

CSE 412: Database Management

Group Project Phase 2

Generating Data

To generate the data, just run `./datagen.sh` (after reading this readme). This will do several things:

- Prompt for a database username and password.
- Run `schema.sql`, which will:
 - Create a new database, called CSE412, using `schema`. ***If this database exists already, this script will delete it.***
 - Create the tables in the database, according to the schema.
- Unzip the words lists, which gives several text files. All of them start with a prefix `data_`. If these files already exist in the folder, the previous copy will no longer exist.
- Run the data generation python script, based on the data files. This uses a seed so that the same random data can be generated again.
- Delete the data files. Note that the `.zip` file will remain.

Data generation took a little less than five minutes on my computer.

The following numbers of rows were generated for the following tables:

Album Name	Number of Rows
album	46616
band	4390
band_member	24117
concert	71923
featured_artist	13448
instrument	38
person	61243
plays	13127
song	698519
staff	23999
tour	13111
Total	970531

