**Assignment:** Introduction to Hadoop

**Apache Hadoop Website**

<http://hadoop.apache.org/>

**Cloudera Quick Start VM -installation**

<http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/cloudera_quickstart_vm.html>

**Hadoop Commands**

<http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html>

**Hue:** Hue is the open source web interface to Hadoop

**Login Credentials**: User name and password is cloudera

**Hello World program**

* **Create a simple file and put in HDFS**

|  |  |
| --- | --- |
| **Type This** | **It Means** |
| touch testfile.txt | create an empty file name testfile.txt |
| echo "Hello world in HDFS" >> testfile.txt | display the line of text that is being inserted into the newly created testfile |
| cat testfile.txt | read and print out the file |
| hdfs dfs -put testfile.txt | put the file into HDFS |

**Home Assignment**

Now that you have uploaded and installed the Hadoop VM and have experienced creating a file you should go a step further and make some changes to the existing file. You should use the Hue environment and copy the "Hello world in HDFS" line ten times. Save the newly edited file under the same name and place.

Take a look how large is the file and report back the number of KB this newly updated file size.

**Mac instructions**

Please use the following instructions to download and install the Cloudera Quickstart VM with VirutalBox **before**proceeding to the **Getting Started with the Cloudera VM Environment**video. The screenshots are from a Mac but the instructions should be the same for Windows. Please see the discussion boards if you have any issues.

1. Go to <https://www.virtualbox.org/wiki/Downloads> to download and install VirtualBox for your computer.

2. Download the VirtualBox VM from <https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.4.2-0-virtualbox.zip> .

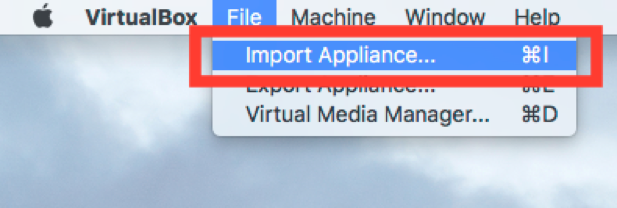
3. Unzip the VirtualBox VM:

**Mac**: Double click cloudera-quickstart-vm-5.4.2-0-virtualbox.zip

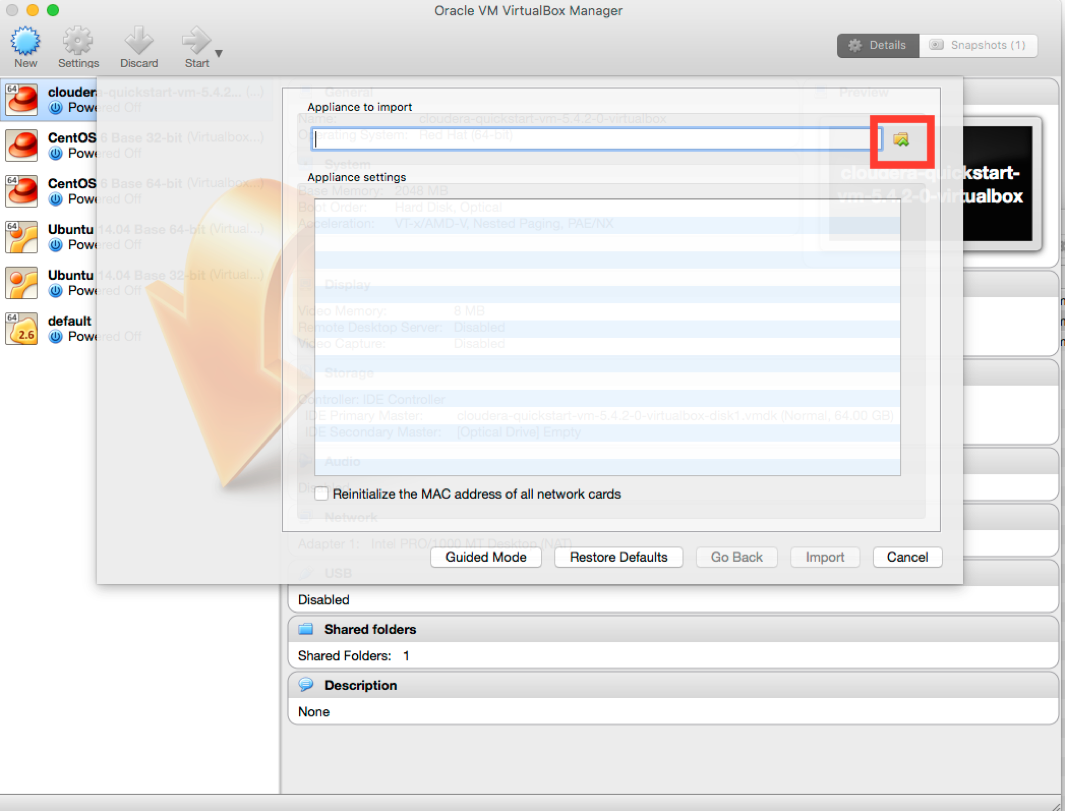
**Windows**: Right-click cloudera-quickstart-vm-5.4.2-0-virtualbox.zip and select “Extract All…”

