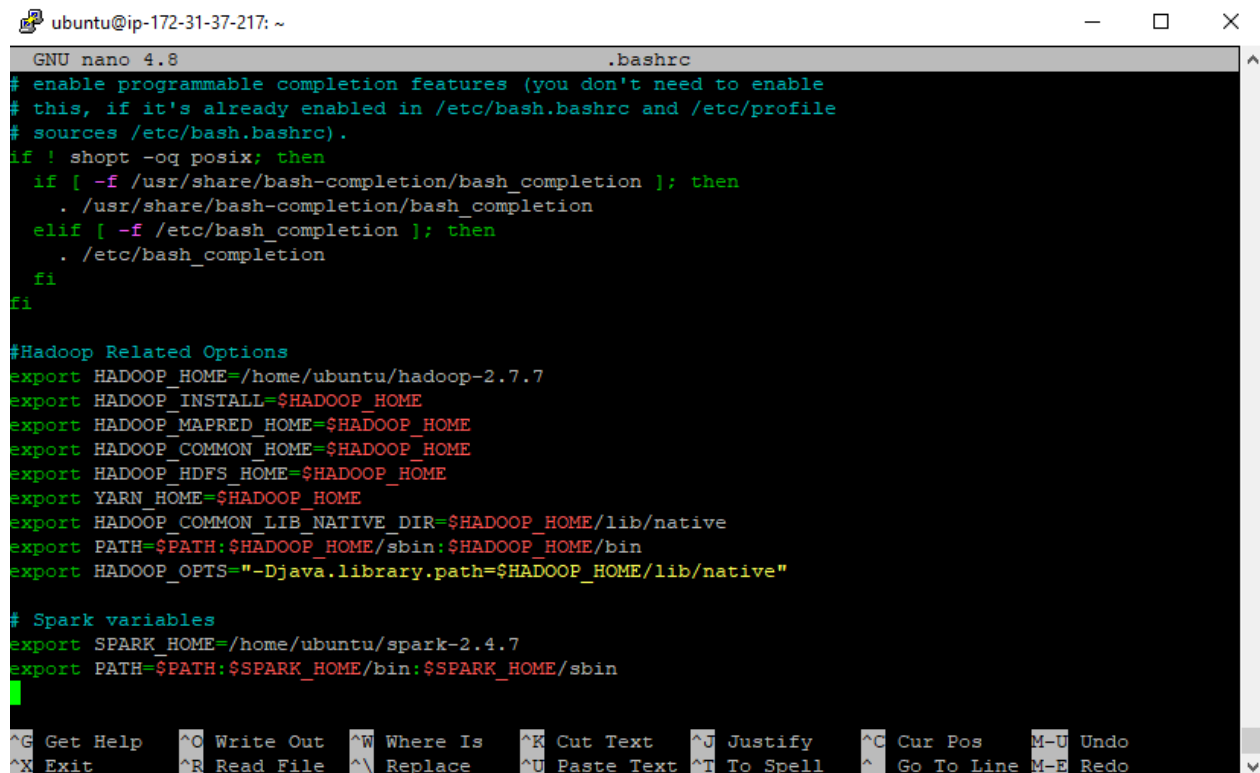


**Question 4. Install Spark (v2) as a pseudocluster mode and Show Word counting example with your own document file (same as Hadoop).**

After setting Hadoop up, started to work on Spark 2.4.7.

<https://ftp.itu.edu.tr/Mirror/Apache/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz>untared it and get the environment variables right.



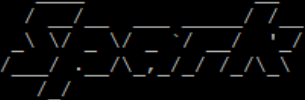
```
ubuntu@ip-172-31-37-217: ~  
GNU nano 4.8 .bashrc  
# enable programmable completion features (you don't need to enable  
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
#Hadoop Related Options  
export HADOOP_HOME=/home/ubuntu/hadoop-2.7.7  
export HADOOP_INSTALL=$HADOOP_HOME  
export HADOOP_MAPRED_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin  
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"  
  
# Spark variables  
export SPARK_HOME=/home/ubuntu/spark-2.4.7  
export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin  
  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos M-U Undo  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

Stopped and started all the nodes again since we made some changes on the paths.

This time wanted to use a different book, Metamorphosis from Franz Kafka.

With the command;

wget <https://www.gutenberg.org/ebooks/5200.txt.utf-8>

```
ubuntu@ip-172-31-37-217: ~  
ubuntu@ip-172-31-37-217:~$ ls  
hadoop-2.7.7  spark-2.4.7  
ubuntu@ip-172-31-37-217:~$ sudo nano .bashrc  
ubuntu@ip-172-31-37-217:~$ exec .bashrc  
-bash: exec: .bashrc: not found  
ubuntu@ip-172-31-37-217:~$ source ~/.bashrc  
ubuntu@ip-172-31-37-217:~$ spark-shell  
21/02/06 23:38:17 WARN NativeCodeLoader: Unable to load native-hadoop library for  
your platform... using builtin-java classes where applicable  
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
Spark context Web UI available at http://ip-172-31-37-217.eu-central-1.compute.internal:4040  
Spark context available as 'sc' (master = local[*], app id = local-1612654704226).  
Spark session available as 'spark'.  
Welcome to  
 version 2.4.7  
Using Scala version 2.11.12 (OpenJDK 64-Bit Server VM, Java 1.8.0_275)  
Type in expressions to have them evaluated.  
Type :help for more information.  
scala>
```

I wanted to use scala for this example since i dont have any prior experience on it.

Executed the bash again and entered spark-shell;

Spark-shell

1. `val input=sc.textFile("Metamorphosis.txt")`
2. `val wordcounts = input.flatMap(line => line.split(" ")).map(word=>(word,1)).reduceByKey(_ + _);`

