Assignment 1

Dimensional Data Modeling

Submission Date: Friday 9-Sep-2016 (Start of DW-A Lab i.e. in Lab#1 at 2pm)

Question 1. Why is the entity-relationship modeling technique not suitable for the data warehouse? How is the dimensional modeling different?

Question 2. What are junk dimensions? Are they necessary in DW?

Question 3. Differentiate between slowly and rapidly changing dimensions?

Question 4. Suppose there are 4000 products sold by the store, 5 brands and each brand has 800 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 systems.

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 1st 10 days and 20th to 25th 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.

Bank (Account System)
The following queries shall be generated through your design:

1. We want to see five years of historical data on every account. For all prior months it will be sufficient to see the end of month snapshot.
2. For the current month, we want a valid snapshot as of yesterday. We don't need the other prior days in the current month.
3. Every type of account has a primary balance. There is a significant need to group different kinds of accounts in the same analyses and compare primary balances.
4. Every type of account has a list of custom dimension attributes and custom numeric facts that tend to be quite different from account type to account type.
5. Every account is deemed to belong to a household. Upon studying the historical production data, we conclude that accounts and the individuals who own the accounts come and go from households as much as several times per year for each household.
6. Since some of the accounts were created many years ago, and by different production systems, we find that our records of the individual account holder' names and addresses differ from account to account in many case.
7. In addition to the household identification, we are very interested in demographic information as it pertains to both the individual account holders and the households. We also capture and store behavior scores relating to the activity in each of the accounts.

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.