4. Start VirtualBox

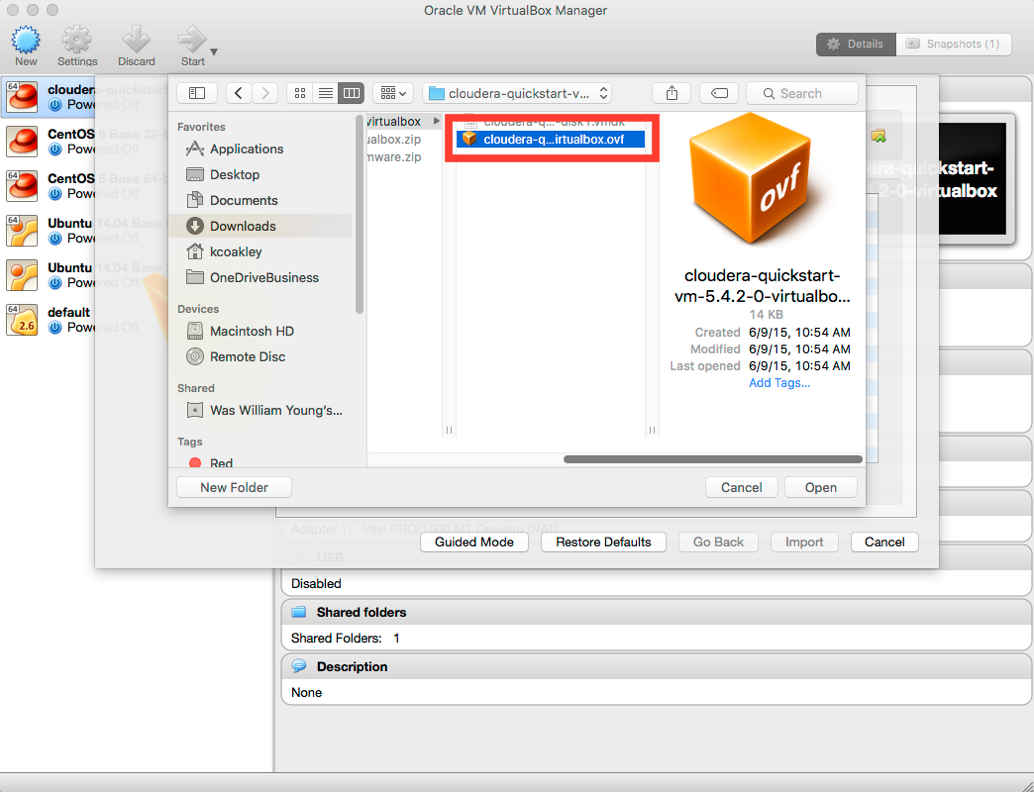
5. Import the VM by going to **File** -> **Import Appliance**



