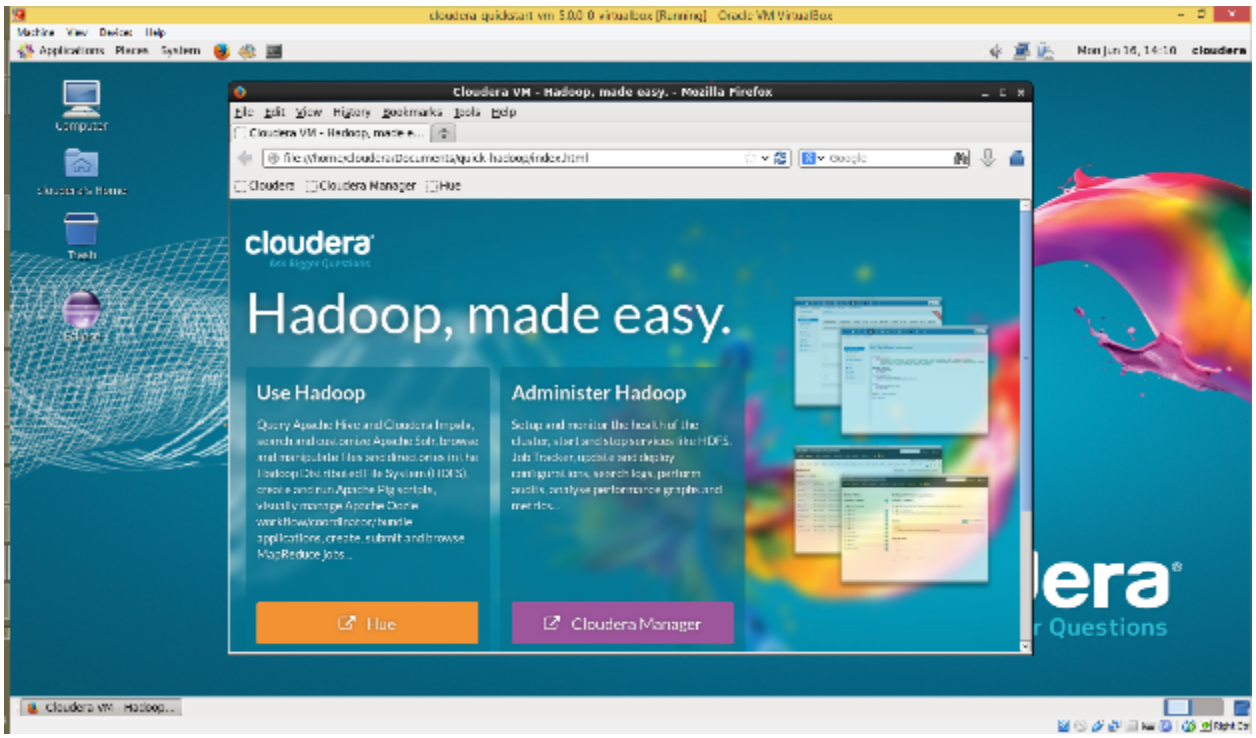
Lab 1 Part 2 Cloudera

Installing Cloudera on LocalMachine

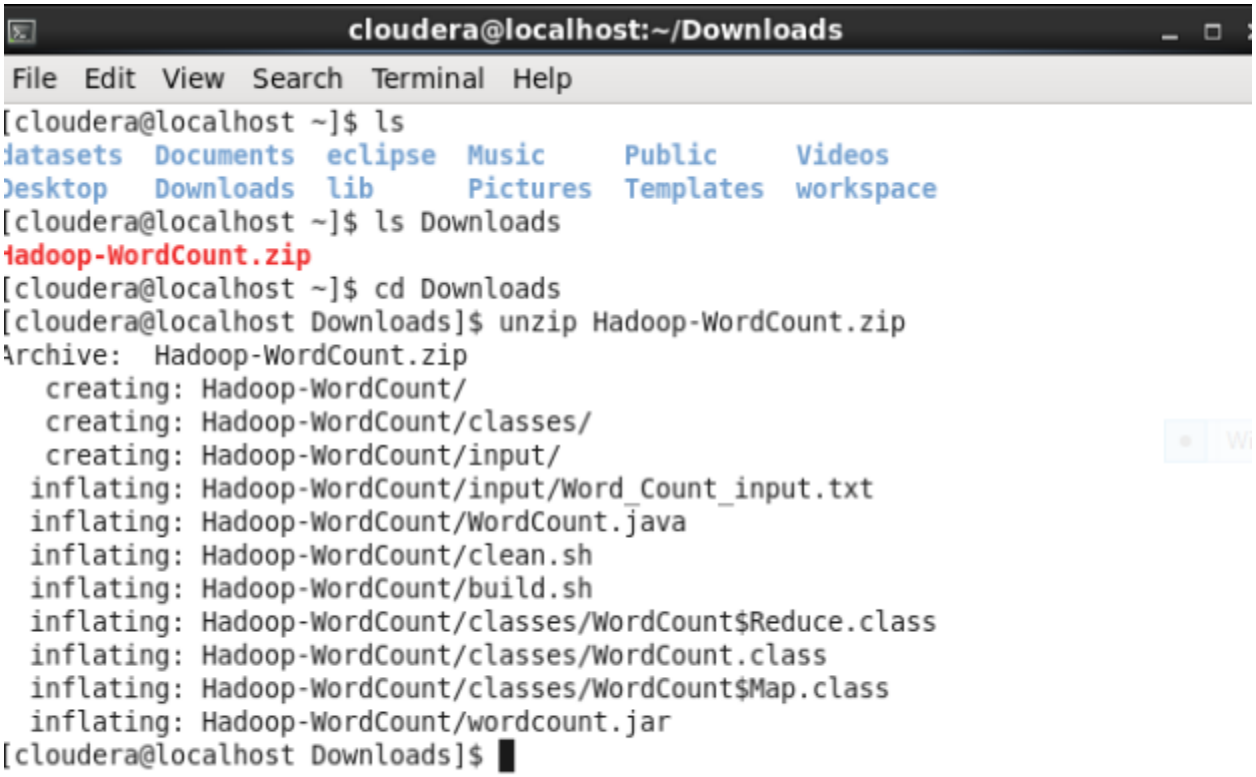
A Cloudera 5.0 VM image is download from
http://www.cloudera.com/content/support/en/downloads/quickstart_vms/cdh-5-0-x.html

