INF 551 – Spring 2016

Homework #3 (100 points)

Due: 11:59pm on 3/23/2016 to Blackboard

For this homework, please create a MySQL database "inf551", with user name "inf551" and password: "inf551". Please also install MySQL Java and Python connectors (you may refer to the lecture slides for some details).

1. [SQL: setup and data loading, 20 points]

   Write a SQL script "setup.sql" to do the following:

   a. Create the following tables in the database. Underlines indicate primary keys of tables. Here are foreign keys: buyer and seller in Purchase refer to name of Person table; maker of Product refers to cname (company name) of Company.
      - Company(cname, stockPrice, country)
      - Person(name, phoneNumber, city)
      - Product(name, price, category, maker)
      - Purchase(buyer, seller, store, product)

   b. Insert data in the provided spreadsheet "purchase.xlsx" into the tables.

      Describe how you turn the data in the spreadsheet into "insert into …" statements. For example, which text editor you use, and what are the steps of conversion, e.g., how to add single quotes to the string value, etc. You may even write a program to do the conversion. Note that manually tuple-by-tuple conversion is a time-consuming process, especially when the dataset is large. It should be avoided in this homework.

      Note that the data loading process here is similar to ETL (Extract-Transform-Load) in data warehousing.

2. [SQL: querying, 40 points]
   Write an SQL query for each of the following questions.
   a. Find names of people who bought cell phones.
   b. Find names of people who bought US products.
   c. Find names of people who bought US products but did not buy Chinese products. For this question, write two different queries.
   d. Find names of people who bought US products and do not live in Los Angeles.
   e. Find names of people who bought stuff from David or bought products from a company whose stock price is more than $100.
   f. Find the name, price, and category of the most expensive product.
   g. Find names of products with at least 3 units/copies sold.
   h. Find the name of the person who spent the most (in terms of the total purchase amount) in purchasing products.
   i. Find the number of products sold for each product category and buyer city, e.g., cell phone, los angeles, 5; cell phone, san gabriel, 3; etc.

j. Find the number of products sold for each product category.

Views (http://dev.mysql.com/doc/refman/5.7/en/create-view.html):

k. Create a view for question i.

l. Use the view to answer question j.

3. [Relational algebra: 20 points]
   Write a relation algebra expression for queries a, b, c, d, e, i, and j in Question 3.

4. [MySQL connectors: 20 points]
   a. Write a Python script “purchase.py” that outputs a list products bought by a given person.
      For example, “python purchase.py Mark” may output “ms office 2010”, etc.
   b. Write a Java program “Purchase.java” that does the same as part a.