

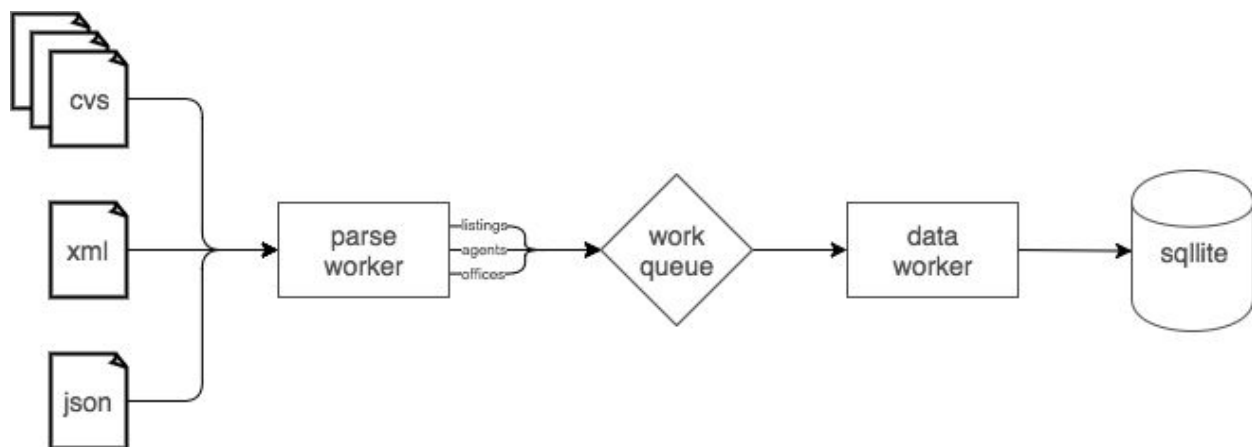


Homes.com SE Applicant Project

Data Processing Pipeline Simulation

You will be given a set of data files and a SQLite database (An SQLite browser can be downloaded from <http://sqlitebrowser.org/>) that only has table structure without data. The goal of the project is to build a processing system that will take the data in the data files and push it into the appropriate tables in the SQLite db.

The data files will be split into three separate categories, each representing a different “MLS board”. Each board will use one of three styles of data representation (CSV, XML, JSON).



(Fig. 1) Control Flow Diagram

According to the flow diagram above, you will be required to create three applications:

- Parse Worker -- this application will take existing source data files, read them, and send the data to a work queue (described below).
- Work Queue -- this application will take in a set of data as a job, storing it, and allowing for later retrieval of the data exactly as it was stored.
- Data Worker -- this application will retrieve jobs from the work queue and push the data contained in those jobs into the correct tables in the database.

Once these applications have been run successfully against a clean SQLite database, the data will be evaluated against expected results.

You are expected to complete this project on your own, using digital resources available to you. It should reflect your personal experience, ingenuity and coding style.

You should provide your code, example output, and any needed instructions via email prior to your interview.