SYSTEMS DEVELOPMENT FOR COMPUTATIONAL SCIENCE

LECTURE 22

Fabian Wermelinger
Harvard University
CS107 / AC207

Tuesday, November 15th 2022

LAST TIME

- Introduction to databases
- Data models
- Structured query language (SQL)
- Python and SQLite

TODAY

Main topics: ***Databases, SQL and SQLite***

Details:

- Hands-on exercises using Python and `sqlite3`:
  - Reading data into tables
  - Queries
  - Sorting
  - Selecting columns
  - Altering tables
  - Aggregation
  - Deleting rows

AGENDA CHECK:

- Milestone 2 deadline has been moved to *Tuesday, November 22nd 11:59pm*. You can find the milestone details at https://harvard-iacs.github.io/2022-CS107/project/M2.

SQLITE EXERCISE I

- The exercise sheet and data is located in the class repository: https://code.harvard.edu/CS107/main/tree/master/lecture/code/lecture22
- We will work in a Jupyter notebook (`lecture22.ipynb`) for this exercise.

Deliverables:

1. Copy the Jupyter notebook along with `candidates.txt` and `contributors.txt` into `lab/pp12` in your private Git repository and commit on your default branch.
2. For each step in the exercise notebook, there are instructions labeled ***"Do the following:"*** (except for Setup and Interlude). Put all the code for those instructions in a *code cell(s) immediately following the instructions*. The code in that cell should be regular Python code. You should place comments where appropriate that describe your intentions.
   → ***Note:*** to get the `pandas` tables to display in a cell, use `display()`.
3. Save and close your database. Be sure to upload your database in `lab/pp12` as well. Please name your database `lecture22.sqlite`.

RECAP

- Hands-on exercises using Python and `sqlite3`:
  - Reading data into tables
  - Queries
  - Sorting
  - Selecting columns
  - Altering tables
  - Aggregation
  - Deleting rows

Further reading:

- E. F. Codd, *"A relational model of data for large shared data banks"*, ACM, 1970 (https://dl.acm.org/doi/10.1145/362384.362685, pdf also in class repository)
- SQLite home: https://www.sqlite.org/index.html
- SQLite docs: https://www.sqlite.org/doclist.html
- SQLite3 module in Python (and tutorial): https://docs.python.org/3/library/sqlite3.html
- James R. Groff, Paul N. Weinberg and Andrew J. Oppel, *"SQL The Complete Reference"*, McGraw-Hill, 3rd edition, 2010