Instructions for Execution

1) create project directory : (**wordcount**)
    a) Browse to home directory
        cd /home/iitp
    b) Create project directory
        mkdir **wordcount**

2) Create source file: **(WordCount.java)**
    a) Browse into wordcount directory
        cd /home/iitp/wordcount
    b) nano WordCount.java
        paste the lines from the source code provided

3) Create input directory (**inputdata)** for input files
    cd /home/iitp/wordcount
    mkdir **inputdata**

4) **Copy the input file (input.txt) into inputdata folder**
    cd /home/iitp/wordcount/inputdata
    nano input.txt
        paste the lines from the provided input file

5) **Start all hadoop services**
    a) Browse to hadoop installation sbin directory
        cd /home/iitp/hadoop-2.6.0/sbin
    b) start all services
        ./start-all.sh
    Enter password as required

6) **Create input directory on HDFS**
    a) browse to hadoop installation bin folder
        cd /home/iitp/hadoop-2.6.0/bin
    b) create directory
        ./hadoop fs -mkdir /wordcount

c) create sub-directory inside wordocunt on HDFS

./hadoop fs -mkdir /wordcount/inputdata

7) **Copy the input text file from local directory to HDFS**

a) browse to hadoop installation bin folder

cd /home/iitp/hadoop-2.6.0/bin

b) Copy from Local

./hadoop dfs -put /home/iitp/wordcount/inputdata/input.txt /wordcount/inputdata/

8) **Compile the Java Source File**

a) export the Hadoop classpath

type in terminal:

export HADOOP_CLASSPATH=/usr/lib/jvm/java-1.8.0-openjdk-amd64/lib/tools.jar

b) browse to bin folder of hadoop installation

cd /home/iitp/hadoop-2.6.0/bin

c) compile

./hadoop com.sun.tools.javac.Main /home/iitp/wordcount/WordCount.java

9) **Create Jar file**

a) Browse to wordcount directory on your VM

cd /home/iitp/wordcount

jar cf wc.jar WordCount*.class

10) **Running the WordCount program**

a) browse to the bin directory of hadoop installation

cd /home/iitp/hadoop-2.6.0/bin

b) Running in terminal

./hadoop jar /home/iitp/wordcount/wc.jar WordCount /wordcount/inputdata/ /wordcount/outputdata

b) Finding outputs

./hadoop fs -cat /wordcount/outputdata/part-r-00000