EXAM – Business Intelligence

The project work can be done **by groups of 3 people**.
Exercise 1 & 2 must be delivered **within 16 March.** Since the second part is based on the result of Exercise 2, when every group will deliver the two exercises I will publish the correct operational schema so that you can do the next exercises.

DEADLINE OF THE WHOLE PROJECT 8 MAY 2017.

PART I

Exercise 1

Design the **conceptual model of a database** respecting the following requirements:

The database will register information about the orders of wines regarding an Italian mail order company sells bottle of wines labelled as DOCG – Denominazione di Origine Controllata e Garantita (Controlled Designation of Origin Guaranteed). A DOCG is also called a wine class.
The order processing system is based on the operational database with the following characteristics:
- The database must contain information about customers identified by a code. For each customer we have the last name e first name, the customer address, his/her phone and his/her orders.
- Each order is characterized by a customer, the delivery date, the order date, the total price spent by the customer for that order and the list of wines ordered specifying for each wine the number of bottle ordered and the revenue.
- Each wine is characterized by the wine class, the name, the bottle volume, the price per bottle and the price per case. Some wines are vintage.
- For each wine class we want to store the name, the alcohol content, the color and the region of origin

Exercise 2

Design the **logical model of the database** in Exercise 1.

Exercise 3

Create in My SQL the above database without instances.
Deliver a script file with SQL Instructions

Exercise 4

Solve the following three queries on the database Sakila.
1. List of actors of English films
2. List of films with the title containing the letter "f" and released after 1950 or after 1980.
3. List of actors worked only in English film
4.
Deliver a script file with text of query and SQL query

PART II

Exercise 5

Given the order processing system based on the operational database in Fig. 1. There is a general feeling among the directors and senior managers that the company is losing its market share. Within the past three months, two more companies have been formed and their presence in the market is already being felt. Also, recently, more customers than usual appear to be leaving the company and new customers are being attracted in fewer numbers than before.

Examples of business questions the directors of the company wish to ask of their data, in order to reach decisions about their future strategy, are:

1. Which bottle wines are increasing in popularity and which are decreasing, over the year 2014?
2. Which bottle wines are seasonal over the last 3 years?
3. Which italian customers place the same bottle wine orders on regular basis over the last 3 years? 4
4. Are some bottle wines vintage more popular in different italian customer regions?
5. Do italian customers from different regions tend to purchase a particular class of wines over the year 2014?

Note that the business questions must usually be "interpreted" before they can be translated into an SQL query. For instance, the business question "Do italian customers from different regions tend to purchase a particular class of wines over the year 2014?" can be interpreted as "For the year 2014 and italian customers, the total number of bottles ordered at more than twice the quantity as the average for all classes of wine, by customer region, by wine class name. The result is sorted by customer region and wine class name".

With respect to the above business scenario, answer the following questions:

1. Rewrite all business questions in the following form: "For a data subset to use, the metrics to compute, by ..., ..., by .... The result is sorted by ...".

2. Design a conceptual schema for the data mart to support the business questions. Your schema should at least be able to satisfy the above mentioned analysis requirements. You may motivate other suitable attributes for the dimensions. Clearly specify the fact granularity. (**Deliver the conceptual design**)

3. Give a relational data mart design and a SQL script that creates the table in MySQL (no need to populate the tables). (**Deliver a SQL script file with the instructions**)

4. Write SQL queries over the relational data mart for all the interpreted business questions. (**Deliver a SQL script file with the question in the form of point 1 and the SQL solution**)
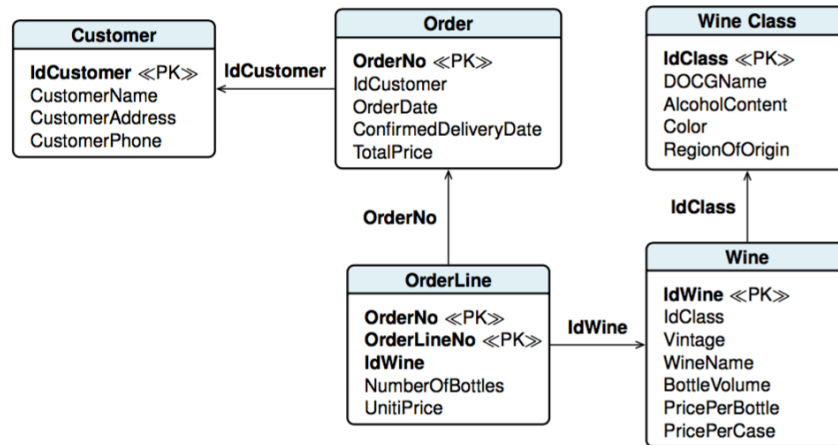
Figure 1: Relational Database Schema

Exercise 6 (ETL)

Consider the SAKILA database. A rental by a customer was done during a travel if the store of the rental was not in the city of residence of the customer.
- Create the database mastermains in MySQL
- Develop an ETL project using Pentaho Data Integration which creates a table ontravel in the mastermains database of the localhost MySQL server and populate it with a row for every customer with three columns:
  1. the customer full name;
  2. the total amount for the customer (this is 0 if the customer made no rentals);
  3. the ratio of total amount of rentals done during travels over the total amount of rentals (treat 0/0 as NULL).

  All the work must be done in the ETL project. **Deliver the Transformation file.**

- Solve the previous point with a single SQL script in MySQL. (**Deliver the SQL script file with the solution**)