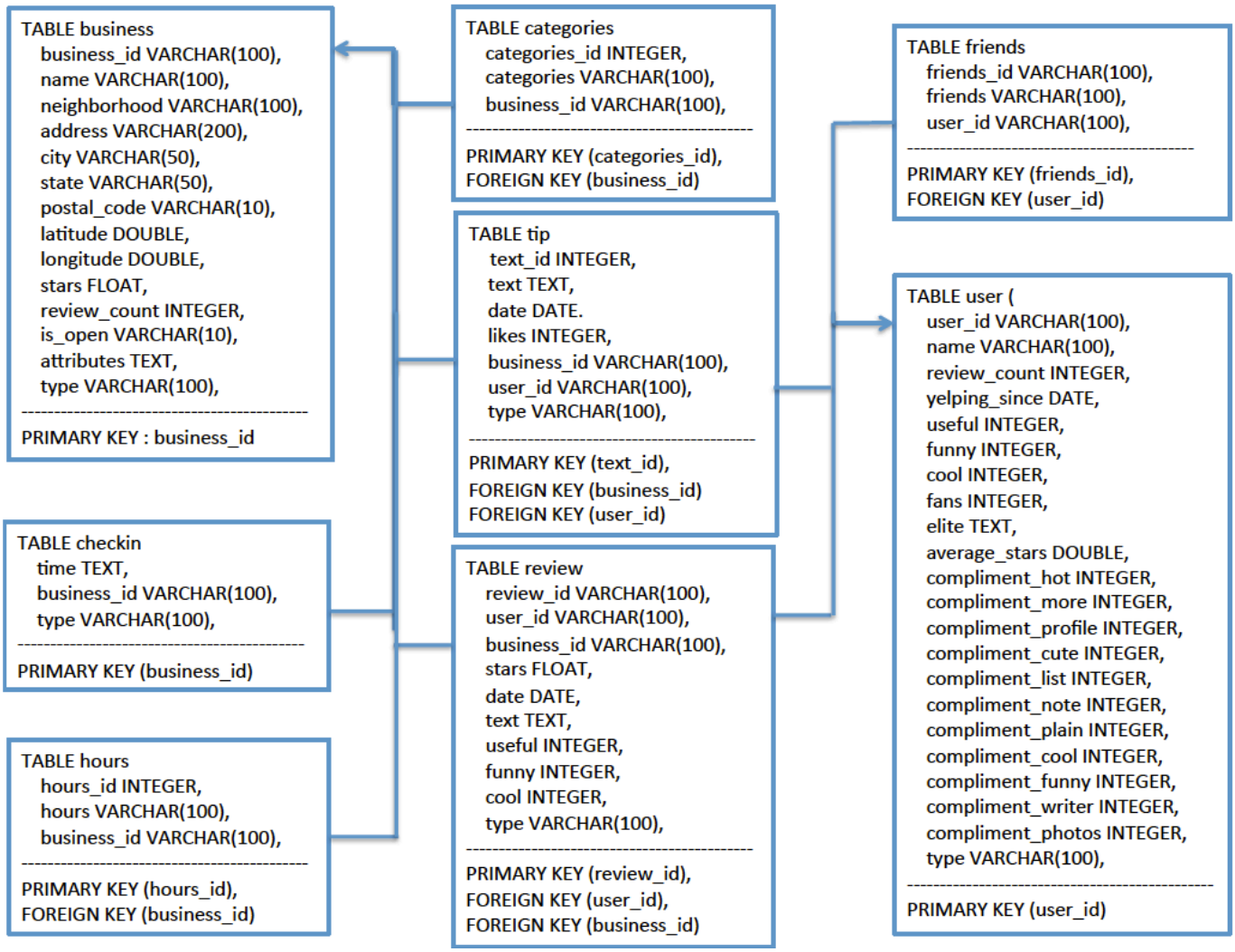
ECE 656
Shala Chen
Feb 28th 2017

Assignment 1 part 3 (a)(b)

I used python 2.7 and imported sys, csv and json.

a.i: relational schema



a.ii:

I separated categories and hours from business table and link them back to as foreign keys. I also took friends from user table and set up an individual table for it.

- Table business has a primary key “business_id”, which it's unique.
- Table checkin has a primary key “business_id”, which it’s unique, and it also serves as foreign key to “business_id” in table business.
- Table hours has a primary key “hours_id” which is incremental and unique. The foreign key “business_id” refers to “business_id” from table business. Both of them are unique.
- Table categories has a primary key “categories_id”, which is incremental and unique. The foreign key refers to “business_id” from table business.
- Table tip has a primary key “text_id” which is incremental and unique, the foreign key “business_id” refers to “business_id” from table business, the foreign key “user_id” refers to “user_id” from table user.
- Table review has a primary key “review_id” which is incremental and unique, the foreign key “business_id” refers to “business_id” from table business, the foreign key “user_id” refers to “user_id” from table user.
- Table friends has a primary key “friends_id” which is incremental and unique, the foreign key “user_id” refers to “user_id” from table user.
- Table user has a primary key “user_id”, which it's unique.

d.

Size and loading time of each table

Table	Rows	Load time
Business	144072	5.28s
Categories	527604	10.47s
Hours	665535	12.94s
Review	4153150	4h3m33.18s
User	1029432	37.27s
Friends	29723247	1h17m20.44s
Checkin	125532	8.38s
Tip	946600	9.69s

business

```
mysql> source /Users/shalachen/Desktop/yelp/businessload.sql
Query OK, 0 rows affected (0.01 sec)

Query OK, 0 rows affected (0.02 sec)

Query OK, 144072 rows affected (14.14 sec)
Records: 144072  Deleted: 0  Skipped: 0  Warnings: 0
```

categories

```
mysql> source /Users/shalachen/Desktop/yelp/categoriesload.sql
Query OK, 0 rows affected (0.92 sec)

Query OK, 0 rows affected (0.03 sec)

Query OK, 527604 rows affected (56.68 sec)
Records: 527604  Deleted: 0  Skipped: 0  Warnings: 0
```

hours

```
mysql> source /Users/shalachen/Desktop/yelp/hoursload.sql
Query OK, 0 rows affected (0.02 sec)

Query OK, 0 rows affected (0.03 sec)

Query OK, 665535 rows affected (1 min 44.40 sec)
Records: 665535  Deleted: 0  Skipped: 0  Warnings: 0
```

user

```
mysql> source /Users/shalachen/Desktop/yelp/userload.sql
Query OK, 0 rows affected (0.26 sec)

Query OK, 0 rows affected (0.03 sec)

Query OK, 1029432 rows affected (4 min 59.54 sec)
Records: 1029432  Deleted: 0  Skipped: 0  Warnings: 0
```

checkin

```
mysql> source /Users/shalachen/Desktop/yelp/checkinload.sql
Query OK, 0 rows affected (0.29 sec)

Query OK, 0 rows affected (0.06 sec)

Query OK, 125532 rows affected (1 min 44.66 sec)
Records: 125532  Deleted: 0  Skipped: 0  Warnings: 0
```

tip

```
mysql> source /Users/shalachen/Desktop/yelp/tipload.sql
Query OK, 0 rows affected (0.14 sec)

Query OK, 0 rows affected (0.02 sec)

Query OK, 946600 rows affected (9.98 sec)
Records: 946600  Deleted: 0  Skipped: 0  Warnings: 0
```

Load time is a little bit different, it was much slower when I ran them all together when I wanted to take screen shot.