

# Accessing HDFS

---

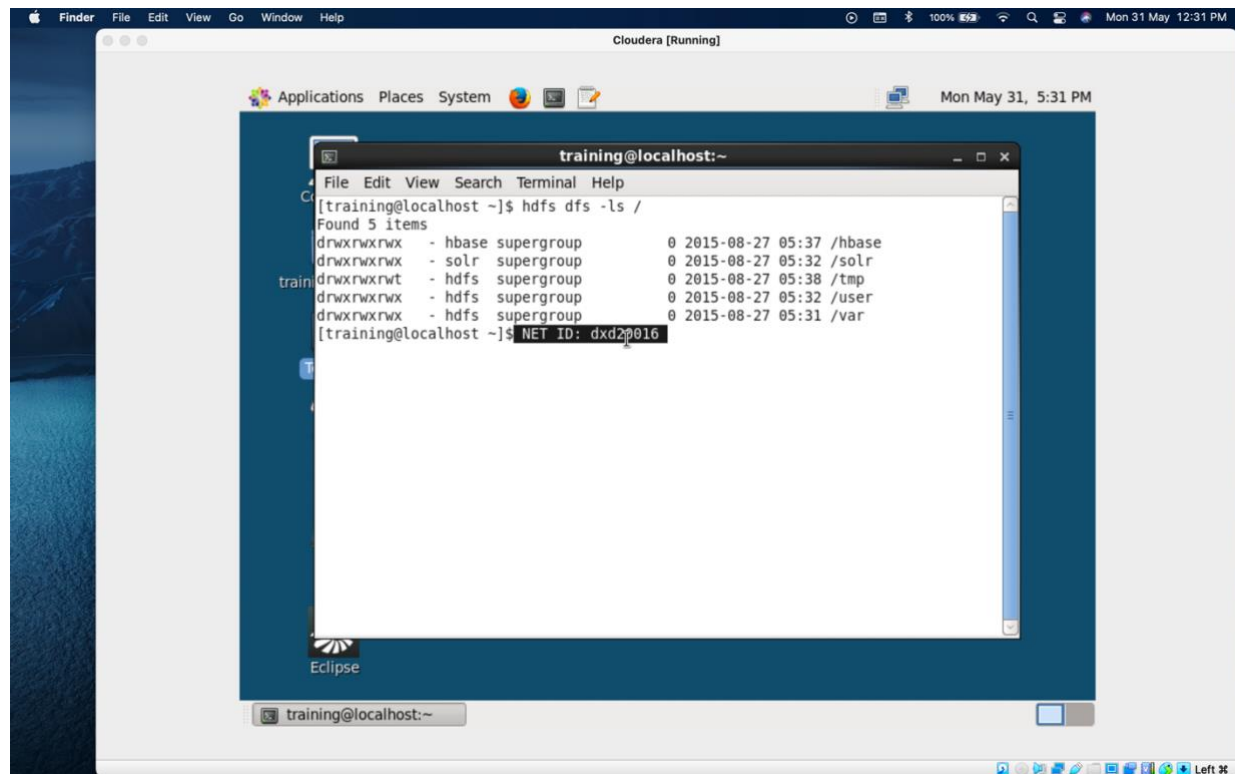
Divyansh Dahiya

## LAB Chapter 03: ACCESS HDFS

### Access HDFS with Command Line and Hue

In this lab I practiced working with HDFS. I used command line and Hue Interface to manipulate files stored in HDFS.

1. First I opened the terminal window and entered the following command: `hdfs dfs -ls /`  
After running this command I was able to see the contents of root directory in HDFS.  
There were total of 5 entries.

