1DV013 – Database Theory

Practical Assignment 2

Deadline: October 6, 2013

Read in the movie data from the previous assignment into your database (use only means of database). It can be done in 2 steps:
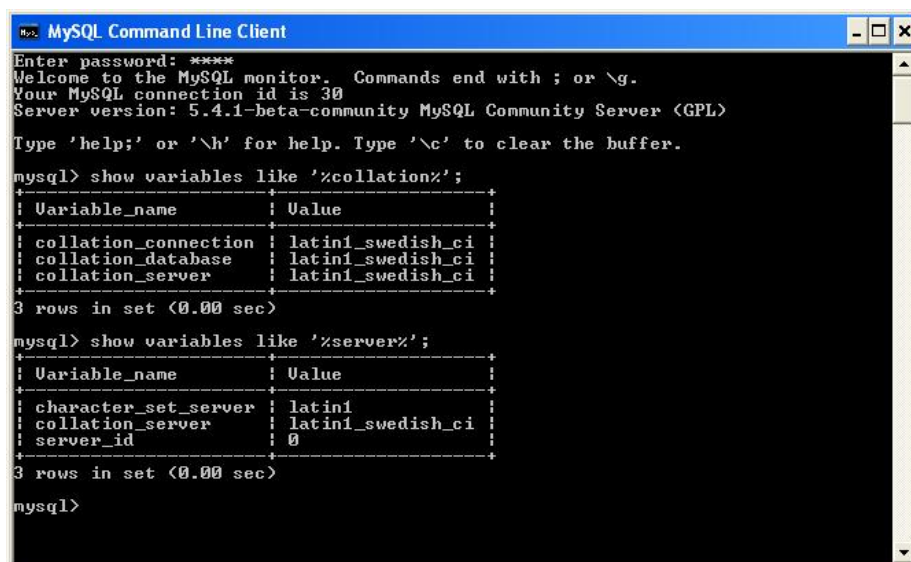
1. Load data into database using temporary created tables. For this, use the Load Data command;
   ```
   mysql> load data infile file_name ignore into table table_name
   fields terminated by '\t';
   ```
   Read what "ignore" and "replace" commands do.

2. Fit the data into your database schema created before. Use Data Manipulation Language (i.e. *insert, delete, update*).

If you run into performance problems, e.g. your data is too big to fit into a table by simple insert query, try to insert data in the process of creating table, e.g.

> create table table_name select * from table_name2;

Don't forget to assign primary keys for tables, it speeds up *join* operation between tables.

If you run into encoding problems while loading the data, make sure that your language for non-unicode programs (Control Panel -> Regional and Language options -> Advanced) corresponds to your database settings. For example, if your regional language is Swedish, then your database setting should look like this:

Figure 1: Encoding settings

To change them read MySql documentation.

If you run into other problems – e.g., timeouts – you need to experiment a little, just like in real life.

When reading in the data, follow these rules:

- A movie is defined uniquely by its *name* and *year*. This allows for, e.g., remakes.

- If you find a movie which is not contained in *movies.txt*, then ignore this movie, i.e., ignore datasets that reference movies not contained in *movies.txt*.

- In *times.txt*, ignore *episodes* and other non-integer information (e.g., "fps") and set the time of the corresponding movies to 0 (for unknown runtime). If a movie has more than one runtime specified, take the first one that you find in the file. For example, "Boy Who Turned Yellow, The 1972" has two times specified, so take "17" here (you can do it by using *ignore* in *load data* command).

Create SQL statements for the following queries:

1. What year did *Kasiping* air?

2. In how many movies has *Abel, Alfred* starred?

3. In which movies that *Aaronson, Evan* has produced did *Beers, Thom* star in?

4. In how many horror movies have people with last name *Clark* acted?

5. Which movies have a play time of 3 1/2 hours or longer?

6. How many movies have been produced, directed, and written by one and the same person?

7. What year was issued the maximum number of movies?

8. How many movies the minimum number of actors acted in?

Hand in your *Load, DML and SQL* statements that populates the database, and the SQL statements together with the answers to the questions.

Assignment must be submitted via Moodle submission system. Email submission attempts will be ignored.