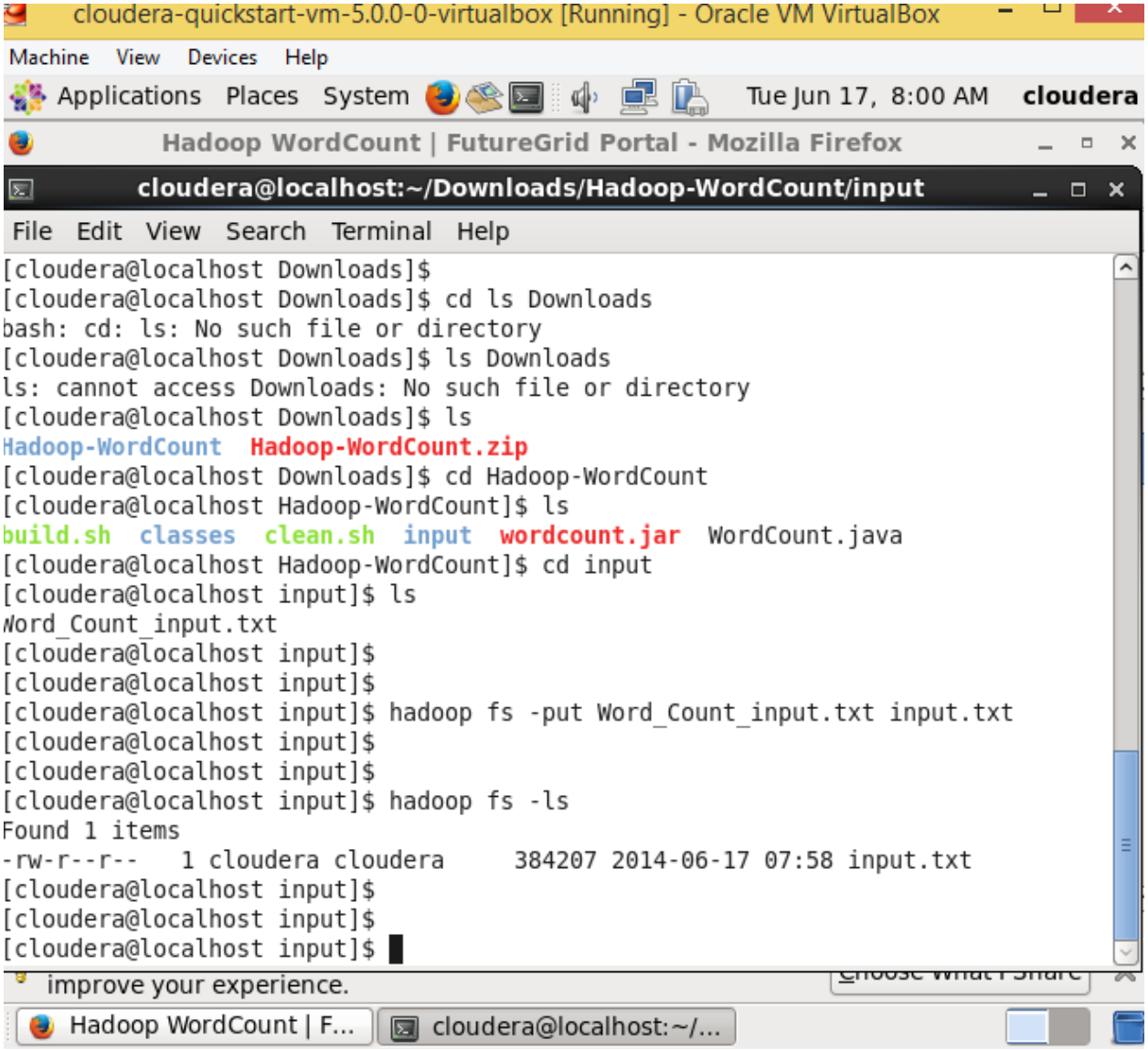
Now the VM is booted up from **VirtualBox** and is ready to perform Hadoop jobs.



Transferring files to Hadoop

Every Command on Hadoop can be performed by
`hadoop -fs`
command, and is same as Unix (Linux) Command


cloudera@localhost:~/Downloads
File Edit View Search Terminal Help
[cloudera@localhost ~]$ ls
datasets Documents eclipse Music Public Videos
Desktop Downloads lib Pictures Templates workspace
[cloudera@localhost ~]$ ls Downloads
Hadoop-WordCount.zip
[cloudera@localhost ~]$ cd Downloads
[cloudera@localhost Downloads]$ unzip Hadoop-WordCount.zip
Archive: Hadoop-WordCount.zip
creating: Hadoop-WordCount/
creating: Hadoop-WordCount/classes/
creating: Hadoop-WordCount/input/
inflating: Hadoop-WordCount/input/Word_Count_input.txt
inflating: Hadoop-WordCount/WordCount.java
inflating: Hadoop-WordCount/clean.sh
inflating: Hadoop-WordCount/build.sh
inflating: Hadoop-WordCount/classes/WordCount$Reduce.class
inflating: Hadoop-WordCount/classes/WordCount.class
inflating: Hadoop-WordCount/classes/WordCount$Map.class
inflating: Hadoop-WordCount/wordcount.jar
[cloudera@localhost Downloads]$



For running WordCount Example, Two files from **Local FileSystem** to **HDFS** are transferred