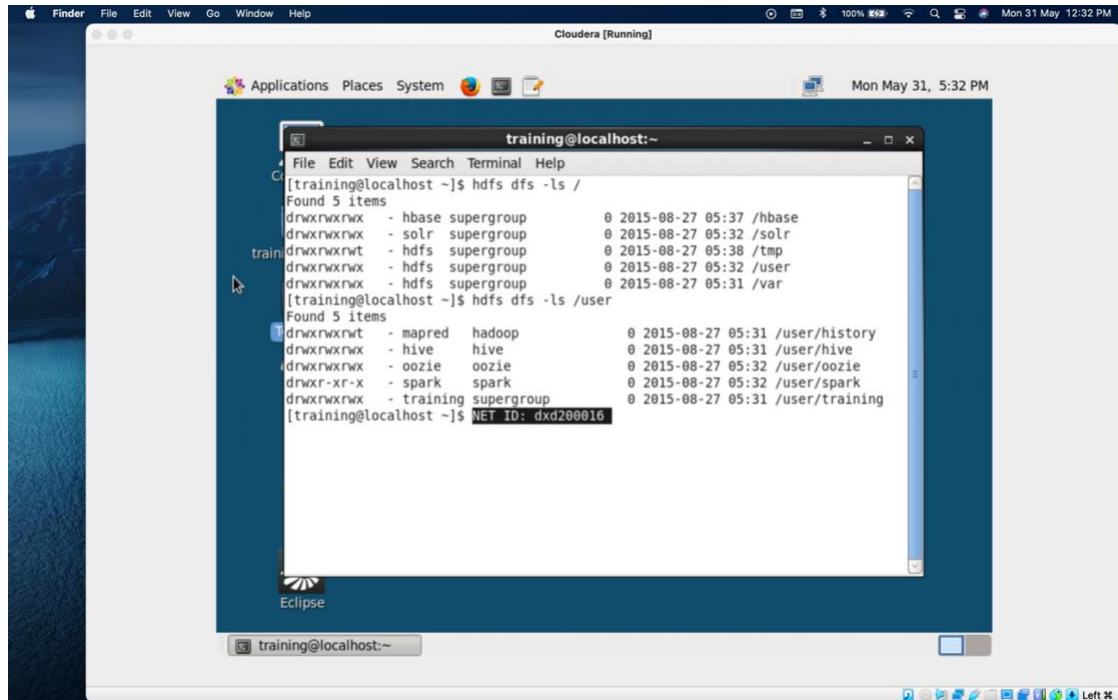


The screenshot shows a terminal window titled "training@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The command `hdfs dfs -ls /` has been executed, resulting in the following output:

```
[training@localhost ~]$ hdfs dfs -ls /
Found 5 items
drwxrwxrwx - hbase supergroup      0 2015-08-27 05:37 /hbase
drwxrwxrwx - solr supergroup        0 2015-08-27 05:32 /solr
drwxrwxrwt - hdfs supergroup        0 2015-08-27 05:38 /tmp
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:32 /user
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:31 /var
```

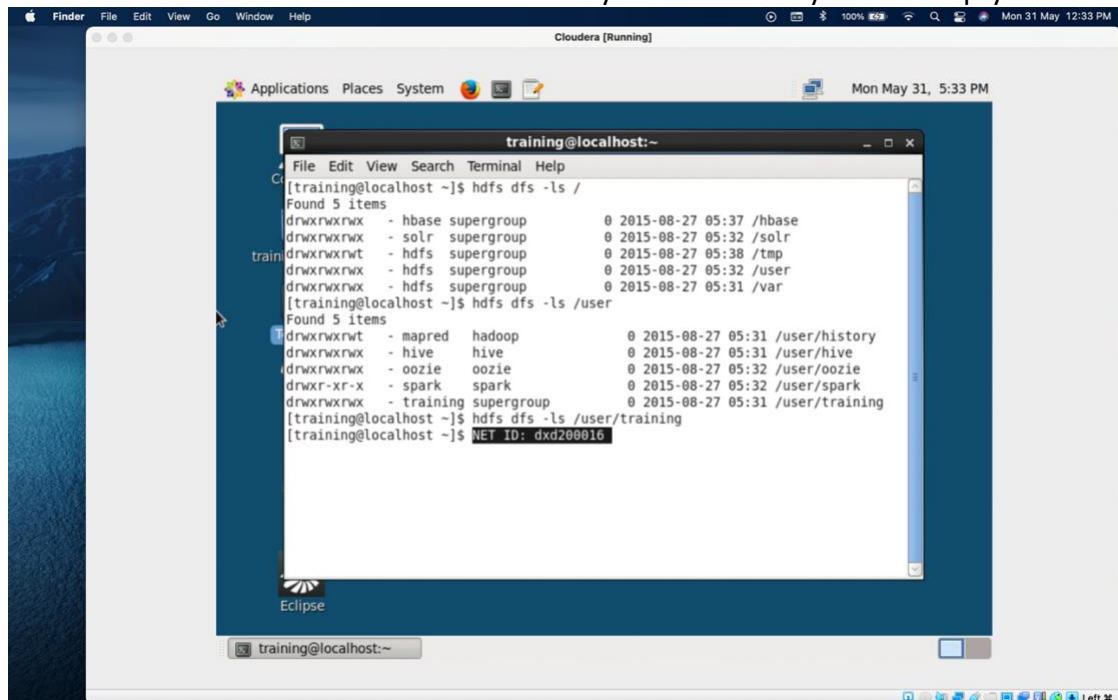
The terminal window is running within a Cloudera [Running] environment. The system clock shows Mon May 31, 5:31 PM. The terminal window is titled "training@localhost:~" and has a menu bar (File, Edit, View, Search, Terminal, Help). The command `hdfs dfs -ls /` has been executed, resulting in the following output:

