**Installation Guide**

**Creating a Cloudera Setup:**

1. Initially we need to download a VMPlayer from vmware website or any trustworthy website and make sure it is a non-commercial purpose version
2. After installing the vmware player download the cloud era image from the following link
   1. <http://www.cloudera.com/content/support/en/downloads/download-components/download-products.html?productID=F6mO278Rvo>
3. Then from vmplayer settings open the image and set its memory usage i.e. RAM utilization to 2.5 GB so that your system doesn’t become slow.
4. Once you load the image you can see the cloudera web pages displayed with several tools as bookmarks in Mozilla Firefox
5. Login to cloudera manger with credentials as cloudera and cloudera respectively for username and password.

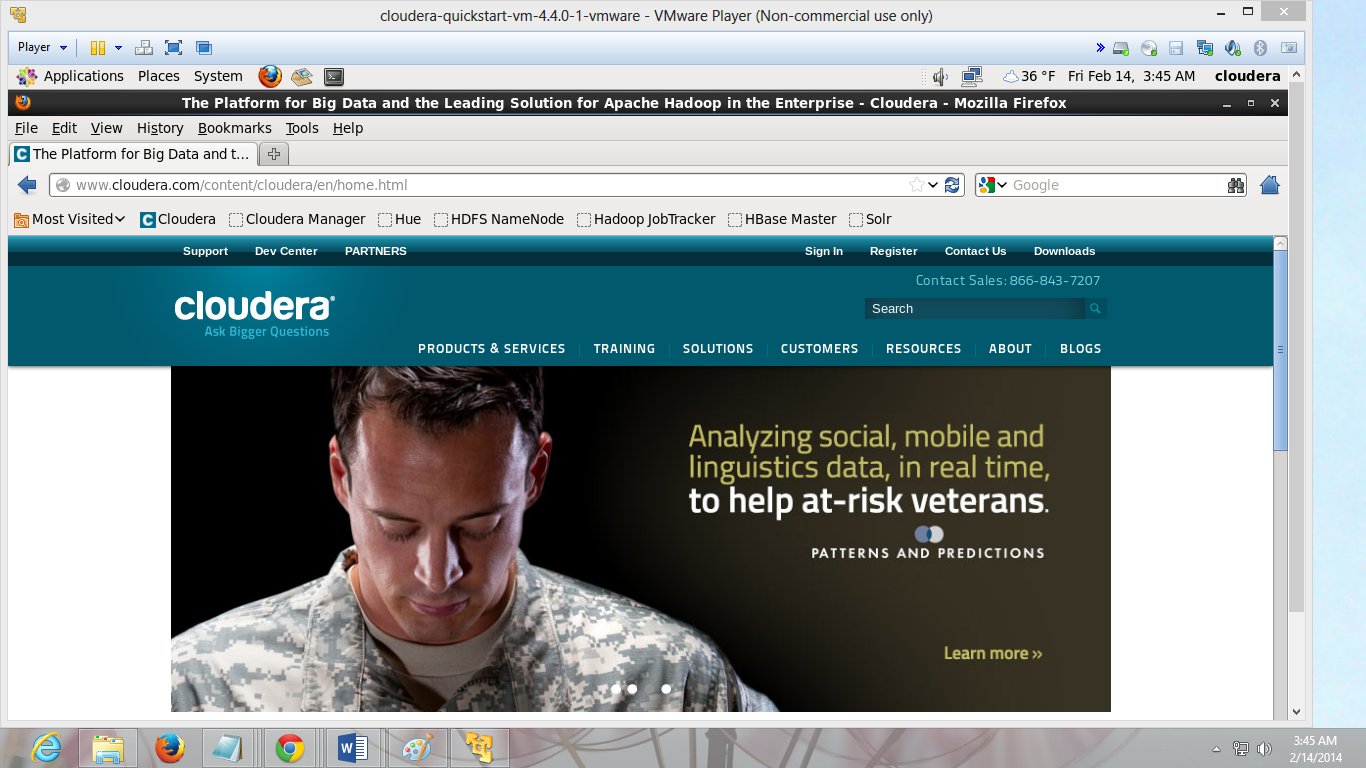


Figure Cloudera Image

**Troubleshooting:**

* If there is an issue while loading the image such as Intel VX enabled.
* Please restart your machine and press F2 which will guide you to BIOS settings.
* Inside BIOS settings make Virtual Technology enabled.

