Database Homework 4 Report

陳冠穎 0316317

A. Database Design

The first part is to import the CSV file.

Since airport name is from A–Z (two digit number in ASCII code) , I using a simple hash function that combine airport character in ASCII, for example ABC–>656667. Therefore by scanning original CSV, I can get three integer (Arrival Delay, Origin, Dest) for every record.
Then write the data in to database file.

Since every record contain only three integer. Therefore I can write three integer sequentially in binary form. For example, I want to get the third record. Then I can seek to 12*(3–1)=24 bytes from database file and read 12 bytes. First 4 bytes is delay, then origin and dest.

B. Index Design

I use unordered_map as hash table. The key of unordered_map is a Long Long number combining of origin and dest. For example, origin–> 656667, dest–>717273, then key –> 656667717273. The value of unordered_map is a vector of integer which save the index number of record in database file. Therefore when creating index, iterating the whole database file, pushing the position of record to the corresponding vector (key=origin+dest). To query with index, just find the corresponding vector and iterate the position and get the value from database file.

C. Screenshot

```
┌[kevinchen@KevinChenMBP] - [~/dev/DB_HW4] - [2016-06-18 10:37:06]
└[0] <git:(master 1480b88x**)> ./test
# 1
Time taken for importing data: 12.8326s
Time taken for querying 1 without index: 18.3168s
Time taken for querying 1 without index: 18.7849s
Time taken for querying 1 without index: 18.6239s
Time taken for querying 1 without index: 18.2674s
Time taken for querying 1 without index: 18.6995s
*1: 17.2076
*2: 11.479
*3: 11.2612
*4: 12.6459
*5: 7.19692
Time taken for query no index: 92.6926s
Time taken for creating index: 24.6913s
Time taken for querying 1 without index: 0.0039s
Time taken for querying 1 without index: 0.0141s
Time taken for querying 1 without index: 0.0257s
Time taken for querying 1 without index: 0.0035s
Time taken for querying 1 without index: 0.0112s
1: 17.2076
2: 11.479
3: 11.2612
4: 12.6459
5: 7.19692
Time taken for making queries: 0.0584s
100
Time taken for making 500 queries: 6.1831s
Avg Time taken per five query: 0.06183149s
```

D. Result Table

Origin	Dest	Avg Arrdelay	Query time (no index)	Query time (with index)	Index Time
IAH	JFK	17.2076	18.3168s	0.0039s	
IAH	LAX	11.479	18.7849s	0.0141s	
JFK	LAX	11.2612	18.6239s	0.0257s	24.6913s
JFK	IAH	12.6459	18.2674s	0.0035s	
LAX	IAH	7.19692	18.6995s	0.0112s	