2. Then I ran the following command: `hdfs dfs -ls /user`  
After running the command I was able to see user directory and the all the listings in the directory.



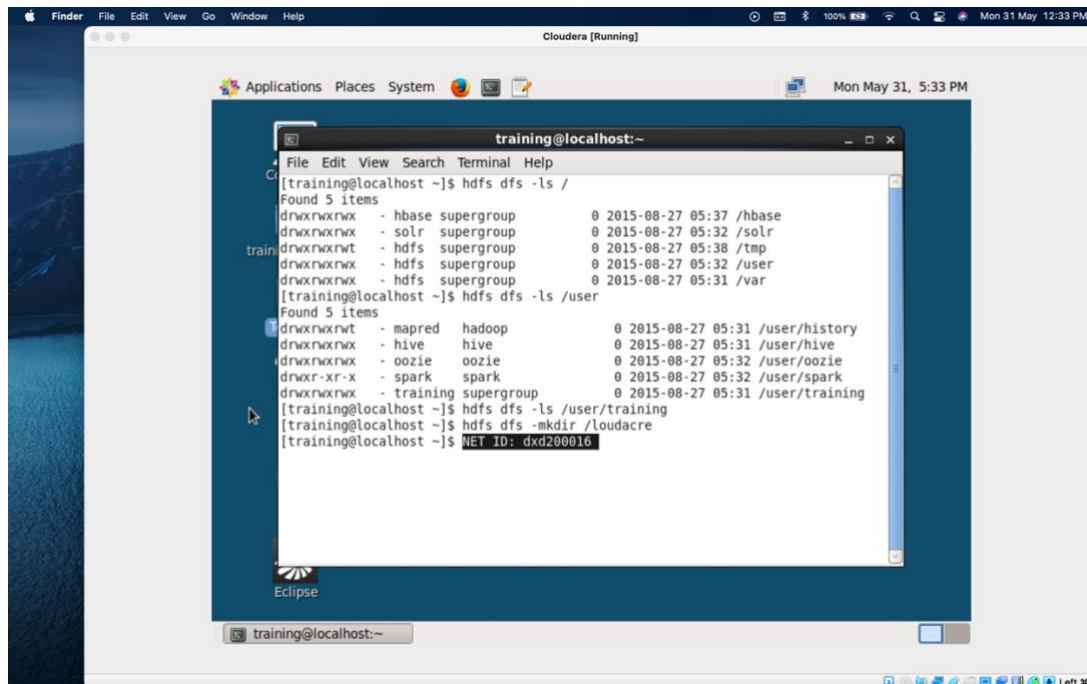
```
training@localhost:~$ hdfs dfs -ls /
Found 5 items
drwxrwxrwx - hbase supergroup      0 2015-08-27 05:37 /hbase
drwxrwxrwx - solr supergroup        0 2015-08-27 05:32 /solr
drwxrwxrwt - hdfs supergroup        0 2015-08-27 05:38 /tmp
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:32 /user
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:31 /var
training@localhost:~$ hdfs dfs -ls /user
Found 5 items
drwxrwxrwt - mapred hadoop          0 2015-08-27 05:31 /user/history
drwxrwxrwx - hive hive              0 2015-08-27 05:31 /user/hive
drwxrwxrwx - oozie oozie            0 2015-08-27 05:32 /user/oozie
drwxr-xr-x - spark spark            0 2015-08-27 05:32 /user/spark
drwxrwxrwx - training supergroup    0 2015-08-27 05:31 /user/training
training@localhost:~$
```

