Predictable Crystal Ball

**Part 1. [2 points]  Set up a single node cluster and optionally an eclipse development environment to create and test your programs.**

1. Get VMWare or VirtualBox (install)

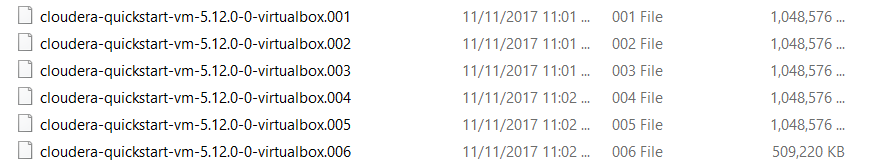
Download VirtualBox from this link: <https://www.virtualbox.org/wiki/Downloads>

and install it.

1. Get Cloudera (install)

Download Cloudera from this link: <http://www.cloudera.com/downloads.html>

There are 6 zipped files as below



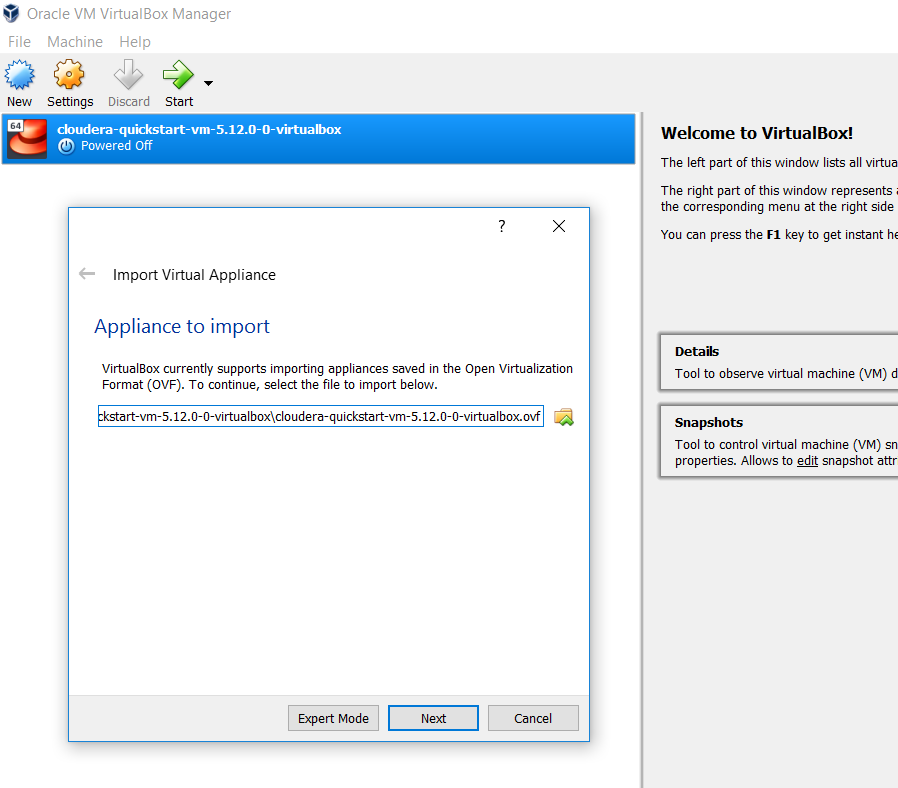
Then choose the first file and unzip it (it will extract all the content of all files to a sole folder). There are 2 important files extracted like that

