# INF 551 – Fall 2015

## Homework #3 (70 points)

## Due: 11:59pm on 10/30/2015 to Blackboard

In this homework, we ask you to build a data warehouse to analyze the sale of beers by a selected number of bars. ☺

Here are the details on the transactional database that we provide to you. The database can be set up by running olap.sql on MySQL.

* + Its **beers**, **bars**, and **drinkers** table store the relevant information about these three entities.
  + Its **sells** table stores the price a particular bar sells a particular beer. We assume that the prices of beers do not change in the periods of time when we analyze the sales data (even amid the economic downturn :).
  + Its **purchase** table records how many bottle of beers (its quantity attribute) a drinker purchased the beer in a particular bar at a specific time (ptime or purchase time). Note that in this sample database, there are only two months of sales data: September and October of 2015.

We ask you to build a warehouse to analyze the sales data available in the purchase table. The warehouse uses a star schema as follows.

* It has a fact table “sales” which has all the attributes in purchase table, plus an additional attribute for sales amount, which is the unit price of the beer (found in the sells table) multiplied by the quantity.
* It has three dimension table, for beers, bars, and drinkers. You can simply use the tables from the transaction database for this purpose.
* Set up foreign keys in the sales table to properly refer to dimension tables.

We then ask you to perform some OLAP operations using this sample warehouse, as specified below.

Here are questions and deliverables for this problem.

1. Draw a diagram depicting the star schema of the warehouse. Indicate the keys and foreign keys in the tables.
2. Submit a SQL script that creates the “sales” table and load data (using “insert” command) into the table. You do not need to recreate dimension tables. . You should ignore the tuples in the Sells table with unknown prices, i.e., price is null.
3. Write SQL queries to perform the following analysis.
   1. Compute a breakdown of total sales amount by month. Which month has a better sale?
   2. Slice result in 1) on the month of September and further drill down to bars dimension.
   3. Slice result in 1) on the month of October and further drill down to bars dimension.
   4. Based on the analysis results so far, did you observe some bars sold more beers than other? If so, what are they? In which month?
   5. Now compute a breakdown of total sales amount by the manufacturer of beers. Which company was best performing (in terms of the total sales amount of its beers)?
   6. Drill-down the result in 5) by bars. Which bars helped most in selling the beers from the best performing company?
   7. Now compute the breakdown of total sales amount by the bars. Which bars generated most of the revenue (in terms of total sales amount by the bar)?
   8. Drill down result in 7) by location of bars. Where the most lucrative bars located?
4. To speed up OLAP operations, data warehouse often anticipates the analytical needs and precompute some analysis results (stored in materialized views) which can then be utilized to answer new analytical queries. A key problem is to determine which views can be used to answer a given query. In this question, suppose that we have the following view:

create view bar\_mon\_view as

select bar, extract(month from ptime) mon, sum(amount) amt

from sales

group by bar, mon;

1. Can this view be used to compute the query “compute a breakdown of total sales amount by month”? If yes, write a SQL script that uses the view to answer the query.
2. Can this view be used to compute the query “compute a breakdown of total sales amount by bar for the month of October”? If yes, write a SQL script that uses the view to answer the query.

Requirements: Your scripts should run correctly on MySQL. You can download and install MySQL by following this link: <http://dev.mysql.com/downloads/mysql/>. Or you may use MySQL on your Amazon computing instance which you can set it up for your account.