

## Assignment 1 Dimensional Data Modeling

**Submission Date: Thursday 14-Sep-2017 Friday 15-Sep2017 (Start of Lab)**

**Question 1.** Differentiate between fully additive and semi additive measures. Provide Examples.

**Question 2.** Discuss the design impact of choosing between star schema and snow flake schema with respect to following factors:

- i) Performance of loading data from source to data warehouse (ETL).
- ii) Performance of querying/analyzing the data.

**Question 3.** What will be your design decision:

- i) If two departments of the same business use different attributes of the same dimension i.e. both the departments have different definition of the dimension.
- ii) If a dimension is used only by one fact table, will you store it directly into the fact table or create a separate dimension table for it?

**Question 4.** Suppose there are 3000 products sold by the store, 5 brands and each brand has 500 products each, there are 10 store locations in the country, also assume there are at least one sale per product per store per week. Estimate the number of rows of fact table retrieved and summarized for following types of queries:

	Product	Store	Time	# of Rows retrieved
Query 1	1 product	1 store	1 week	
Query 2	1 product	All stores	1 week	
Query 3	1 brand	1 store	1 week	
Query 4	1 brand	All stores	1 year	
Query 5	All brands	All stores	1 year	

For which of the above queries Aggregate fact tables should be used and why?

Also draw the appropriate dimensional model showing aggregate fact tables.

Suppose you created an aggregate fact table for the third query... Then how many rows you need to retrieve for Queries 3, 4 and 5?

**Question 5.** You are required to design a Dimensional Model in the way that it fulfills the requirement for the following system.

### Grocery System (POS)

The following queries shall be generated through your design:

1. Total sales of a particular product from all stores in the last quarter

2. Total sales by product by store by month
3. Yearly profit generated by stores in the north region
4. How customer deviates from store to store with particular products
5. When I promote one thing how does it affect the other
6. Check if more Products are sold on 1<sup>st</sup> 10 days and 20<sup>th</sup> to 25<sup>th</sup> date of the month than the whole month.
7. Average daily sales (in dollars) of product categories.
8. The total number of customers purchasing a particular product.
9. The total number of customers visiting a particular store in a month.
10. Count how many people buy with coupon.

### Design Requirements

Here is the eight points of the complete dimensional modeling design:

1. The processes, and hence the identity of the fact tables
2. The grain of each fact table.
3. The dimensions of each fact table
4. The facts, including pre-calculated facts
5. The dimension attributes with complete descriptions and proper terminology
6. How to track slowly changing dimensions
7. The historical duration of the database
8. The urgency with which the data is extracted and loaded into the data warehouse.