PSG College of Technology, Coimbatore -641 004

Department of Applied Mathematics and Computational Sciences

8th Semester MSc TCS

18XT87 Data Mining Lab

Problem Sheet – 5

Data Warehouse

1.

Table 3.3 A 3-D view of sales data for *AllElectronics*, according to the dimensions *time*, *item*, an *location*. The measure displayed is *dollars_sold* (in thousands).

	location = *"Chicago"*				***location*** = *"New York"*				***location*** = *"Toronto"*				***location*** = *"Vancouver"*			
	item				***item***				***item***				***item***			
	home				*home*				*home*				*home*			
time	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*
Q1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
Q2	943	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
Q3	1032	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
Q4	1129	992	63	870	1142	1091	54	984	978	864	59	784	927	1038	38	580

Consider the above data and perform the following operations

(a) Roll-up on location (from cities to countries)

(b) Drill-down on time (from quarters to months)

(c) Dice for (location='Toronto' or 'Vancouver' and (time='Q1' or 'Q2') and (item='home entertainment' or 'computer')

(d) Slice for time = 'Q1'

2. Suppose that a data warehouse for *Big_University* consists of the four dimensions *student, course, semester* and *instructor* and two measures *count* and *avg_grade*. At the lowest conceptual level (e.g., for a given student, course, semester and instructor combination), the avg_grade measure stores the actual course grade of the student. At higher conceptual levels, avg_grade stores the average grade for the given combination.

   (a) Draw a snowflake schema diagram for the data warehouse

   (b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (e.g., roll-up from semester to year) should you perform in order to list the average grade of CS courses for each Big_University student.

   (c) If each dimension has five levels (including all), such as 'student<major<status<university<all', how many cuboids will this cube contain (including the base and apex cuboids)? What are the values of these cuboids?