Athens University of Economics and Business
Department: Computer Science
Course: Databases
Academic Year: 2019-2020
Instructor: Vasilis Vassalos

2nd Assignment: Airbnb Staging Tables

Goals:

In this first assignment, we will create a database for storing and exploring Airbnb data. The database will consist of a collection of tables and their relationships. We will define the schema based on the format of the input data and visualize it through an ERD. We will create the tables in SQL and load them with the \copy command. We will also create foreign key constraints for referential integrity.

Input:

https://drive.google.com/file/d/1omHRHh8IGSaRKydFO9qPfaZWtmtCLYSt/view

Desired Outputs:

- ER Diagram of the Airbnb tables, relationships and attributes
- Airbnb database created in Postgres Cloud instance, or locally.
- Airbnb tables created, 1 table per csv file.
- Airbnb tables populated with input data.
- Airbnb tables contain primary key and foreign key constraints.

Tools you Need:

- Draw.io
- Postgres psql client / pgAdmin

Programming Style:

- The first letter of each table must be in uppercase (e.g. Calendar, etc.).
- The attribute names must be in lowercase (e.g. host_id, etc.).
- Place each *create table* statement into its own file. For example, create_calendar.sql.

- Create a main script, create_tables.sql, that calls the individual create table scripts that match each csv file.
- Place all alter table statements into a single file, alter_tables.sql.
- ER should be done in Draw.io and saved locally. Name the file airbnb_ERD.png

Hints:

- Connect to your newly created database, via psql shell or pgAdmin in order to create the tables corresponding to each csv file.
- Use **gen_ddl_python3.py** (for Python3) to auto-generate the initial create table statement. This python script takes as an argument a .csv file e.g. listings.csv for Listings table and produces a .sql file that contains the create table script for this specific table. Note that some manual adjustments have to be made in the sql file, like the primary key constraints.
- Use the command \i <filename> in psql to run a SQL command from a script. For example, \i create_tables.sql. Alternatively, in pdAdmin choose your database, press the query tool button and run a sql script as follows: press the "Open file" button, import the sql script from your computer and then press the "thunder" image in the query tool bar.
- Use the \copy command in psql or the Import/Export process in pgAdmin to import all data from the csv files. Note that the first line of each csv file is a header. Set a parameter in the copy command to avoid the headers. For example, \copy Listing FROM 'airbnb dataset/listings.csv' DELIMITER ',' CSV HEADER;
- Before running copy, run the command set client_encoding to 'utf8'; in your psql client to avoid character encoding issues.
- Add the foreign key constraints after the tables have been populated.
- The Listings table has ~90 attributes, choose a small subset of attributes to include in your ERD.

Useful Links:

- Create table command: https://www.postgresql.org/docs/9.6/sql-createtable.html
- Copy command: https://www.postgresql.org/docs/9.6/sql-copy.html
- Alter table command: https://www.postgresql.org/docs/9.6/sql-altertable.html
- Postgres meta commands, like \copy: https://www.postgresql.org/docs/9.2/app-psql.html
- pgAdmin query tool: https://www.pgadmin.org/docs/pgadmin4/dev/query_tool.html

- pgAdmin import:
  https://www.pgadmin.org/docs/pgadmin4/dev/import_export_data.html

Additional Notes:

All .sql files and the ER diagram must be in the same folder.