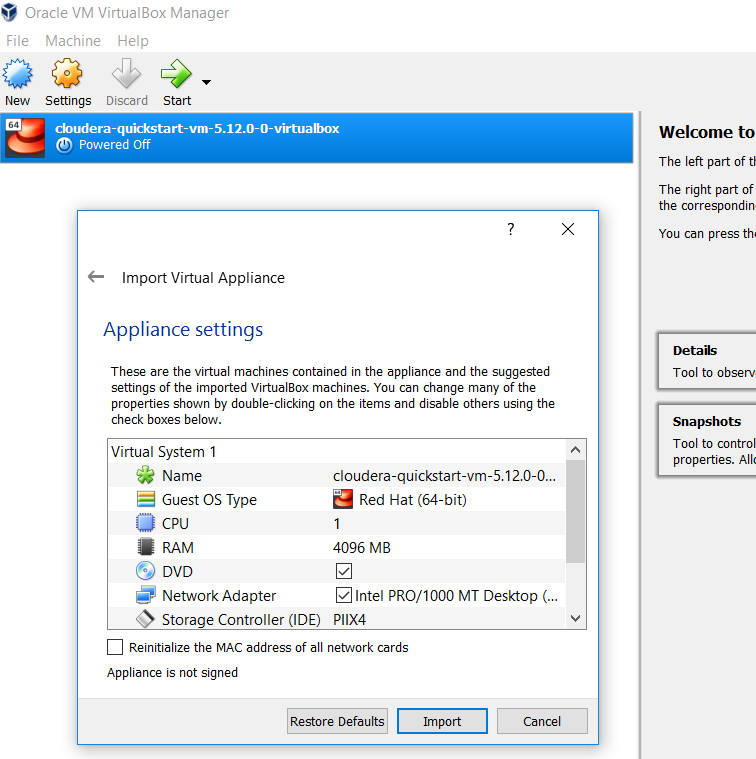


(b’) Import and start Cloudera

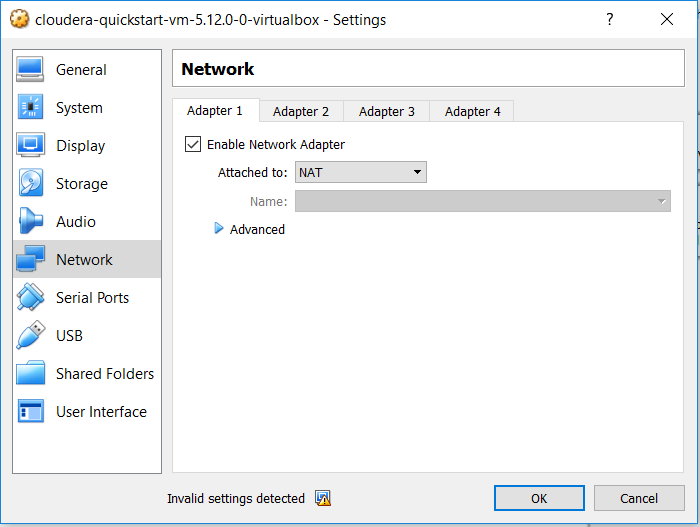
* + Import Cloudera image into VirtualBox

Start VirtualBox -> on the menu, click on File -> Import Appliance -> browse the cloudera extracted (.ovf) file to import it into VirtualBox

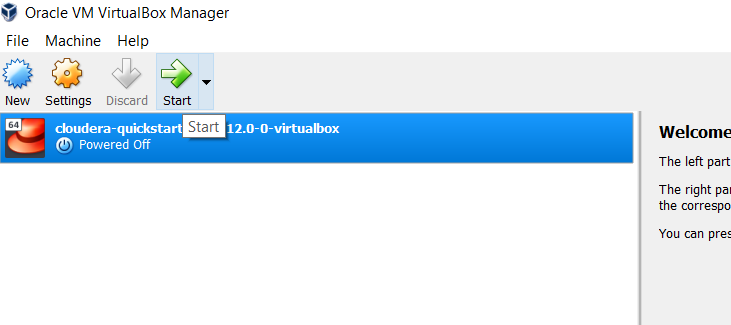


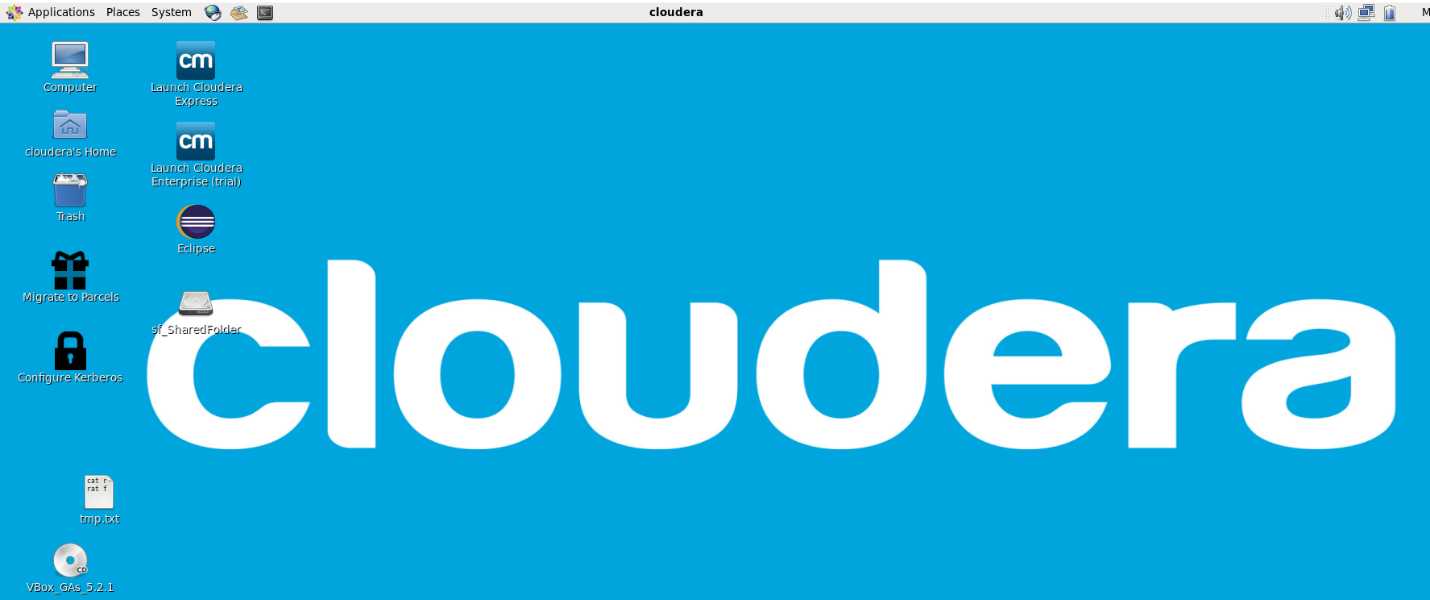


* + Enable network access go to Machine > Settings > Network and choose NAT in Attached to option as below

