and 200, respectively. If we perform multi-way array aggregation, then calculate the minimum memory requirement for holding all relevant 2-D partial sums in chunk memory when the chunks are brought into memory in the order: 1, 17, 33, 49, 5, 21, ..., 13, 29, 45, 61, 2, 18, ...

**[20 Marks]**

2. A chain of departmental stores called “India-Mart” having operations only in India, plans to develop a data warehouse for effective decision-making about their sales and different promotion schemes. India-Mart puts some of their products on promotional sales from time to time. There may be a large number of different types of promotions like coupon sales, end-of-the-aisle display, buy-two-get-one-free, etc.

India-Mart would like to analyze how item sale is affected by the promotions at each store, in each state and across the entire country.

With respect to the above business scenario, answer the following questions.

**[15+5+5+5+5+5]**

- a. Design a star schema for the data warehouse clearly identifying the fact table(s), dimension table(s), their attributes and measures along with the primary key and foreign key relationships.
- b. Write an SQL query by which you can display year-wise, promotion-wise, product-wise total sales in the entire country from your schema.
- c. Draw a cuboid that would display the result of the query specified in Q. b above.
- d. From the cuboid of Q. c above, if we want to find the total amount of promotional sales made during the years 2002 and 2003 for the states of Karnataka and Maharashtra, which sequence of OLAP operations would you need to perform?
- e. Draw possible schema hierarchies for each dimension that you have designed.
- f. Based on the schema hierarchies drawn in Q. e above, determine the total number of cuboids, considering all the aggregation levels.