

CSE 6331 Cloud Computing Spring 2013, © DL, UTA, 2013

Programming Assignment 2 Introduction to Cloud Computation: Hadoop
Due: February 24, 2013, 23:00 (UTA time)

Description: One of the common uses of "Clouds", is large computation. A very popular and effective tool to aid in creating understanding of large data is Hadoop.

You are to take the data set "TexasWeatherFor50YearsBig" which is a file, in CSV format, of Weather reports from Texas weather stations, for somewhat more than the last 50 years. There is a description of: the format of the data, as well as, the location of the stations (in Texas).

Please compute yearly arithmetic means, maximums and minimums, for Temperature, Wind and one of the following: (rain, tornado, or any other reported data):

1. Each Year, for all of Texas.
2. Each year for each station.
3. For each station (all years)

Please use Hadoop, which you may obtain from <http://hadoop.apache.org/> (or similar), you may use your computer, or any computer available to you.

You should run (map) at least 20 jobs, or more, to be efficient. You may use, and refer to, online (or printed) references, programs and other tools, as long as ***you clearly site references.***

You may, if you wish, work in groups of two, or you may work alone.

Bonus: You may add "value-added" features (look at the service provider for ideas - such as scheduled backup, alerts, etc.)

Please, Email ONLY to the class account. All work must be your own (or team of two).

You must e-mail this lab, working (or partially) by the due date. The e-mail subject should clearly state the lab number. You may (optionally) demonstrate this lab, working (or partially) to the GTA before the due date.

Your program should be well commented and documented, make sure the first few lines of your program contain your name, this course number, and the lab number. Your comments should reflect your design and issues in your implementation. Your design and implementation should address error conditions.