Database Lab

1   Task1: Application description:

You are a database engineer employed by the Chicago police department. For quite some time, the police have been investigating a gang of elusive bank robbers who have been operating in Chicago district. They have collected quite a lot of information about the gang, some from an informant close to one of the gang members. So far, they have been keeping the data in a set of spreadsheets, but they realize that they cannot do many of the queries they want in the spreadsheet and they are also worried that the data entry is introducing errors and inconsistencies that the spreadsheet does not check for. From the spreadsheets, they have produced a collection of tab-separated files of data. They now want you to convert the data into a well-designed relational database and generate the queries given at the end.

The data files: They contain information about the gang and the banks in the cities where they have been operating:

- banks.txt: lists all the bank branches in the Chicago district. The banks are specified by the name of the bank and the city where the branch is located in. The data file also includes the number of accounts held in the bank (an indicator of size) and the level of the security measures installed by the bank.

- robbers.txt: contains the name (actually, the nickname), age, and number of years spent in prison of each gang member.

- hasaccounts.txt: lists the banks at which the various robbers have accounts.

- hasskills.txt: specifies the skills of the robbers. Each robber may have several skills, ranked by preference what activity the robber prefers to be engaged in. The robbers are also graded on each skill. The file contains a line for each skill of each robber listing the robbers nickname, the skill description, the preference rank (a number where 1 represents first preference), and the grade.

- robberies.data: contains the banks that have been robbed by the gang so far. For each robbery, it lists the bank branch, the date of the robbery, and the amount that was stolen. Note that some banks may have been robbed more than once.

- accomplices.txt: lists the robbers that were involved in each robbery and their estimated share of the money.

- plans.txt: contains information from the informant about banks that the gang is planning to rob in the future, along with the planed robbery date and number of gang members that would be needed. Note that gang may plan to rob some banks more than once.

Each of these files could be converted directly to a relation in the database. However, this would not be a great design because of the following issues.

1.1   The nickname problem

The robbers are currently identified by their nicknames. Although the current list has no duplicates, it is quite possible to have two robbers with the same nickname. It would be better to give each robber a unique Id, and to use the Id for identifying the robber in all the tables. This way, adding a new robber with a duplicate nickname would not require redesigning the whole database schema.

1.2   The skills problem

The list of robber skills uses the descriptions of the skills. There should be a finite set of possible skills, and we would like to ensure that data entry does not misspell skills. Misspelled skills constitute a severe concern since queries might then miss out some answers because of the misspelling. One approach is to define a constraint on the skill attribute of the Has_skills table that checks that every value is one of the possible skills. However, if we then wanted to add a further kind of skill, we would have to change the database schema. A better design can be achieved if we introduce an additional Skills table listing all the possible skills, and define a constraint on the Has_skills table to ensure that every skill there is also in the new Skills table.

1.3 Further assumptions

The banks are identified by their name and city rather than by an Id. The business rules set by the local banking authority ensure that the combination of name and city is unique, so that it is not necessary to create an Id for the banks.

2 Task 2

Now you need to define the database **RobbersGang** using MySQL. Populate the schema using the data files. Create views for each of the following queries-

1. v1: to list the robbers who have spent more than 2 years in jail.

2. v2: to list the robbers who were involved in robbery between 2005 and 2010

3. v3: to list for each skill the senior most robber (in terms of age)

4. v4: to list the banks which were robbed more than once

5. v5: to list the recent most five robberies in Chicago city

6. v6: to list the robbers and the total amount they have robbed so far.

7. v7: to list the robbers and the number of skills they have

8. v8: to list the banks that may be robbed in next 3 months from the current date

9. v9: to list the suspected robbers who may rob in next 3 months. Assume that suspected robbers are those who have accounts at the same bank branch.

10. v10: for each bank, provide a summary of security level and the number of times it was robbed so far.

After creating each view, show the contents using

```
select * from viewname;
```

3 Submission

Write all the relevant MySQL queries that you have used to perform *Task 2*. Also, create a backup file assign5Backup.sql using mysqldump. Use openssl and encrypt the file. Add these commands in your submission file. Submit the queries using a assign5.txt file. Pls, submit the assignment using the following submission link only.

```
http://172.16.1.252/~samrat/CS355/submission/
```