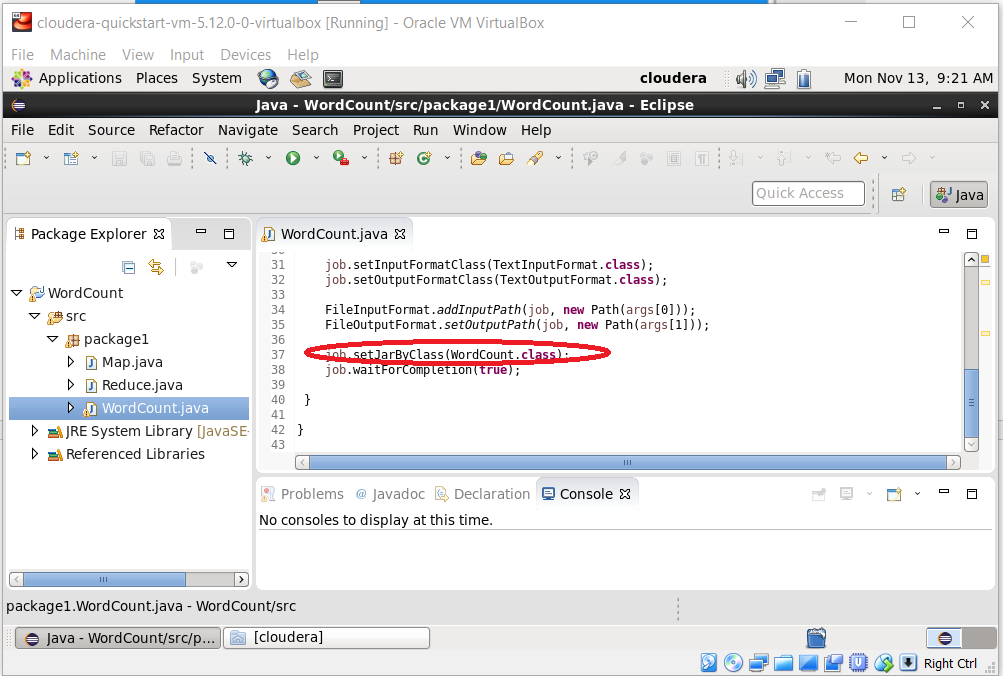


* + Start Cloudera and the desktop will appear like that





1. Get WordCount (test run)
   * Open Eclipse in Cloudera
   * Create WordCount project with the source code copied from [**https://wiki.apache.org/hadoop/WordCount**](https://wiki.apache.org/hadoop/WordCount)
   * Modify the source code a little bit, such as separate Map and Reduce into 2 classes and add 1 more line of code in WordCount class like this



* + Run the application to create the project configuration (it will have exception but ignore it)
  + Export the project as Runnable jar file (use the above configuration)
  + Copy the exported jar file to Home folder
  + Run the following command lines to create folder input

/user/cloudera/wordcount/input

* $ sudo su hdfs
* $ hadoop fs -mkdir /user/cloudera
* $ hadoop fs -chown cloudera /user/cloudera
* $ exit
* $ sudo su cloudera
* $ hadoop fs -mkdir /user/cloudera/wordcount /user/cloudera/wordcount/input
  + Create an input file named tmp.txt in the Desktop with the following content

[cat rat mat pat mat

rat fat cat bat fat]

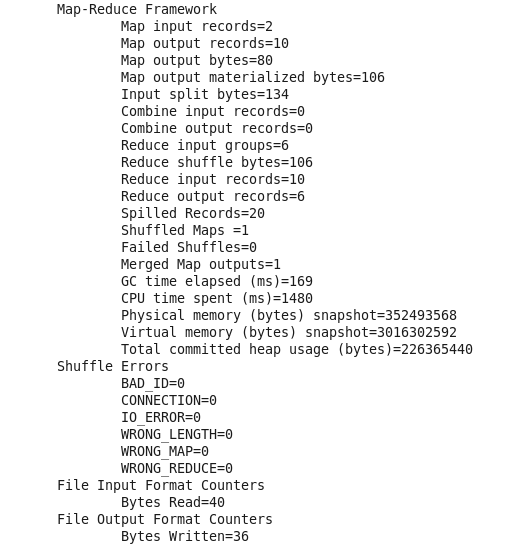
* + Copy this file into input folder by the command line

*hadoop fs -put /home/cloudera/Desktop/tmp.txt /user/cloudera/wordcount/input*

* + Execute this command line to run jar file

*hadoop jar wordcount.jar /user/cloudera/wordcount/input /user/cloudera/wordcount/output*

**

**