

Homework 2

SQL Tasks in MySQL

Goal:

You need to do some SQL query in “Airline” dataset
(<http://stat-computing.org/dataexpo/2009/>)

Dataset:

1. Download the data from **2004** to **2008** year (totally five years)
(<http://stat-computing.org/dataexpo/2009/the-data.html>)
2. Download **airports.csv** and **plane-data.csv**
(<http://stat-computing.org/dataexpo/2009/supplemental-data.html>)

Table Schema:

You can get schema here:

<http://stat-computing.org/dataexpo/2009/sqlite.html>

Query Task:

【Every task need to contain the **SQL**, **time** and the **answer**】

Example : 0. How many records in the dataset

SQL: SELECT COUNT(*) FROM ontime Time: 86.71592 sec Ans: 35,874,731

1. Please list the number of all different route (origin → destination) and contain “**DISTINCT**” syntax in usage of SQL
請列出所有不同飛行航線 (起點→終點) 的數量，請包含 **DISTINCT** 語法
ex: 1000
2. Please list the number of flights which destination is JFK and the actual flight time is between 1 to 3 hours and contain “**BETWEEN**” syntax in usage of SQL
請列出目的地為 JFK 且實際飛行時間介於 1~3 小時的航班數量，請包含 **BETWEEN** 語法
ex: 1000

3. Please list both the aircraft type name and amount which aircraft type name beginning with "737" and contain "LIKE" syntax in usage of SQL

請列出機型名稱開頭為 737 的機型與其數量，請包含 LIKE 語法

ex: 737-724, 34

737-401, 12 ...

4. Please list the average speed(use the actual arrival time) of all aircraft type and contain "AVG" in usage of SQL

請列出所有機型的平均速率(用實際時間算)，請包含 AVG 語法

(Compare the time if you "create index TailNum on Plane(TailNum)")

ex: A330-223, 6666.652

737-401, 6551.656 ...

5. Please list the top 50 time zone count and route (origin → destination) by the count of routes which across time zone(assume one time zone is formed by 15 degree longitude and not related to country) in decrease order and contain "ORDER BY" in usage of SQL

請根據跨越的時區(假設每 15 度為 1 個時區，不看國家)數量由大到小列出前五十名航線(起點→終點)與數量，請包含 ORDER BY 語法 (若數量一樣根據航線(起點→終點)名稱以字典排序)

ex: 4, ABQ, DAL

4, ABQ, ELP

2, AMA, DAL ...

6. Please list the aircraft type which never appear after 2009 (appeared before)and contain "NOT IN" syntax in usage of SQL

請列出在 2009 年再也沒出現的機型 (在之前出現過) 請包含 NOT IN 語法

ex: A330-223

737-401 ...

7. Please list the manufacturer which airplanes' delay time is more than half an hour

找出飛機平均延誤時間大於半小時的製造廠商

ex: BOMBARDIER INC

CANADAIR ...

8. Please list the average delay time of week (7 day) and time (24 hour), and when is the best time to fly (that has minimum delay)

請列出一週 7 天以及一天 24 小時中的飛機平均延誤時間，什麼時間搭飛機遇到的 Delay 會最少？

ex: Monday, 10	→ Best day	01hour, 18	→ Best Time
Tuesday, 11		02hour, 19	
...		...	
Sunday, 17		24 hour: 20	

9. Do older planes suffer more delays? (Answer by your own view)

比較老的飛機 Delay 時間會比較久嗎? (此題為開放式答案，請利用 SQL 找出的結果闡述你的觀點)

10. How well does weather or season predict plane delays? (Answer by your own view)

天氣或是季節會影響 Delay 時間嗎? (此題為開放式答案，請利用 SQL 找出的結果闡述你的觀點)

11. (Bonus) Can you detect cascading failures as delays in one airport create delays in others? Are there critical links in the system?

請找出飛機發生的連環錯誤的 critical links (ex:因為某班的 Delay 造成另一個航線也 Delay)

12. (Bonus) Feel free to think. Any valuable observation with explanation

請自由發想 SQL，找出有趣的資訊並解釋它的含意

Submission:

➤ At first, please record the OS, RAM and storage of your computer, for example:

- OS : Ubuntu 14.04 LTS
- RAM : 8GB
- Storage : 500GB

Then, please answer the 10+2 query tasks that contain the **SQL**, **time** and the **answer of question** information

➤ Be sure to upload your homework to E3 by due date and name your file as "HW2_XXXXXXX.pdf", where XXXXXXX is your student ID.

Deadline:

➤ 2016/4/22 23:59