**Name: Mandar Gadkari**

**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 |

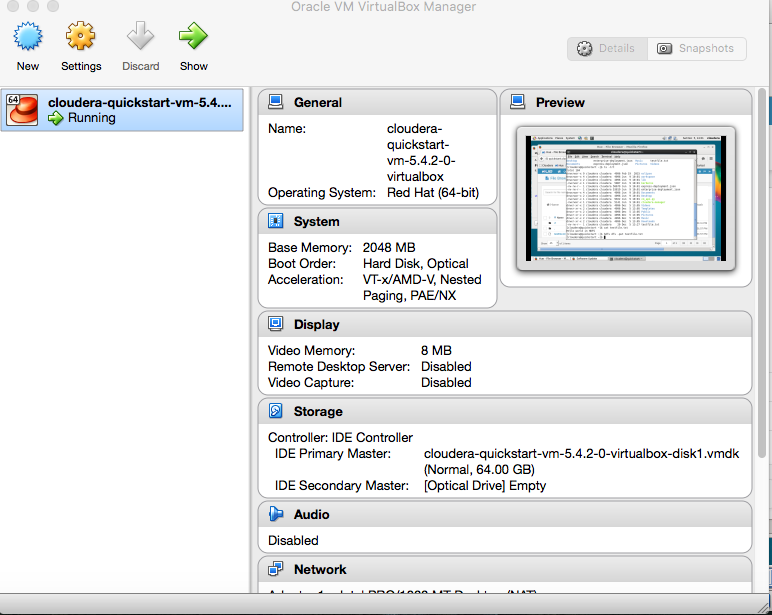
**Home Assignment**

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.

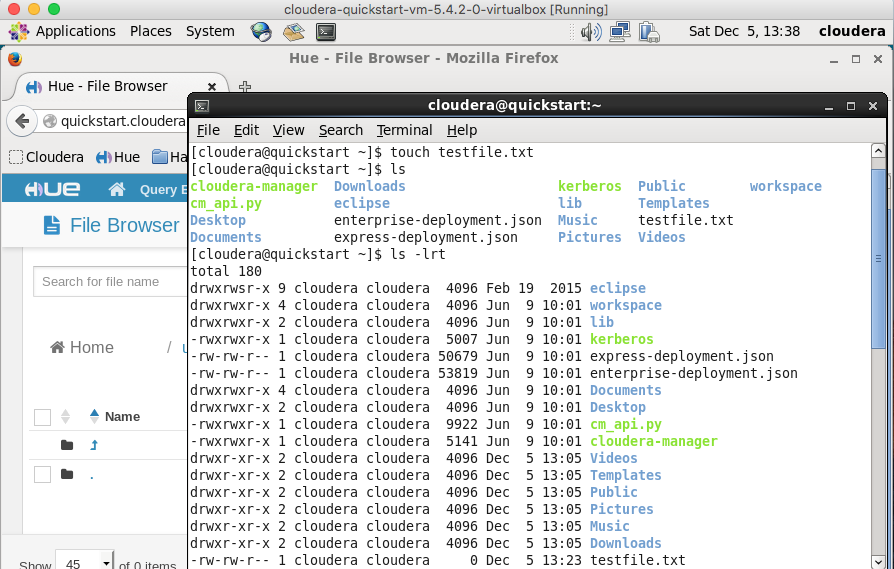
**Solution:**

After installing VM BOX and Cloudera quickstart VM, you log in into application “Hue” and create a file “testfile.txt”.



Initially there is no data in this file hence the file size is 0 bytes. As shown in the picture below:

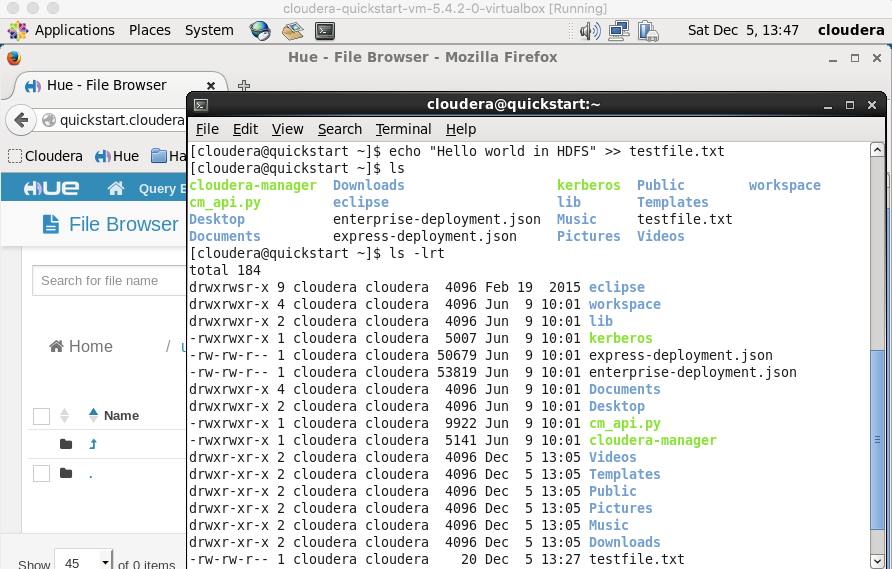
Command: touch testfile.txt



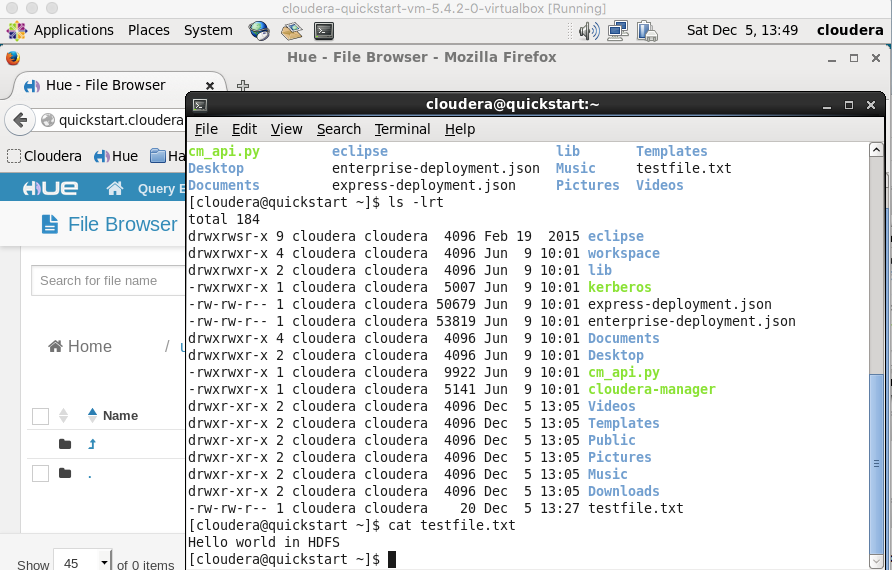
After that , you write a line in this test file using the command.

Command: echo "Hello world in HDFS" >> testfile.txt

Now we see that the file size has increased to 20 Bytes as shown in the figure below:



To display the content of the test file on the terminal, you use the below command as shown in the screen shot. Command: cat testfile.txt

