Jeffrey Burmood

ISDA 608

May 11, 2016

Final Project Description

My company is a small startup working to develop, manufacture, and sell Atmospheric Water Generators (AWGs) which are devices that condense liquid water out of the surrounding air. AWGs have an effective range of ambient temperature and relative humidity where they can generate liquid water.

This project will perform the first step in an AWG market analysis by collecting weather data from across the US, processing the data, creating water generation estimates, and generating visualizations of estimated monthly water output to assist in identifying the best candidate locations for AWG sales based on ambient weather conditions.

The dataset to be used for the project is historical weather data from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI). The link to be used is:

http://www.ncdc.noaa.gov/orders/qclcd/

The dataset is comprised of a series of ZIP files, where each ZIP file contains detailed weather data files for over 2,100 reporting weather stations across the US. Each ZIP file contains monthly weather and station data dating from July 2007 to the present. The specific data files to be used in the project are the monthly weather data and the station data.

A Python script is used to download the ZIP archive files from the NOAA site into a local directory. The script then extracts just the weather and station data from the archive file, removes any rows with missing data, performs some basic data validation, and then stores the data into a MongoDB database.

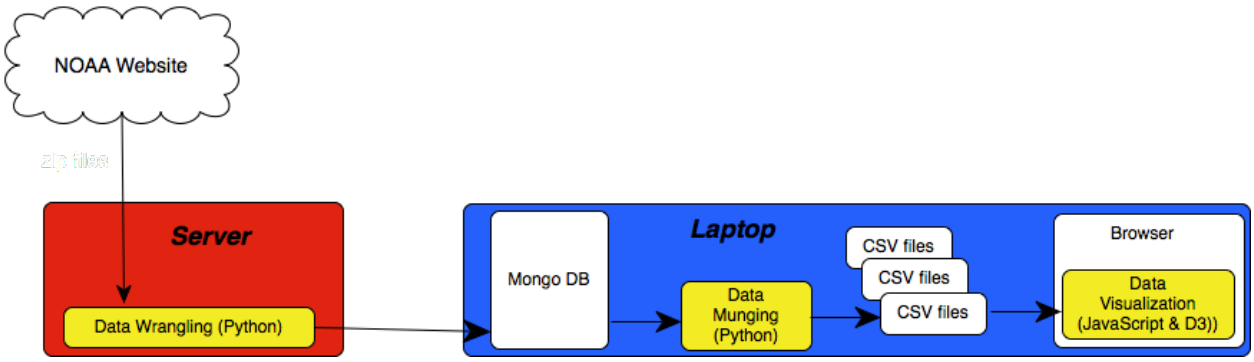
A separate Python script is then used to retrieve the weather and station data from MongoDB, calculate the relative humidity value for each row of the monthly weather data using the retrieved average air temperature and average dew point temperature values, and then, using the monthly average relative humidity data, calculate the estimated monthly water production in liters. Finally, a CSV file is created for each month's data and stored in the local directory. Normally, a back-end server script would be used to retrieve data from the MongoDB database for an HTML client, but a series of CSV files are used to simplify the project for now.

An HTML file contains all of the HTML and Javascript code to allow the user to load and display the estimated water production data for all reporting weather stations for a selected month and year. The data for the weather visualization is retrieved from the appropriate CSV file.

The estimated monthly water production calculation is based on a competitor's published monthly water production claims for a similar system based on the ambient relative humidity value.

RUN NOTEs:

Please refer to the following figure.



Back-End

The project “back-end” is comprised of the following two Python 2.7 scripts:

- DataWrangling.py - download weather data from the web and place in the DB
- DataMunging.py - extract data from DB and create CSV files for client

In order to run the back-end scripts, the runtime environment requires the following:

- Python 2.7 interpreter
- MongoDB server
- pymongo DB Python interface
- pandas Python library
- an internet connection
- 6GB free disk space

After the MongoDB server is started, execute the DataWrangling.py script. When it completes, execute the DataMunging.py script. The result will be a series of CSV files containing the required client data organized by month/year.

Front-End

The project “front-end” is comprised of the following HTML file:

- DataVisualization.html

The front-end file utilizes HTML, javascript, and the D3 library. The following files must be in the local directory where the file resides:

- the CSV data files
- us-states.json

The user may select any month and year combination available via the provided pull-down menu. The front-end code will retrieve the related CSV file, if one is available from the NOAA site, and use the data to display the estimate monthly water production data on the US map.

The circles on the US map represent reporting weather stations, and hovering on a circle will display relevant information about the estimated water production and the station location.