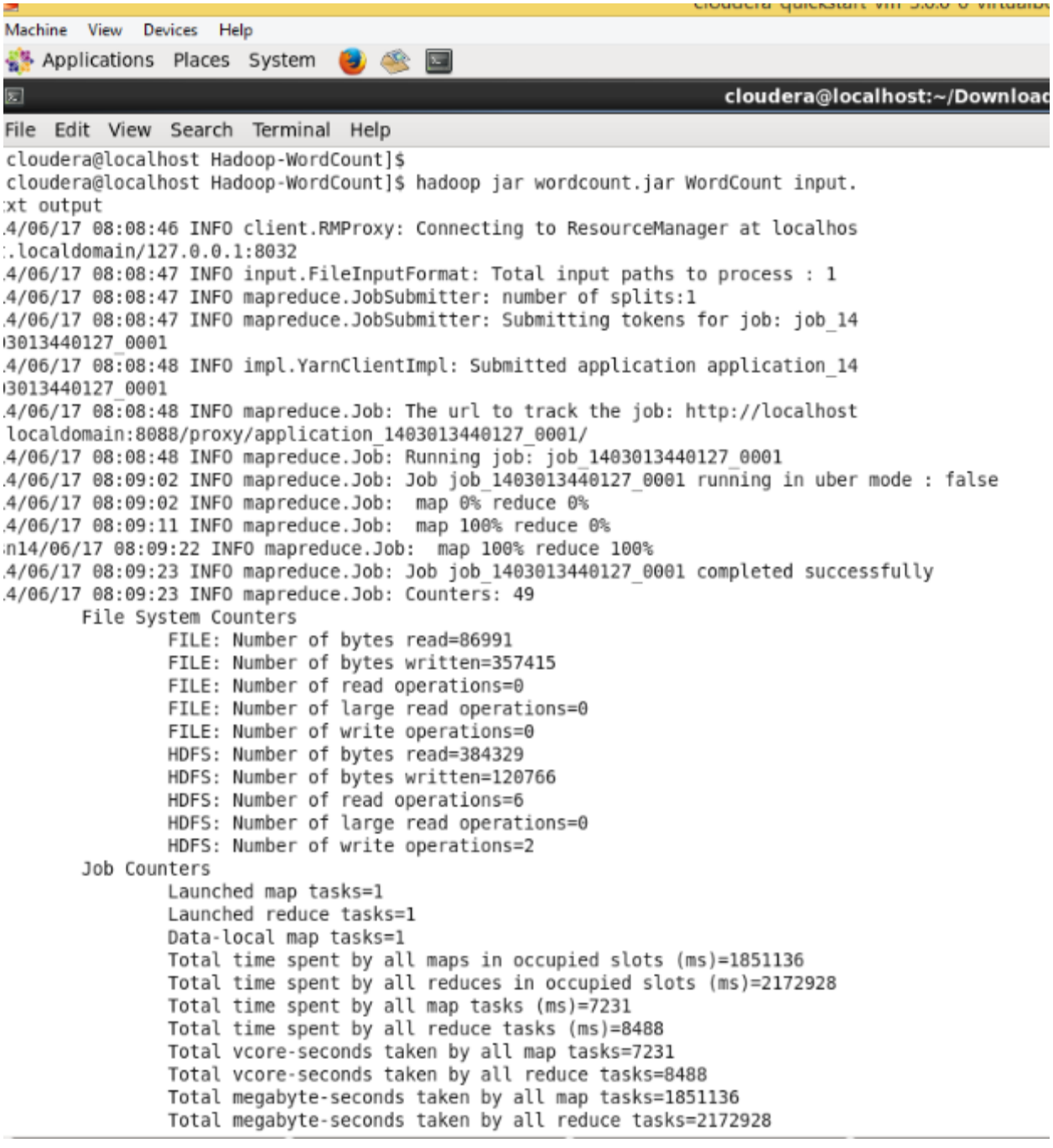
```
hadoop fs -put Word_Count_input.txt input.txt
```

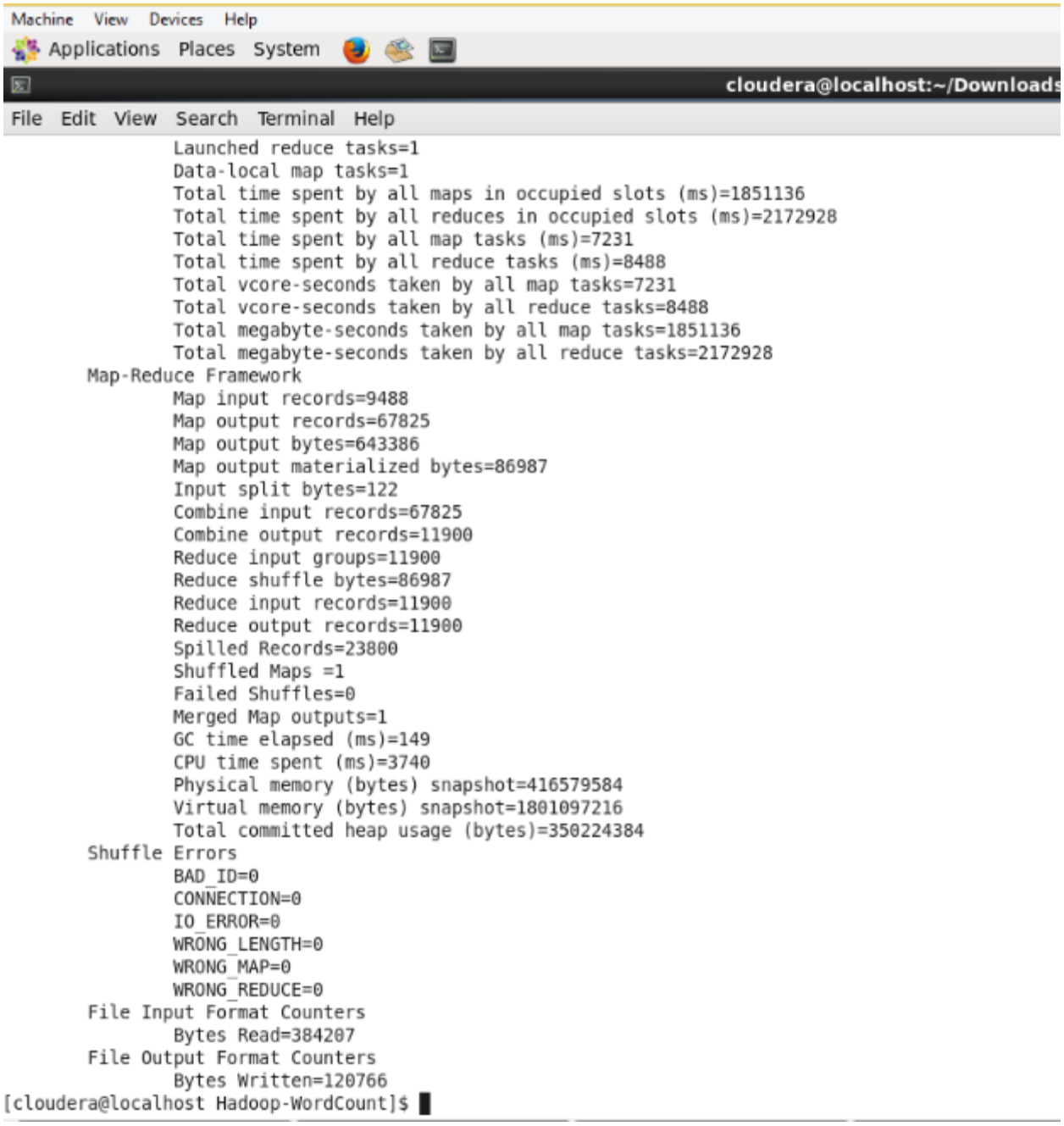
```
hadoop fs -put wordcount.jar worcount.jar
```