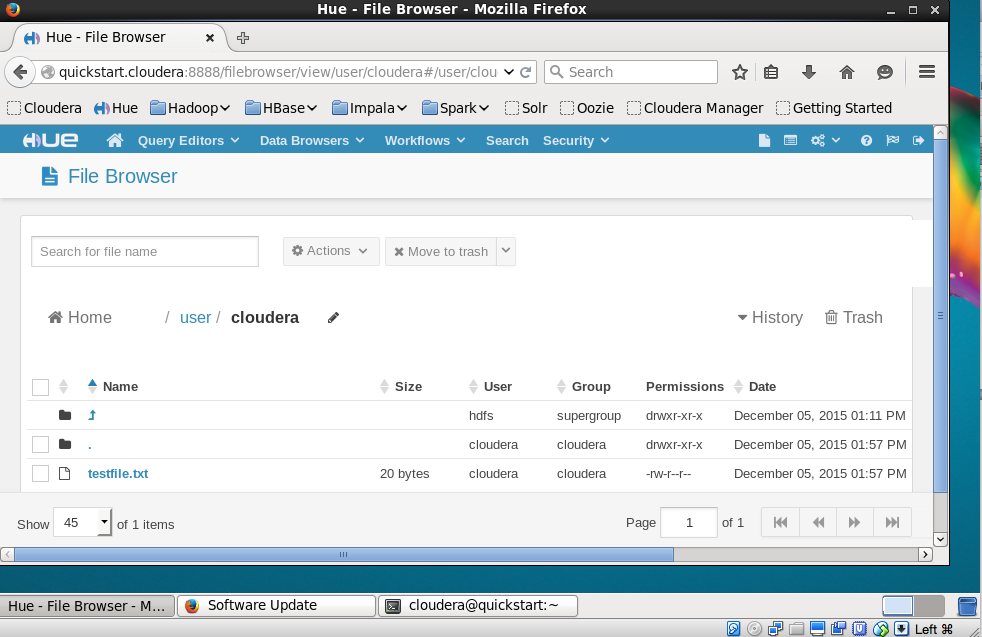


You can now, load the file from the terminal to HDFS with the following command,

Command: hdfs dfs -put testfile.txt

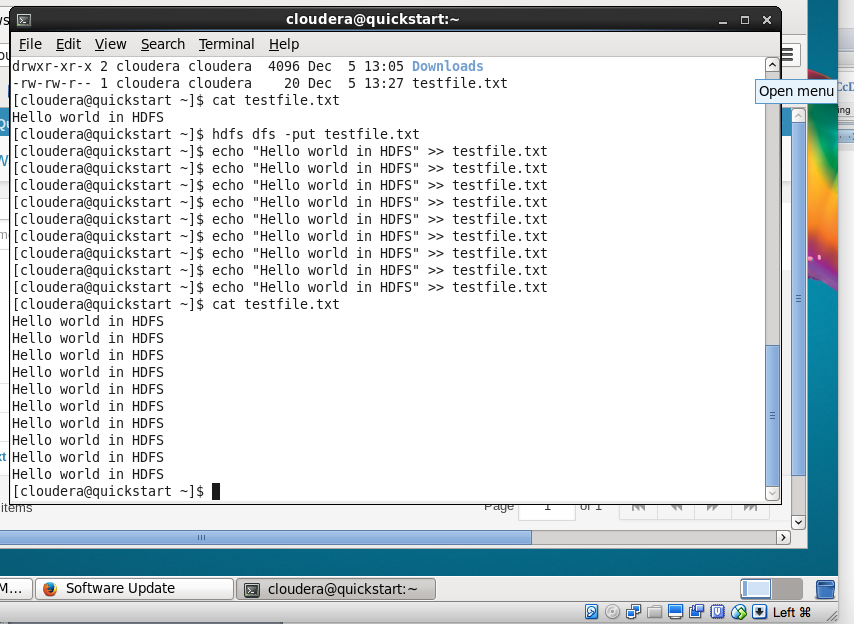


After you upload the file in HDFS database, you can see it in the directory. It shows the file size in the HDFS directory as 20 bytes as well. See the screen shot below.

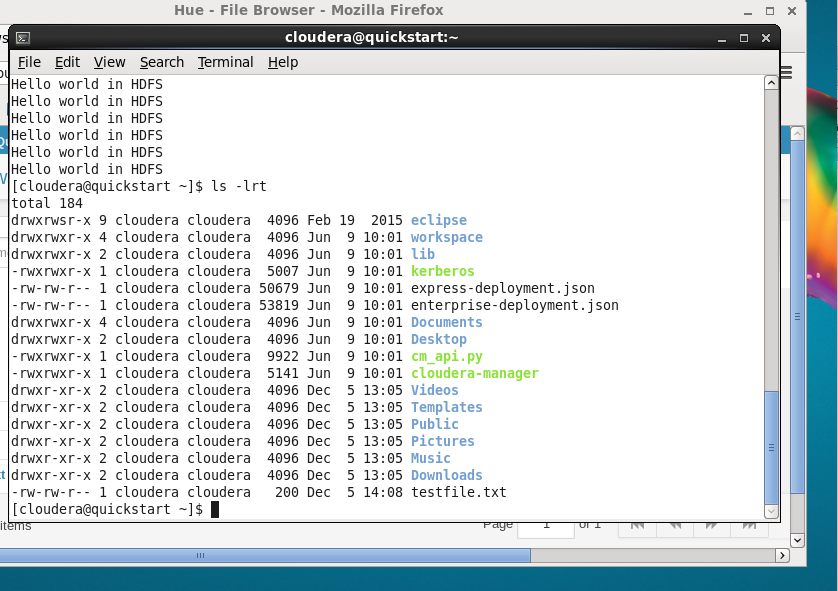


Now you go back to terminal and edit the testfile .txt by adding 9 more lines using the same command 9 more times. You can display the contents of the testfile.txt

Screenshot: New testfile.txt



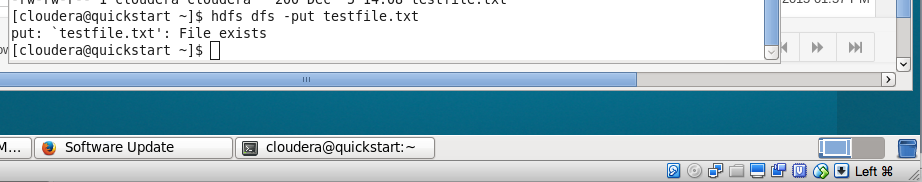
Now the file size is 200 Bytes as shown below:



You can now, load the file from the terminal to HDFS with the following command,

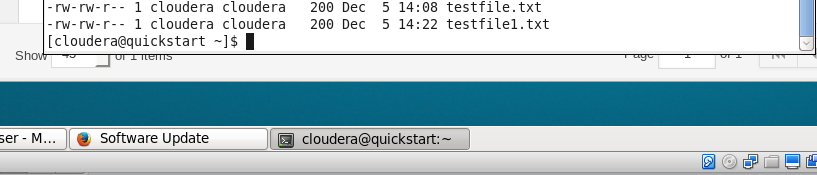
Command: hdfs dfs -put testfile.txt

It gives an error saying the file testfile.txt already exists.

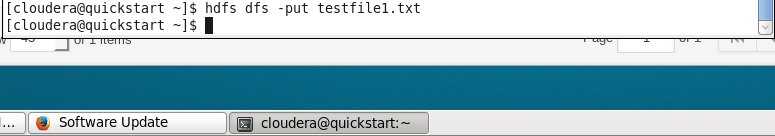


The testfile.txt still shows in the HDFS directory structure as 20 bytes.

You can save the testfile.txt as testfile1.txt in terminal and then upload it on HFDS directory.



Upload testfile1.txt in the HDFS directory.



Now if you see in the HDFS directory the new file “testfile1.txt” and older file “testfile.txt” both show up.

testfile.txt is 20 Bytes

testfile1.txt is 200 Bytes.

Screen Shot:

