HW 3 Hadoop Mapreduce

- (a) (90%) Write a map reduce application using Python to do the log analysis of apache2 web server access log

- (b) (10%) Write the Java version of part (a)

  - The output of the application: visits per hour

```
2012-11-18 T 16:00:00.000        1
2012-11-18 T 17:00:00.000       21
2012-11-18 T 18:00:00.000        3
2012-11-18 T 19:00:00.000        4
2012-11-18 T 20:00:00.000        5
2012-11-18 T 21:00:00.000       21
...
2012-12-17 T 14:00:00.000       30
```

  - Use the access_log sample (http://hpc.ee.ntu.edu.tw/html/IntelligentClouds/webAccessLog/access_log) or the one chosen by you

  - Run the application in a Hadoop Docker container (can be in Google cloud, AWS EC2, Azure cloud).

  - Submission: 1) a report with discussion, 2) source codes in github, 3) run commands

  - Ref:

    Applied Big Data Analysis in the Real World with MapReduce and Hadoop: http://www.informit.com/articles/article.aspx?p=2017061