

CSCI 3287: Project 1

Introduction to MySQL

Due Date

9:55p Sun. September 25th, 2016

No late submissions will be accepted.

Delivery Method

You will hand in four files that are contained in a tarball. The first file, named `queries.txt`, will contain the text of the 3 queries you generate. The remaining three files will contain the redirected output generated by each of the three queries.

Description

As an introduction to mysql, you will import and submit queries against a sample employee database provided with mysql. The database contains about 300,000 employee records with 2.8 million salary entries. The database size is approximately 167 MB, which is not huge, but heavy enough to be non-trivial for testing.

This project will use interview grading. 40% of your grade will be based on the submitted materials. The remaining 60% will come from the interview portion. You must schedule an interview slot with a grader. If you do not schedule or attend a slot the interview portion will be scored zero.

Procedure

In this assignment you will import this database, perform various SQL operations on it and submit the resulting output. You can find the database file on Moodle.

You are responsible for getting access to a mysql environment. You can use the CS virtual machine or any other machine you want that you can install mysql on. It is an exercise for the student to install mysql on whatever machine they choose.

In your virtual machine (or whichever Linux environment you are using), install the sample database into mysql. Follow the instructions on this github page for installation.

<https://dev.mysql.com/doc/employee/en/employees-installation.html>

It's a good idea to run the validation script that generates and compares checksums to ensure your installation is correct.

Once you have the database installed, start a mysql client and execute the following commands:

```
mysql> show databases;
```

There should be an `employees` database listed.

```
mysql> use employees;
mysql> show tables;
+-----+
| Tables_in_employees |
+-----+
| current_dept_emp      |
| departments           |
| dept_emp              |
| dept_emp_latest_date  |
| dept_manager          |
| employees             |
| salaries              |
| titles                |
+-----+
```

You are encouraged to explore the database. Throw some generic queries, e.g. `select *` from, against various tables to look at their contents. You can use the `describe` query to explore the tables schemas or refer to the documentation on the `mysql` website for a description of the schemas.

You will create queries and redirect their output to files to be submitted on Moodle. To redirect the output of your queries to a file, append the following clause to your queries:

```
INTO OUTFILE <filename> FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n'
```

You need to develop and execute the following queries:

1. We want to recognize our newest employees. Create a query that contains all the employees that have been with the company for less than a month. Assume the last employee hired was hired today. Create an output file called **new_hires.csv** for this query.
2. Create a view called **high_salaries** of our high salary employees. A view can be created as follows:

```
mysql> create view <name> as <query>;
```

Define a high salary to be any employee making greater than \$120,000.00 annually. This view should appear as a table in `show tables`; when you are finished. Once created, this view can be used like any other table. After creating the view, execute the query: `select * from high_salaries` and redirect the output to **high_salaries.csv**.

3. Create a view called **high_salary_names** that is a list of the first and last names of our high salary employees. The list should contain no duplicates and be in sorted order.

After creating the view, execute the query: `select * from high_salary_names` and redirect the output to `high_salary_names.csv`.

When you are finished, copy the text of each query you came up with into a file called `queries.txt`. Then put `queries.txt`, `new_hires.csv`, `high_salaries.csv`, and `high_salary_names.csv` into a directory and create a tarball. If the files are in a directory, `test`, `cd` to the parent directory and create the tar archive:

```
~/fwmiller/test/$ cd ..  
~/fwmiller/$ tar czvf test.tgz test
```

This will create a tarball called `test.tgz` that you should submit to Moodle.