**Running Hadoop WordCount:**

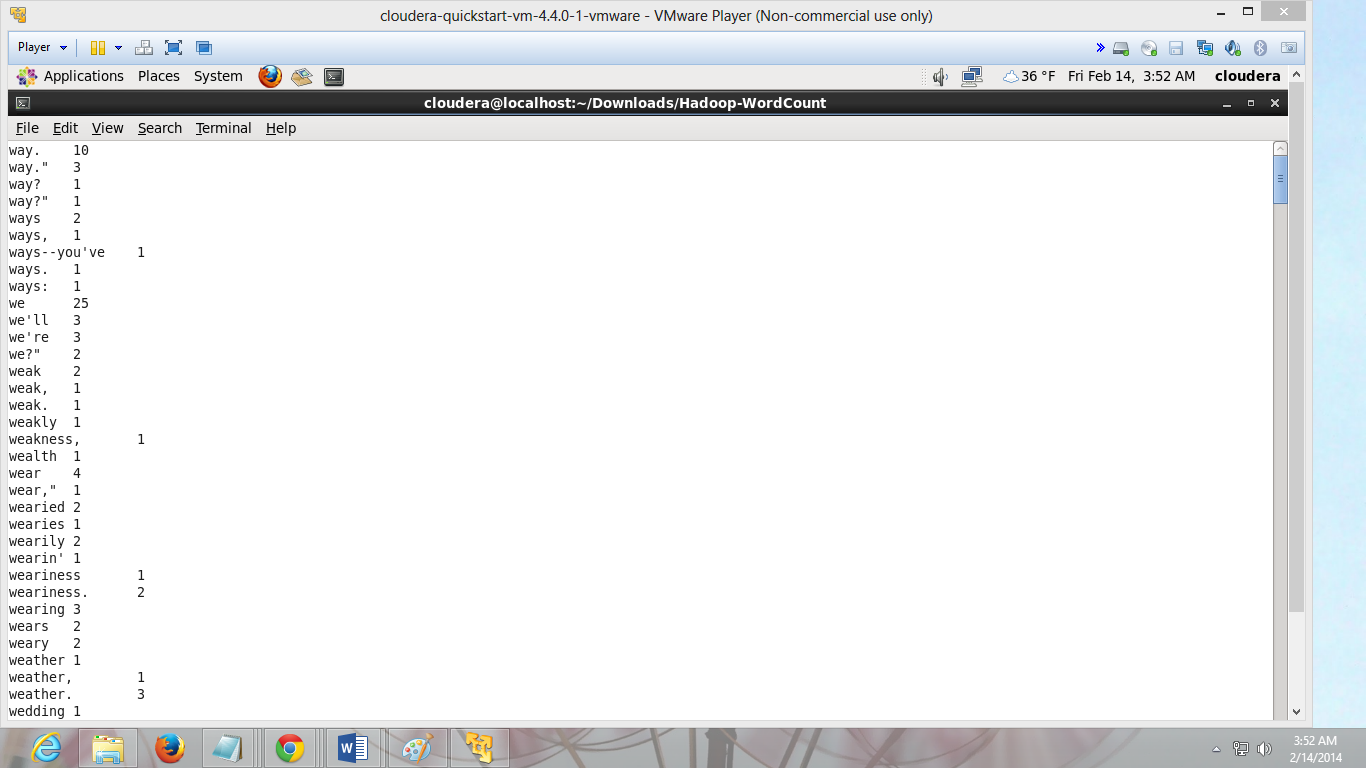
* Download the wordcount file from the following link:
  1. <https://portal.futuregrid.org/manual/hadoop-wordcount>
* Extract the wordcount file and run the following commands inside the word count folder
  1. Hadoop fs –put input input
  2. hadoop jar wordcount.jar WordCount input output
  3. Hadoop fs –cat output/\*
* 

Figure Word Count

* You can see the count of each word after running the commands.

