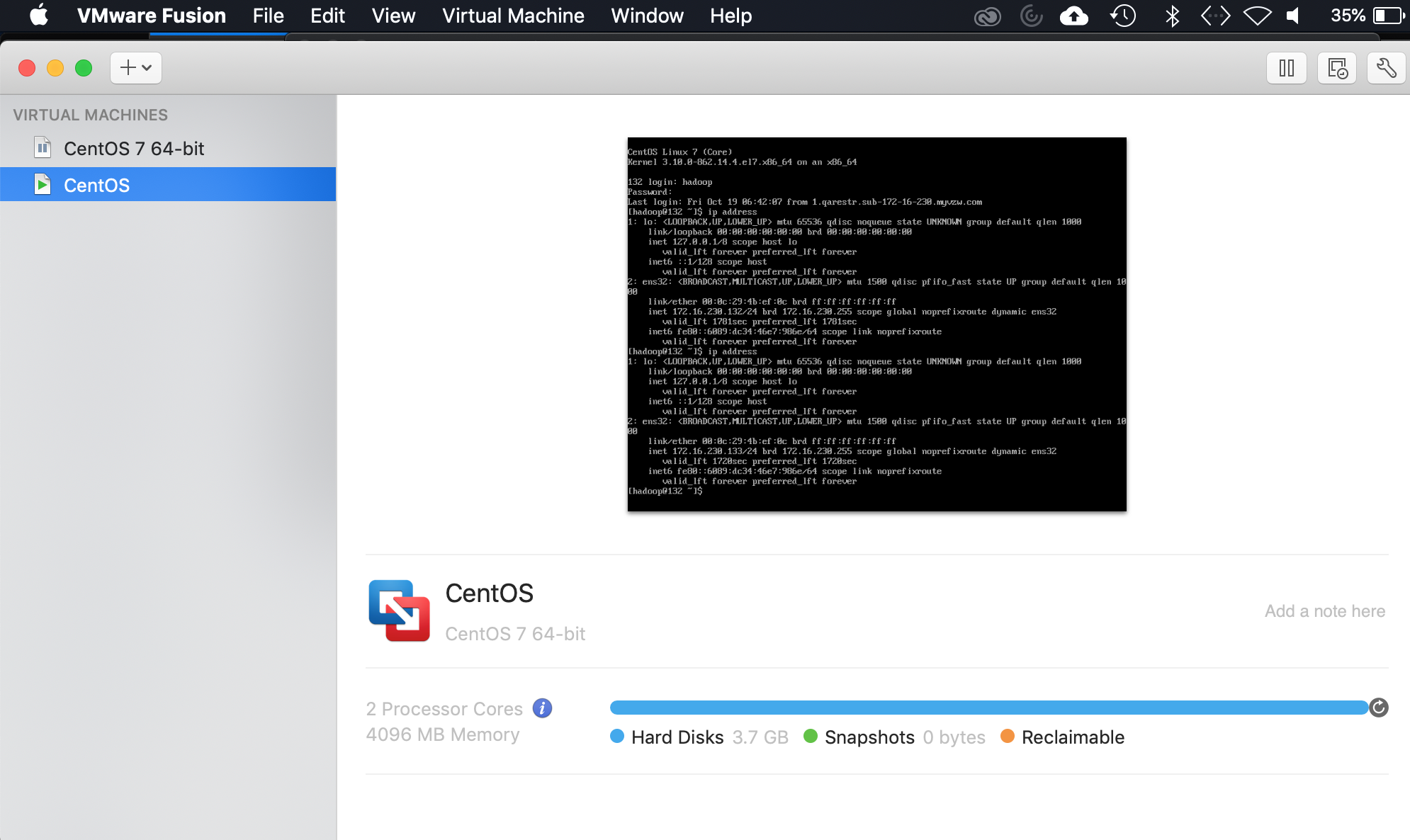
Exercies – Using Hadoop

1. Successfully downloaded VMware Fusion for Mac.
2. Installed a virtual machine with Linux operating system using .ovf file from DropBox.



1. Acquired the machine’s ip address and opened the virtual machine using Terminal and completed project as shown below. Code lines are underlined for clarity. Comments are written in red.

##connect to the virtual machine

users-MacBook-Pro-2163:~ etanan$ ssh hadoop@172.16.230.128

The authenticity of host '172.16.230.128 (172.16.230.128)' can't be established.

ECDSA key fingerprint is SHA256:KIGt2OCXyFCGjYGMVpoTy22KbVz47rER+Ps5eMjKDM4.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added '172.16.230.128' (ECDSA) to the list of known hosts.

hadoop@172.16.230.128's password:

Last login: Fri Oct 12 10:39:25 2018

##Start the Hadoop Distributed File System (DFS), Yet Another Resource Negotiator (YARN) and MapReduce from the core Hadoop components.

[hadoop@132 ~]$ start-dfs.sh

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [132.qarestr.sub-172-16-230.myvzw.com]

[hadoop@128 ~]$ start-yarn.sh

Starting resourcemanager

Starting nodemanagers

[hadoop@132 ~]$ mapred --daemon start historyserver

##Confirm the daemons are running

[hadoop@132 ~]$ jps

2417 SecondaryNameNode

2210 DataNode

2099 NameNode

1605 NodeManager

1497 ResourceManager

2585 Jps

1371 JobHistoryServer

##This line download the text document into the working folder. Shakespeare.txt was already present in the folder. Therefore, this command had renamed the file as ‘shakespeare.txt.1’ and saved it.I noticed this line requires “www” as part of the URL or it will return an error.

[hadoop@128 ~]$ wget http://www.norvig.com/ngrams/shakespeare.txt

--2018-10-12 22:44:40-- http://www.norvig.com/ngrams/shakespeare.txt

Resolving www.norvig.com (www.norvig.com)... 66.96.146.129

Connecting to www.norvig.com (www.norvig.com)|66.96.146.129|:80... connected.

HTTP request sent, awaiting response... 200 OK

Length: 4538523 (4.3M) [text/plain]

Saving to: ‘shakespeare.txt.1’

100%[======================================>] 4,538,523 1.15MB/s in 4.3s

2018-10-12 22:44:45 (1.01 MB/s) - ‘shakespeare.txt.1’ saved [4538523/4538523]

##The next three lines of codes create directories

[hadoop@128 ~]hadoop fs -mkdir -p /user/hadoop

[hadoop@128 ~]hadoop fs -mkdir shakespeare

[hadoop@128 ~]hadoop fs -mkdir shakespeare/input

##Here, this line copies Shakespeare.txt from the local file location into the Hadoop file system inside the Shakespeare/input directory.

[hadoop@128 ~]hadoop fs -copyFromLocal shakespeare.txt shakespeare/input

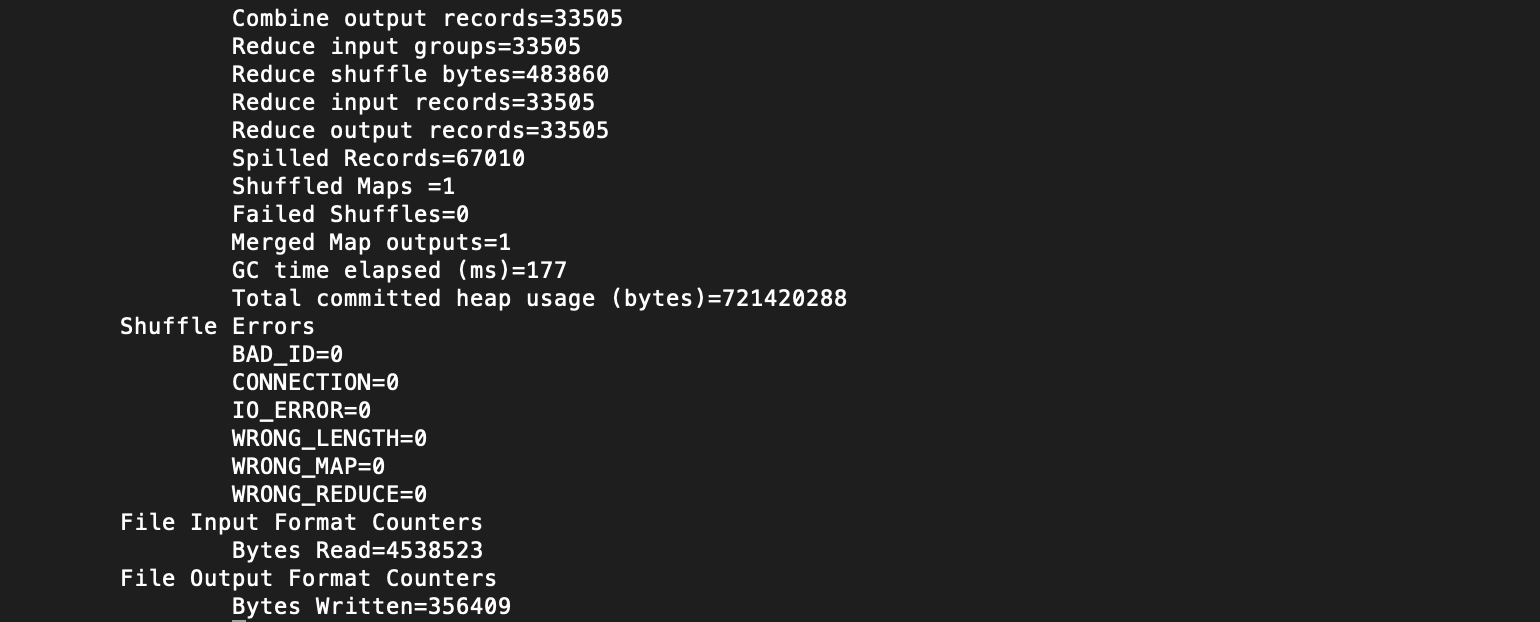
[hadoop@128 ~]hadoop fs -ls shakespeare/input

Found 1 items

-rw-r--r-- 1 hadoop supergroup 4538523 2018-10-24 14:53 shakespeare/input/shakespeare.txt

##The line of code here calls from a Java Archive file in the local file /home/hadoop/hadoop-3.1.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.1.jar and runs a wordcount where the standard input file is read from Shakespeare/input and the output will be stored in Shakespeare/output

[hadoop@128 ~]hadoop jar /home/hadoop/hadoop-3.1.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.1.jar wordcount shakespeare/input shakespeare/output



##It shows here that the two files in the folder are ready for further processing.

[hadoop@128 ~]hadoop fs -ls shakespeare/output

Found 2 items

-rw-r--r-- 1 hadoop supergroup 0 2018-10-12 22:58 shakespeare/output/\_SUCCESS

-rw-r--r-- 1 hadoop supergroup 356409 2018-10-12 22:58 shakespeare/output/part-r-00000

##The output file is then cancatinated and read using –cat command from below.

[hadoop@128 ~]hadoop fs -cat shakespeare/output/part-r-00000



##Output file part-r-00000 in shakespeare/output is now copied from the Hadoop file system as a text file into the local directory as shakespeareoutput.txt

[hadoop@128 ~]hadoop fs -copyToLocal shakespeare/output/part-r-00000 ~/shakespeareoutput.txt

##The tile can be seen here in the local file which is highlighted.

[hadoop@133 ~]$ ls -la

total 833152

drwx------. 10 hadoop hadoop 4096 Oct 12 23:07 .

drwxr-xr-x. 3 root root 20 Oct 8 16:30 ..

drwxrwxr-x. 10 hadoop hadoop 184 Oct 8 16:39 apache-hive-3.1.0-bin

-rw-rw-r--. 1 hadoop hadoop 281005636 Jul 30 12:39 apache-hive-3.1.0-bin.tar.gz

drwxrwxr-x. 11 hadoop hadoop 4096 Oct 8 17:17 apache-mahout-distribution-0.13.0

-rw-rw-r--. 1 hadoop hadoop 228079891 Apr 17 2017 apache-mahout-distribution-0.13.0.tar.gz

-rw-------. 1 hadoop hadoop 3031 Oct 19 16:43 .bash\_history

-rw-r--r--. 1 hadoop hadoop 18 Apr 10 2018 .bash\_logout

-rw-r--r--. 1 hadoop hadoop 193 Apr 10 2018 .bash\_profile

-rw-r--r--. 1 hadoop hadoop 799 Oct 8 20:11 .bashrc

drwxrwxr-x. 2 hadoop hadoop 21 Oct 10 12:45 .beeline

-rw-rw-r--. 1 hadoop hadoop 22196 Oct 8 20:06 derby.log

drwxr-xr-x. 4 hadoop hadoop 34 Oct 10 12:44 Exercises

lrwxrwxrwx. 1 hadoop hadoop 12 Oct 8 16:39 hadoop -> hadoop-3.1.1

drwxr-xr-x. 10 hadoop hadoop 161 Oct 8 19:39 hadoop-3.1.1

-rw-rw-r--. 1 hadoop hadoop 334559382 Aug 8 13:43 hadoop-3.1.1.tar.gz

drwxrwxr-x. 3 hadoop hadoop 18 Oct 8 19:39 hadoopdata

lrwxrwxrwx. 1 hadoop hadoop 21 Oct 8 16:40 hive -> apache-hive-3.1.0-bin

-rw-rw-r--. 1 hadoop hadoop 22 Oct 8 20:06 .hivehistory

lrwxrwxrwx. 1 hadoop hadoop 33 Oct 8 17:18 mahout -> apache-mahout-distribution-0.13.0

drwxrwxr-x. 5 hadoop hadoop 133 Oct 8 20:06 metastore\_db

-rw-r--r--. 1 hadoop hadoop 356409 Oct 12 23:07 shakespeareoutput.txt

-rw-rw-r--. 1 hadoop hadoop 4538523 Aug 16 2011 shakespeare.txt

-rw-rw-r--. 1 hadoop hadoop 4538523 Aug 16 2011 shakespeare.txt.1

drwx------. 2 hadoop hadoop 80 Oct 8 19:39 .ssh

[hadoop@128 ~]more shakespeareoutput.txt

