COEN498/691 (Fall 2017) - Assignment 2
Due by December 2nd 23:55
Individual Assignment

Description

In this assignment, write a MapReduce program to process the following data set.
The Canadian Radiological Monitoring Network – Airborne Radioactivity.
http://open.canada.ca/data/en/dataset/21b821cf-0f1c-40ee-8925-eab12d357668
The description and sample plotting of the data are available on this link.

The program should display the **Maximum, Minimum, Median** values and **Standard Deviation** of the metric **7Be MDC/7Be CMD (mBq/m3)** for a location in each year.

You can choose to code the MapReduce program in Java, C++ or Python. No matter what language you choose to implement, the submission should be packed and executable.

- For Java, all the classes should be packed in a single executable Jar.
- For C++, please provide the make file

This MapReduce program is programmed and executed in a Mapreduce Runtime such as Apache Hadoop or AWS EMR. It CANNOT be a standalone program that mimic the behavior of mapping and reducing.

Pack all your executable in a single archive file.

You can choose the execution environment, either it is on local computer, or in the lab computer, or on the Cloud.

Submit to Moodle site the following :
1. your code in a single package, including both source code and packaged executable

2. A report in PDF that describes
a. Your map reduce algorithm design
b. How to run your program and show results with your sample data with screenshots.
c. Use the logs to estimate the time taken for the map task and reduce task respectively.
d. Discuss if your observation from the logs of data partition and workload balancing.

Marking Criteria
1) Correctness of the program code (40%)

2) MapReduce algorithm design (20%)
3) Quality of the report with clear description (40%)