MSc in Business Analytics

Data Management and Business Intelligence

Study Group Assignment 1

Task 1 (10%):

Design a database for the system described below.

The system facilitates the operation of a mental health institute. Its main use is to serve as a case tracking system. Each case has the following information:

- The date the patient first contacted the institute about this specific case. It doesn't necessarily have to be the same date the first therapy session took place.

- Who knows about the problem. Possible values are {Nobody, My Partner, My parents, My Siblings, My Friends, My Co-workers}. More than one value could be true.

- Each case is associated with one patient. One patient may be associated with more than one case.*

- Each case is associated with a list of prescribed drugs. You will find the list of all the possible drugs in the .zip file of the assignment. One case may be associated with more than one drug. Each drug is prescribed for a certain timeframe (Has a Start/Stop Date).**

- Each case is associated with a list of diagnoses. You will find the list of all the possible diagnosis values in the .zip file of the assignment. One case may be associated with more than one diagnosis.***

- Each case is associated with a list of therapy sessions. One case may be associated with more than one therapy session.

Each therapy session has the following information:

- Duration (minutes).

- The date it took place.

- The outcome. Possible values are {Healed, Improved, Worsened, Stabilized}.

Each patient has the following information:

- Gender.

- Date of birth.

- Address (Based on the street he/she lives in).****

- Current Marital Status (Married, Single, Divorced, In a relationship).

- Weight (in kg).

- Height (in cm).

- Hours of work per week.

- Income per month.

Each patient may have children. Each child has the following information:

- Gender.

- Date of birth.

** One patient may be associated with more than one case. These are treated as totally separate cases. Say for example that one patient was diagnosed with depression when he/she was a teenager. The same person could have visited the institute again when he/she got married, because of marital problems. These are considered as two totally different cases associated with the same patient.*

*** You should store the drug's brand name as well as its description. The description contains keywords that will be used later.*

**** The diagnosis should be stored in the database along with the category it belongs to.*

***** The available street values should be stored in the database along with the towns they belong to. The towns should be stored in the database along with the regions they belong to.*

Note: You should take care so as to avoid data redundancy in the way you store your data.

Task 2 (10%):

Write a java program that will automatically generate and insert to the database 10M different cases.

Task 3 (5%):

Write an SQL query to get the average number of therapy sessions per case, out of all recorded cases.

Task 4 (5%):

Write an SQL query to get the average duration of therapy sessions of all the cases that have at least one recorded therapy session in the Q1 of 2015 grouped by their outcome.

Task 5 (10%):

Write an SQL query to get the number of cases whose date of first contact with the institute was in the first 8 months of 2014 grouped by the diagnosis' category. Display the results in a descending order according to the number of recorded cases per category.

Task 6 (10%):

Write an SQL query to get the number of cases that have been prescribed with at least one drug they should take for more than 60 days.

Task 7 (15%):

Create a view that will return the average number of children, of patients who have been prescribed tranquillisers at least once in their lifetime, grouped by the gender of the child. (Tip: use the drugs' description to search for the "tranquilliser" keyword)

Task 8 (15%):

Write a java program that will receive a four digit number (year) as input. The program should return the average income of patients who were born that year and they first contacted the institute between 2005 and 2010, grouped by the BMI category they belong to. (Tip: Check https://en.wikipedia.org/wiki/Body_mass_index for the formula that calculates the BMI index of a person and the way it's split into different categories)

Task 9 (20%):

Write a java program that will receive the code of a region and will return the average working hours per week of the married male patients that have been diagnosed at least once in their lifetime with a Mood Disorder grouped by their number of children.

Bonus Task (+20%):

Keep track of the execution time of your scripts in the MemSQL Server. Then create a dump of your database and destroy your server. Create a new one with the same specs and setup a MySQL Server. Import your dump file and run the same scripts as you did in MemSQL. Create a comparison list with the script execution times. Comment on the results.

Deliverable

Each team is assigned a Group Number.

Your deliverable will be a .zip file.

The .zip should be named as Assignment1_Group_<GROUP_NUMBER>.zip

The .zip should contain:

1) One folder for each task that you have completed. Each folder should contain the code required for this particular task. You may include your Java programs and/or your SQL queries.

2) A .doc file that will include:
   a. A description of your schema
   b. Comments on the answer of each task.
   c. Any extra necessary info/comments on your deliverable
   d. Your server's credentials:
      i. ip
      ii. username
      iii. password

Note: Any code included should run on your server.

Example structure of the deliverable of the team with the group number 15:

```
Assignment1_Group_15.zip
    |-> Assignment1_Group_15.doc
    |-> Task_1 (Folder)
        |-> Schema_Dump.sql
    |-> Task_2 (Folder)
        |-> MyCode.java
    |-> Task_3 (Folder)
        |-> Queries.sql
    |-> Task_4 (Folder)
        |-> Queries.sql
    |-> Task_5 (Folder)
        |-> Queries.sql
    |-> Task_6 (Folder)
        |-> Queries.sql
    |-> Task_7 (Folder)
        |-> Queries.sql
    |-> Task_8 (Folder)
        |-> MyCode.java
    |-> Task_9 (Folder)
        |-> MyCode.java
    |-> Bonus_Task (Folder)
        |-> Comments.doc
```