Counting words in Alice.txt

1. Create cluster on google cloud

2. Access the master node (name node) via SSH

3. upload the wordcount.java & Alice.txt to the master node

4. Set up environment

    1. create user directory : hadoop fs -mkdir /user/pakin

    2. make java accessible via terminal : export PATH=${JAVA_HOME}/bin:${PATH}

    3. tools for compiling .java to .jar : export
       HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar

5. On vm local machine compline wordcount.java

    1. hadoop com.sun.tools.javac.Main ./wordcount.java

    2. jar cfwordcount.jar WordCount*.class

6. (for me to be able to run I need to upload the Alice to hdfs first) : hadoop fs -put Alice.txt
   /user/pakin/

7. run the wordcount.jar on hadoop : hadoop jar wordcount.jar WordCount
   /user/pakin/Alice.txt /user/pakin/out

8. get the file from hdfs to vm local machine

9. hadoop fs -get /user/pakin/out/

10. downloads back to my laptop