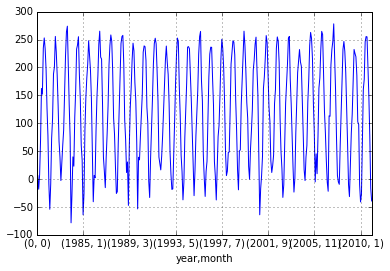
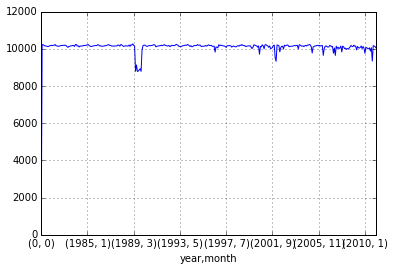
# Assignment 6





1. Enabled googles Apache Hadoop for cloud and installed Hadoop and created a cluster with minimal configurations.
2. Downloaded climate data for hourly details based on a station.
3. Python program to read the data and map the data values to a data frame using pandas has been handled.
4. Reducer calculates the temperature values based on the year and averages them.
5. Yearly and month wise temperatures and sea level pressure were calculated.
6. Data set contains around 200,000+ records.
7. Results are displayed using graphs . x-axis (Year and month) ,y-axis( Temperature ) with scaling factor as 10.
8. And for the other graph y-axis used was sea level pressure.

Commands:

hadoop jar /home/hadoop/hadoop-install/contrib/streaming/hadoop-streaming-1.2.1.jar \

-jobconf mapred.reduce.tasks=2 \

-file ./mapper.py \

-mapper ./mapper.py \

-file ./reducer.py \

-reducer ./reducer.py \

-output myoutput

References:

<https://aws.amazon.com/elasticmapreduce/>

<http://cs.smith.edu/dftwiki/index.php/Hadoop_Tutorial_2_--_Running_WordCount_in_Python>

<http://hadoop.apache.org/docs/r1.2.1/streaming.html#Hadoop+Streaming>

<https://cloud.google.com/hadoop/getting-started>