Homework 2

Group 5

TA: 葉承翰

E-maill: jerry.yeh0531@gmail.com

Introduction

In this homework, you have to load a data set into your database and use SQL to answer the following questions. Here is the link for downloading the data set.

Dataset:

Time series data from sensors

Descriptions:

This data set consists of records produced by multiple sensors of different organizations from 3/2 to 3/5. There are three tables you have to establish before answering the questions.

Tables:

sensors: information of sensors

sid	the id of a sensor
lat	latitude of a sensor
lon	longitude of a sensor
org	the organization that a sensor belongs to

records: the data recorded by sensors

rid	the id of a record
date	the date when the record is created
time	the time when the record is created
h	humidity
temp	temperature
pm1	pm1 data
pm10	pm10 data
pm2.5	pm2.5 data

send: the producer(sender) of the records

sid	the id of a sensor
rid	the id of a record

Questions:

Answers should contain SQL, time, and the results except question 1.

1. Show the tables(schema). You have to follow the schema in previous page and show the datatype of each attribute.
2. How many distinct sensors are there in this data set?
3. Calculate the contribution(number of records) of sensor "74DA38AF48A8" in this data set.
4. Find the record which has the highest pm2.5 value in 3/3.
5. Calculate the percentage of "Airbox" sensors in this data set.
6. From 3/4 to 3/5, compute the average temperature for each sensor.
7. List top 10 sensors that has highest pm2.5 value for each day.
8. Compute the average interval for sensor "28C2DDDD4797" in 3/2. (Suppose A sensor has *n* records in 3/2, then it has *n-1* intervals. Hence, the average interval is compute as sum(*n-1* intervals) / (*n-1*))
9. Count the number of records that they are produced in an area which its coordinates are lat(24.781466~24.801441), lon(120.987374~121.004073).
10. Find the sensor which has the largest interval in 3/1.

Bonus.

Find something that is interesting, meaningful or useful using SQL query, and explain your idea.

Note: Remember to add a column in sensors table called org when you load csv files into your database. The value of the org equals to filename. ex. If a file with name "xxx.csv", then all the data in that file should add a column "org" which it's value is "xxx".

NCTU CSCC Database Service
- https://www.cs.nctu.edu.tw/cscc/mysql/

Due Date: 23:59, April 10, Monday, 2017

If you have any questions, feel free to contact me.