Running WordCount Program on Cloudera

Now as both the files required are moved to HDFS, we're ready to run the program

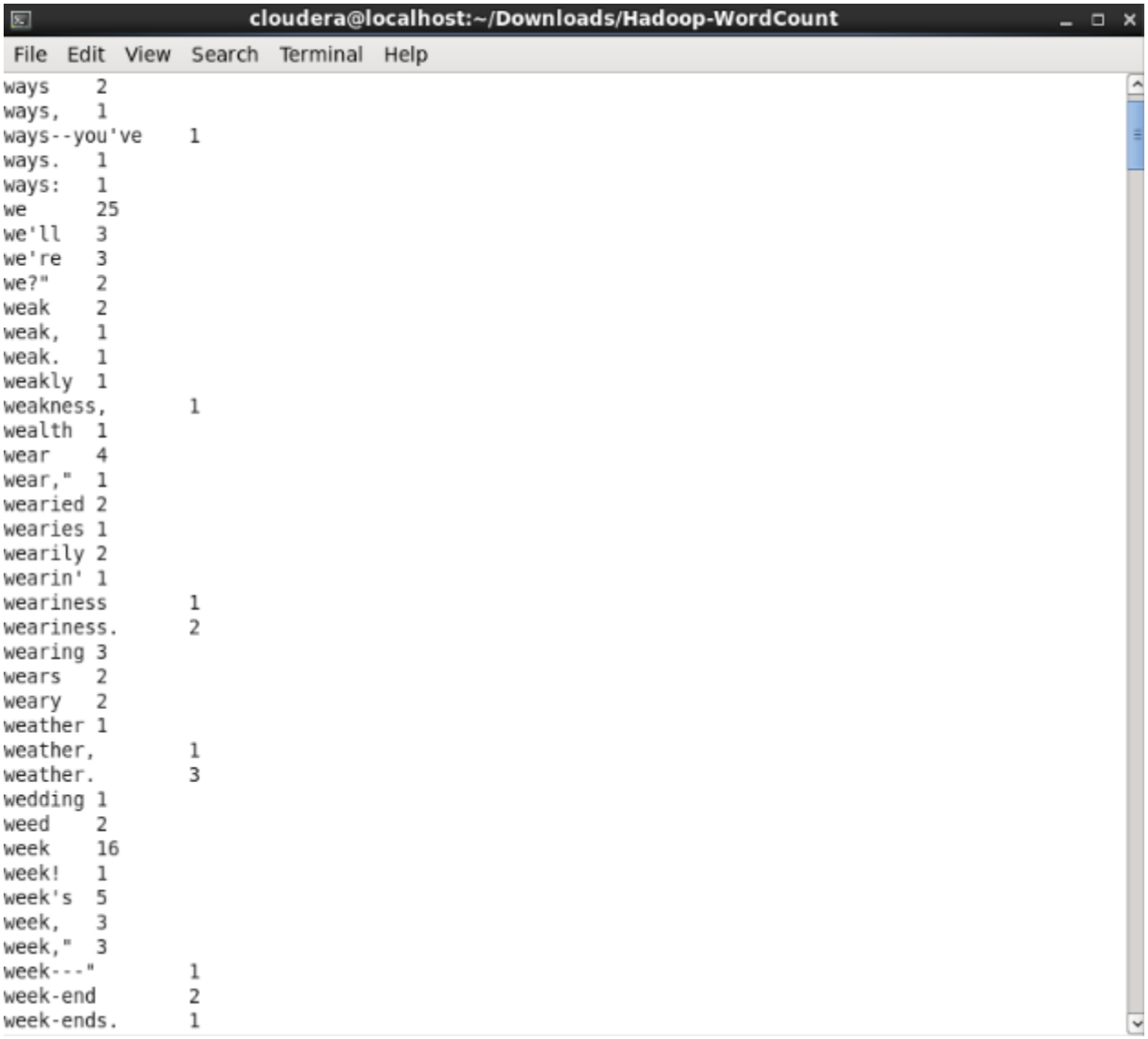
```
hadoop jar worcount.jar WordCount input.txt output
```





The word count is run on the input.txt and the Output is generated on output folder

```
hadoop fs -cat output/*
```



The Terminal Shows the output, occurrence count of every word form the input.txt