CSCD467/567 Homework 6
Cloud Computing, Distributed File System and MapReduce

What is provided
In the start-up folder, you are provided two text files **file1** and **file2** that will be the input files to your program.

Problem description and Requirements
You are required to write a MapReduce program that takes file1 and file2 as input. The **number** at the beginning of each line in file1 and file2 indicates the line number within that file. You can use that number in your program.

Within all input files, the program searches a pattern **MyKey** you specified through command-line argument, and output**s** a full path to the file in which the pattern **MyKey** is found, together with a line number within that file at which the pattern appears.

Two test cases are provided, file **output_luck** and **output_world,** the search results for the pattern **luck** and **world**. These files demonstrate the sample output that your program is supposed to generate. Your output should be similar to the sample output.

Turn In

Turn in all your source files, you input files and a readme.txt file that describes the commands used in shell to compile and run your program.

Turn in screen shots that shows your program can be compiled successfully and runs successfully.
Turn in screen shots that shows the content of the output file(s) generated by your program for the pattern **luck** and **world** in two separate runs.

Please zip all your files into a .zip file and name it as firstNameInitial + Lastname + hw6.zip, submit the zip file on canvas.

Useful Materials
1, about how to obtain the input file name in each mapper.
http://stackoverflow.com/questions/19012482/how-to-get-the-input-file-name-in-the-mapper-in-a-hadoop-program

2, An e-book about Hadoop is also included in the startup folder, HadoopTheDefinitiveGuide.pdf

3, MapReduce API 2.6
https://hadoop.apache.org/docs/r2.6.0/api/org/apache/hadoop/mapreduce/package-summary.html