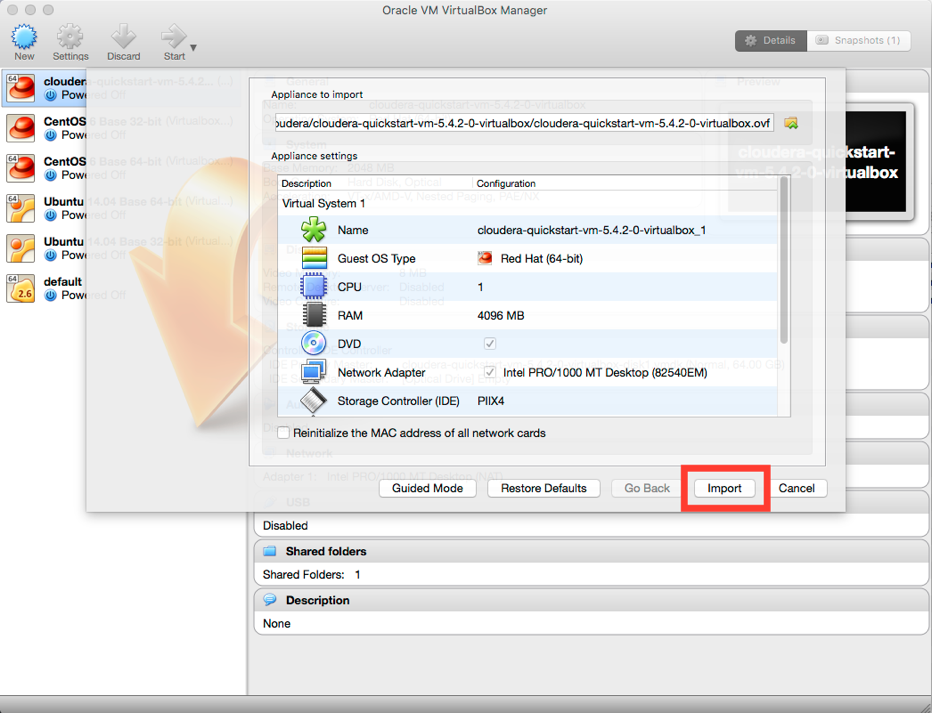
6. Click the Folder icon.



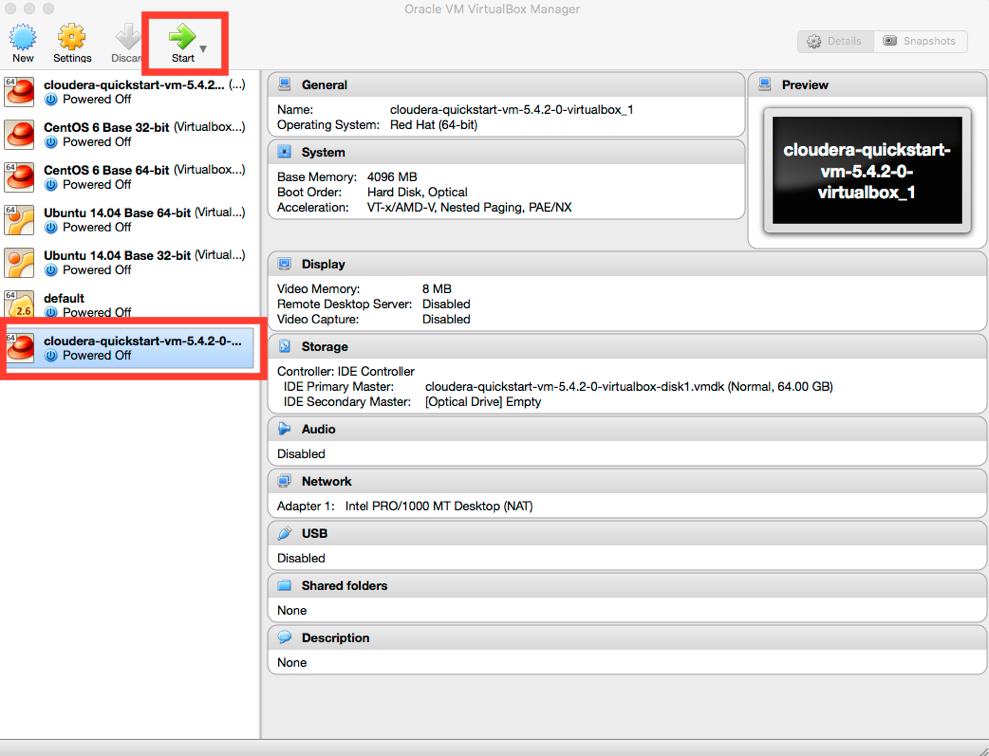
7. Select the **cloudera-quickstart-vm-5.4.2-0-virtualbox.ovf** from the Folder where you unzipped the VirtualBox VM.



8. Click **Import** and wait for the VirtualBox VM to import.



9. Select the **cloudera-quickstart-vm-5.4.2-0**VM from the left nav and click the **Start** button.



It could take a few minutes for the Virtual Machine to start.