**Installation of Maven:**

* Open a terminal in the cloudera image.
* And execute the following command to download maven
  1. wget <http://www.apache.org/dist//maven/binaries/apache-maven-3.0.5-bin.tar.gz>
* Extract the downloaded file using the following command
  1. tar –zxvf apache-maven-3.0.5-bin.tar.gz
* Export the environmental variables M2\_HOME, PATH
  1. Export M2\_HOME=/(downloaded location)/apache-maven 3.0.5
  2. Export PATH=${M2\_HOME}/bin:${PATH}
* You can store these variable sin bash profile so that there is no need to export these variables each and every time when you log in.
* Run the command “mvn”
* And to check for successful installation execute the following command
  1. Mvn –version

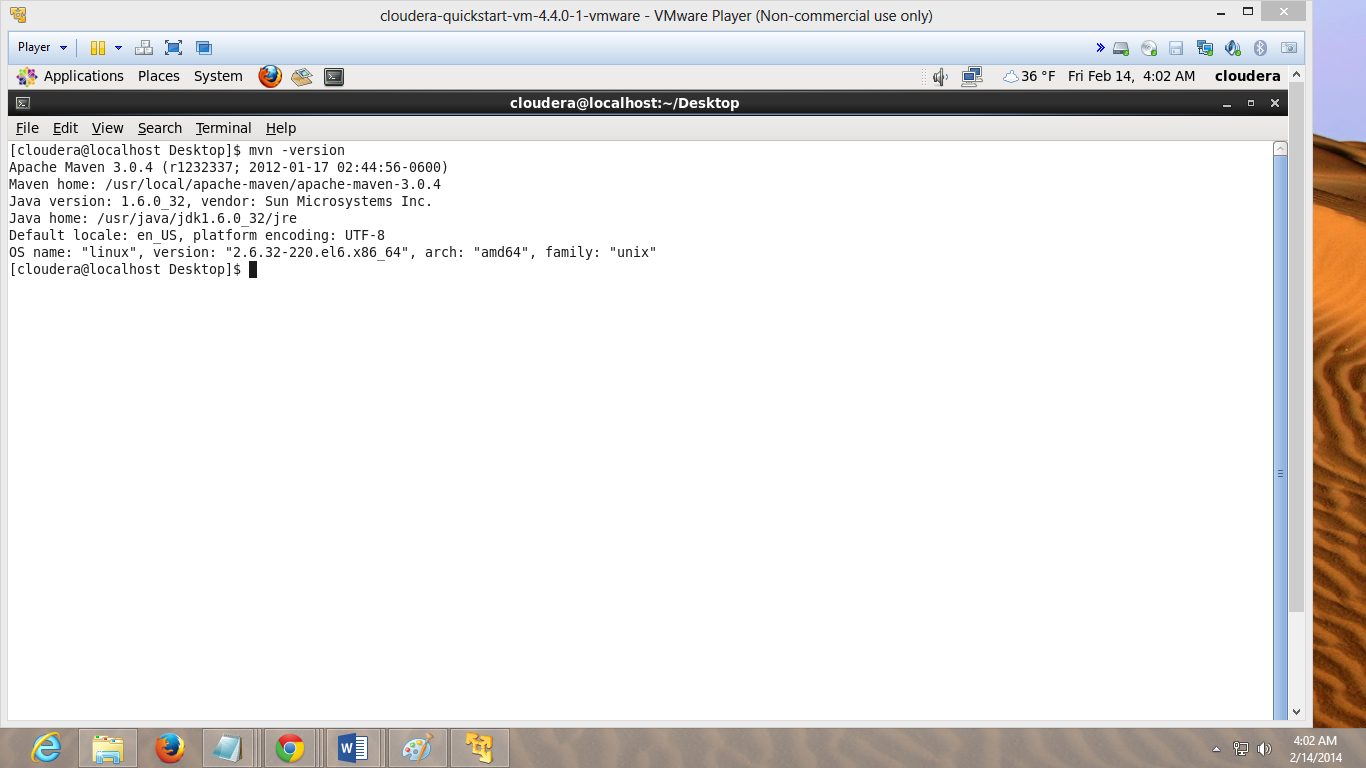


Figure Maven installed successfully

**Installation and Running Mahout:**

1. Initially we need to download the mahout pom file and zip file from the following link
   1. <http://repo2.maven.org/maven2/org/apache/mahout/mahout-distribution/0.6/>
2. Unzip the zip file and go to the mahout distribution folder which has been extracted, place the pom file inside the extracted folder.
3. Then run the following command:
   1. mvn –f mahout-distribution-0.6.pom -DskipTests install
4. Make sure the pom file is in right folder.
5. After successful execution of the command, go to /current directory/examples/bin folder
6. There we need to execute a shell script name “./classify-20newsgroup.sh”
7. Then we need to select either NaiveBayes or SGD

