Database and Big Data Systems - Homework 1
Due: 6/10/2021 11.59pm

In this homework, you will be working with a small dataset from the Bureau of Transportation Statistics. You will use the file named `flights.zip` that is uploaded with this homework on eDimension.

Loading Data [1 points]

Create and load the following tables from the dataset (drop existing tables with the same names). You need to take a look at the data and determine the type for each attribute.

```
flights(fid, year, month_id, day_of_month, day_of_week_id, carrier_id,
        flight_num, origin_city, origin_state, dest_city, dest_state,
        departure_delay, taxi_out, arrival_delay, cancelled,
        actual_time, distance)
carriers(cid, name)
months(mid, month)
weekdays(did, day_of_week)
```

The primary keys for these tables are `fid, cid, mid`, and `did`, respectively.

SQL Queries [5 points]

Write SQL queries for the following.

Q1. Find distinct flight numbers for flights from Seattle WA to Boston MA, operated by Alaska Airlines Inc on Monday.

Q2. Find the top 3 days of the week with the longest average arrival delay. Print name of the day, and average delay.

Q3. Find the names of airlines that operate more than 1000 flights in one day (any day). For example, the result may contain an airline that only on 3rd Dec 2005 has > 1000 flights, but for the rest of the days has < 1000 flights per day.

Q4. Find the total departure delay per airline over the entire dataset. Print out airline name and the delay.

Q5. Find airlines that have flights out of New York NY and have cancelled more than 0.5% of the time. List in ascending order of the percentage.

Submission

Please submit a single file named `hw1.sql` to eDimension. This file is run against our test MySQL database, and it should load the data and execute the five queries. You are supposed to make use of or make reference to the given template `hw1_template.sql`.

The first few lines of this file should contain the following

```
-- STUDENT number. UPDATE ONLY, DO NOT DELETE.
select "1001234";
-- Replace the above with your student number
```

Please do not delete it and update the student number according to yours.

The each question, put your answer in the corresponding section marked with the following comments and statement.

```
-- DATA LOADING Seperator. DO NOT DELETE
select "DATA LOADING";
-- Put your table creation annd data loading SQL statements here

-- QUESTION 1 Seperator. DO NOT DELETE
select "QUESTION 1";
-- Put your Q1 SQL statements here
```

The file `flights.zip` is unzipped into a folder that is readable by `mysql`, e.g. `/tmp/sql/`. The `/tmp/sql/` folder contain no subfolder.

You may test your submission on your machine by running

```
mysql -u root -b test < hw1.sql
```

where `test` is a database name. You must ensure there are no errors.

Troubleshooting

If you encounter the following error when loading the data

```
ERROR 1290 (HY000): The MySQL server is running with the --secure-file-priv option
so it cannot execute this statement
```

1. append the following line to the file

   - Ubuntu: `/etc/mysql/mysql.conf.d/mysqld.cnf`
   - MacOS X (brew): `/usr/local/etc/my.cnf`

   ```
   secure-file-priv = "<folder where the CSV files are located>"
   ```

   e.g.

   ```
   secure-file-priv = "/tmp/sql/"
   ```

2. then restart mysql service. Make sure that the above folder is accessible by MySQL client.