Michael Johnson

CS 440 Spring 2015

Project Report

**Intro**

This project utilized Amazon AWS Elastic Map Reduce service to process a database containing 20 GB of weather data from 1929 to 2009.

**List of Data Collected**

I collected a large amount of data but unfortunately wasn’t able to use it all due to time constraints, however, plotting all the data I collected would not be much of a challenge.

**YOU** can find the data I collected at the following link. Please let me know if you have any trouble accessing it.

s3.amazonaws.com/mjohns39-cs440/

**Data Analysis**

When analyzing the data, I used all the stations and separated them into Northern and Southern Hemispheres…that’s right, I said all the stations.

From there, I did an analysis over the course of the year and for the months for the data provided in the link.

Here is a list though to summarize:

1. North and South
   1. Station and Globally
      1. Annually and Monthly
         1. Max Gust
         2. Max Temperature\*
         3. Max Wind Speed
         4. Mean Wind Speed
         5. Min Temperature
         6. Average Temperature\*
         7. Total Precipitation

A \* next to the list item means I make a D3 plot for that data (eight total).

**Code**

I use Pig to analyze the weather data. My pig script can be found in the folder “Pig Script”.

**Charts**

For my charts, I used D3. Since one of the core strengths of D3 is making plots for web pages, I’ve included several html markups that can be opened to view the charts I’ve made…since copying and pasting the results would be a crime.

You can find the html documents for the various data analyzed (\* in list above) in the folder “Data Charts”.

**Conclusion**

From the results of the data analysis, Global Warming is indeed happening.

I had A LOT of fun with this project. Learning Pig, AWS EMR, JavaScript, HTML, SVG, and D3 has given me a profound since of accomplishment.