Alla Topp
MSDS 696 - Data Science Practicum

How to create tables in yelpdb and import csv files in PostgreSQL

To create a new table in PostgreSQL, the CREATE TABLE statement is used. Next, list the column name, its data type, and column constraint. To import data into created table COPY statement is used.

- Specify the table with column names after the COPY keyword. The order of the columns must the same as in the CSV file.
- Put the CSV file path after the FROM keyword. Since CSV file format is used, DELIMITER as well as CSV keywords should be specified.
- HEADER keyword indicates that the CSV file contains a header line with column names. When importing data, PostgreSQL ignores the first line, which is the header line of the file.

Since the yelp files were stored on Ubuntu VM in home directory, use /home/ubuntu/…

1. Creating table business

```
yelpdb=# CREATE TABLE business (
yelpdb(#     id SERIAL PRIMARY KEY,
yelpdb(#     business_id VARCHAR(100) NOT NULL,
yelpdb(#     name VARCHAR(100) NOT NULL,
yelpdb(#     address VARCHAR(255),
yelpdb(#     city VARCHAR(50),
yelpdb(#     state VARCHAR(5),
yelpdb(#     postal_code VARCHAR(50),
yelpdb(#     latitude FLOAT,
yelpdb(#     longitude FLOAT,
yelpdb(#     stars FLOAT,
yelpdb(#     review_count INT,
yelpdb(#     is_open BOOLEAN,
yelpdb(#     attributes TEXT,
yelpdb(#     categories TEXT,
yelpdb(#     hours VARCHAR(255)
yelpdb(# );
CREATE TABLE
yelpdb=# COPY business(address,attributes,business_id,categories,
yelpdb(# city,hours,is_open,latitude,longitude,name,postal_code,
yelpdb(# review_count,stars,state)
yelpdb-# FROM '/home/ubuntu/business.csv' DELIMITER ',' CSV HEADER encoding 'UTF
8';
COPY 192609
```

2. Creating table checkin

```
yelpdb=# CREATE TABLE checkin (
yelpdb(#     business_id VARCHAR(100) NOT NULL,
yelpdb(#     date TEXT,
yelpdb(#     PRIMARY KEY (business_id)
yelpdb(# );
CREATE TABLE
yelpdb=# COPY checkin(business_id, date)
yelpdb-# FROM '/home/ubuntu/checkin.csv' DELIMITER ',' CSV HEADER encoding 'UTF8
';
COPY 161950
```

3. Creating table tip

```
yelpdb=# CREATE TABLE tip (
yelpdb(#     id SERIAL PRIMARY KEY,
yelpdb(#     user_id VARCHAR(100) NOT NULL,
yelpdb(#     business_id VARCHAR(100) NOT NULL,
yelpdb(#     text VARCHAR(2000),
yelpdb(#     date DATE,
yelpdb(#     compliment_count INT
yelpdb(# );
CREATE TABLE
yelpdb=# COPY tip (user_id, business_id, text, date, compliment_count)
yelpdb-# FROM '/home/ubuntu/tip.csv' DELIMITER ',' CSV HEADER encoding 'UTF8';
COPY 1223094
```

4. Creating table user

```
yelpdb=# CREATE TABLE y_user (
yelpdb(#     user_id VARCHAR(100) NOT NULL,
yelpdb(#     user_name VARCHAR(50),
yelpdb(#     review_count INT,
yelpdb(#     yelping_since TIME,
yelpdb(#     useful INT,
yelpdb(#     funny INT,
yelpdb(#     cool INT,
yelpdb(#     elite VARCHAR(255),
yelpdb(#     friends TEXT,
yelpdb(#     fans INT,
yelpdb(#     average_stars FLOAT,
yelpdb(#     compliment_hot INT,
yelpdb(#     compliment_more INT,
yelpdb(#     compliment_profile INT,
yelpdb(#     compliment_cute INT,
yelpdb(#     compliment_list INT,
yelpdb(#     compliment_note INT,
yelpdb(#     compliment_plain INT,
yelpdb(#     compliment_cool INT,
yelpdb(#     compliment_funny INT,
yelpdb(#     compliment_writer INT,
yelpdb(#     compliment_photos INT,
yelpdb(#     PRIMARY KEY (user_id)
yelpdb(# );
CREATE TABLE
yelpdb=# COPY y_user(user_id, user_name, review_count, yelping_since, useful, fu
nny,
yelpdb(# cool, elite, friends, fans, average_stars, compliment_hot, compliment_m
ore,
yelpdb(# compliment_profile, compliment_cute, compliment_list, compliment_note,
yelpdb(# compliment_plain, compliment_cool, compliment_funny, compliment_writer,
 compliment_photos)
yelpdb-# FROM '/home/ubuntu/user.csv' DELIMITER ',' CSV HEADER encoding 'UTF8';
COPY 1637138
```

5. Creating table review

```
yelpdb=# CREATE TABLE review (
yelpdb(#     review_id VARCHAR(100) NOT NULL,
yelpdb(#     user_id VARCHAR(100) NOT NULL,
yelpdb(#     business_id VARCHAR(100) NOT NULL,
yelpdb(#     stars FLOAT,
yelpdb(#     useful FLOAT,
yelpdb(#     funny FLOAT,
yelpdb(#     cool FLOAT,
yelpdb(#     text TEXT,
yelpdb(#     date DATE,
yelpdb(#     PRIMARY KEY (review_id)
yelpdb(# );
CREATE TABLE
yelpdb=# COPY review(review_id, user_id, business_id, stars, useful, funny, cool
, text, date)
yelpdb-# FROM '/home/ubuntu/review1.csv' DELIMITER ',' CSV HEADER encoding 'UTF8
';
COPY 5764393
```

Now all 5 csv files of yelp data are in the psql table ready for queries!

References

2.3. Creating a New Table. (n.d.). Retrieved from https://www.postgresql.org/docs/8.2/tutorial-table.html

How to List Databases and Tables in PostgreSQL Using psql. (n.d.). Retrieved from https://chartio.com/resources/tutorials/how-to-list-databases-and-tables-in-postgresql-using-psql/

Import CSV File Into PostgreSQL Table. (n.d.). Retrieved from https://www.postgresqltutorial.com/import-csv-file-into-posgresql-table/

Import CSV File Into PostgreSQL Table. (n.d.). Retrieved from https://www.postgresqltutorial.com/import-csv-file-into-posgresql-table/

Importing Data from CSV in PostgreSQL. (n.d.). Retrieved from https://dataschool.com/learn-sql/importing-data-from-csv-in-postgresql/

PostgreSQL Boolean Data Type with Practical Examples. (n.d.). Retrieved from https://www.postgresqltutorial.com/postgresql-boolean/

Ratanapintha, M. (n.d.). Getting Started with PostgreSQL. Retrieved from https://courses.cs.washington.edu/courses/csep544/11au/resources/postgresql-instructions.html