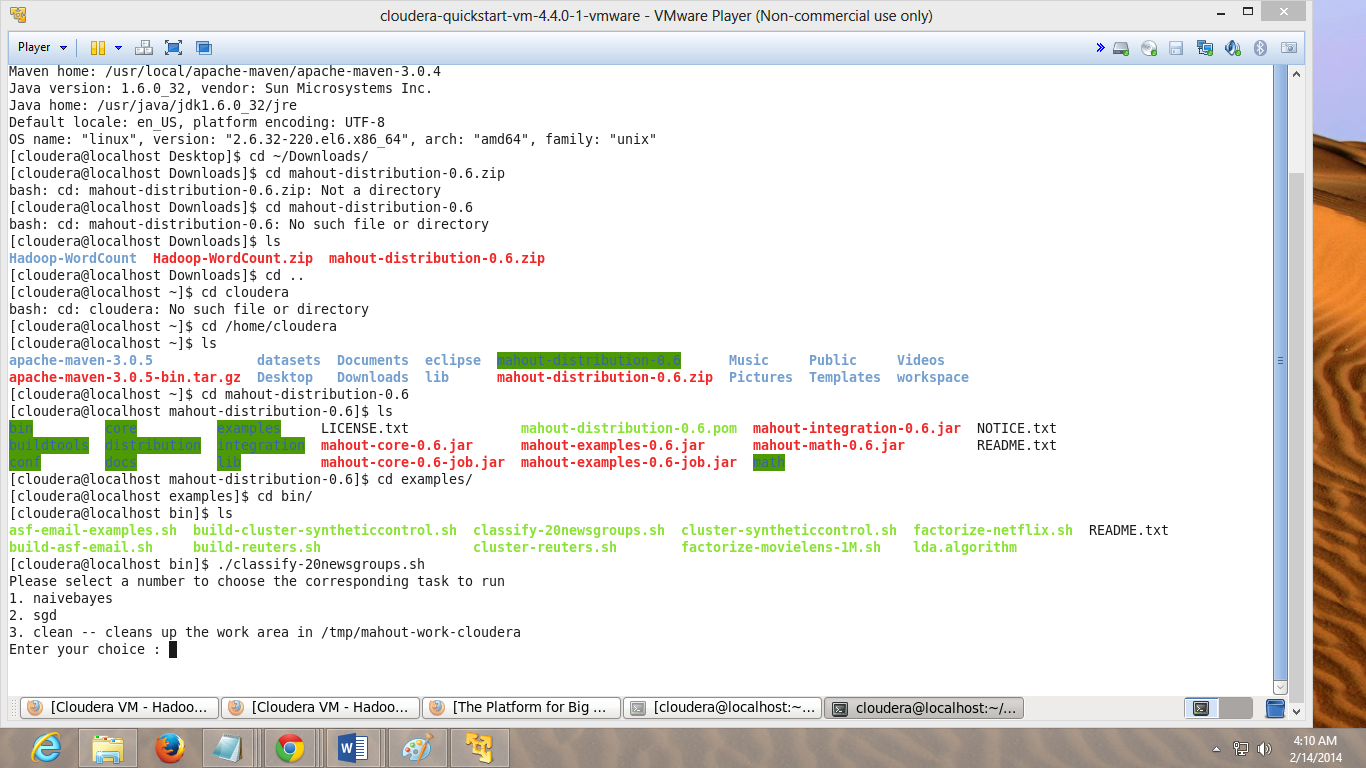


Figure selecting naivebayes

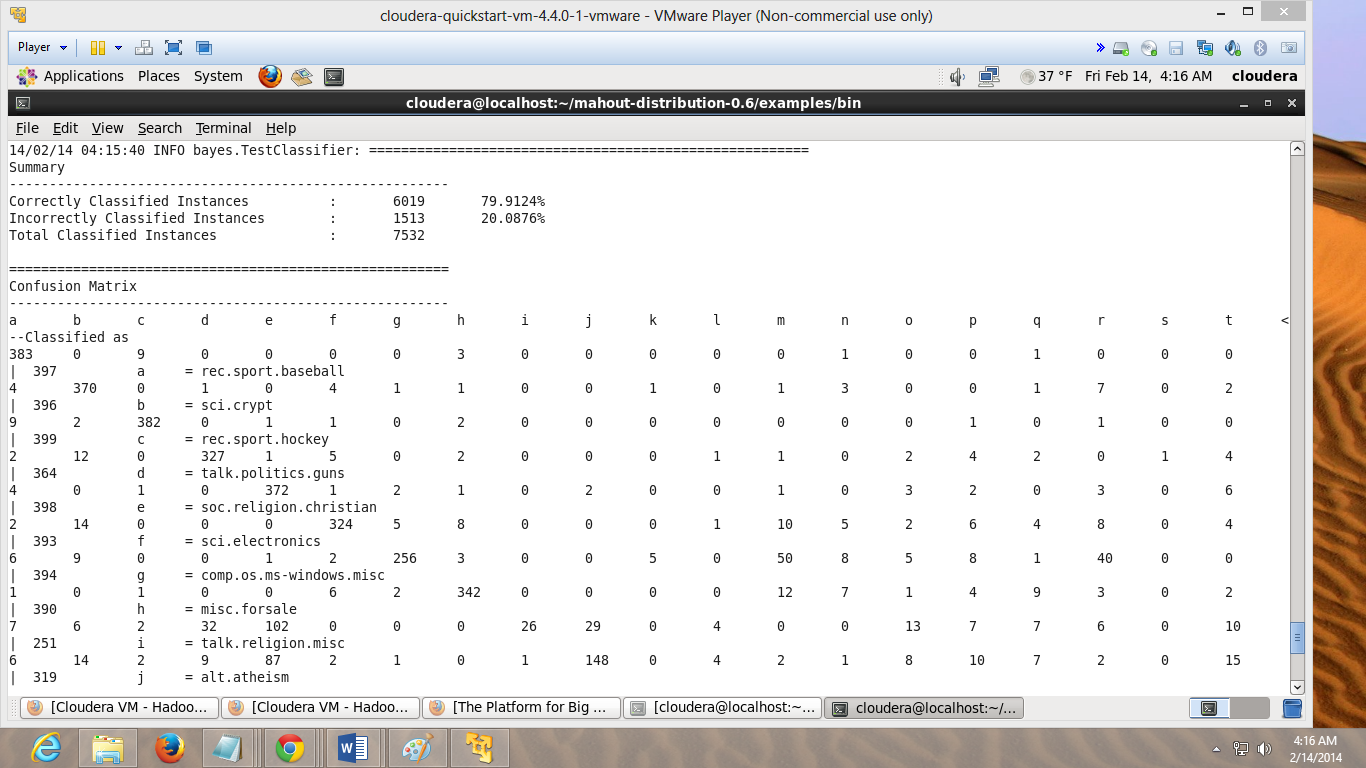


Figure executed naivebayes program

**Troubleshooting:**

* If there are any issues while running the shell script or mvn command with pom file make sure to check their permissions and change them respectively using chown and chmod commands as clodera and 777.