3. Then I ran the following command: `hdfs dfs -ls /user/training`  
I ran this command to view the contents of my home directory. It was empty as there are no files.



```
training@localhost:~$ hdfs dfs -ls /user/training
Found 0 items
training@localhost:~$
```

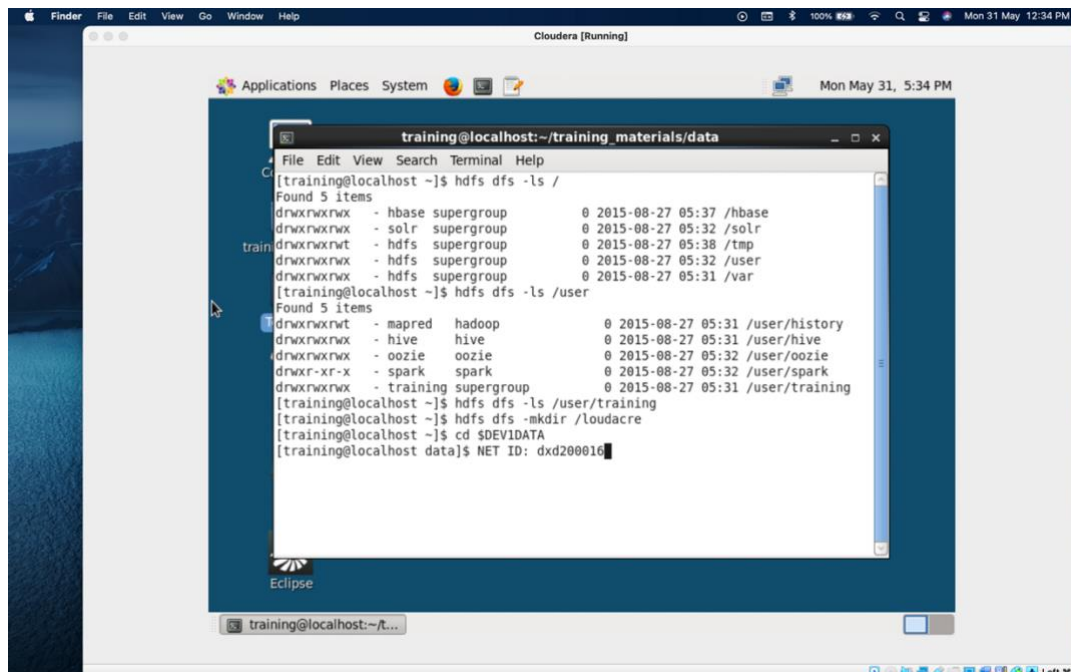
- After that, I ran a command to create a new directory while working on future labs and assignments



The screenshot shows a terminal window titled "training@localhost:~" within a Cloudera [Running] environment. The terminal displays the following commands and output:

```
[training@localhost ~]$ hdfs dfs -ls /
Found 5 items
drwxrwxrwx - hbase supergroup      0 2015-08-27 05:37 /hbase
drwxrwxrwx - solr supergroup        0 2015-08-27 05:32 /solr
drwxrwxrwt - hdfs supergroup        0 2015-08-27 05:38 /tmp
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:32 /user
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:31 /var
[training@localhost ~]$ hdfs dfs -ls /user
Found 5 items
drwxrwxrwt - mapred  hadoop      0 2015-08-27 05:31 /user/history
drwxrwxrwx - hive   hive        0 2015-08-27 05:31 /user/hive
drwxrwxrwx - oozie  oozie       0 2015-08-27 05:32 /user/oozie
drwxr-xr-x - spark  spark       0 2015-08-27 05:32 /user/spark
drwxrwxrwx - training supergroup 0 2015-08-27 05:31 /user/training
[training@localhost ~]$ hdfs dfs -ls /user/training
[training@localhost ~]$ hdfs dfs -mkdir /loudacre
[training@localhost ~]$
```

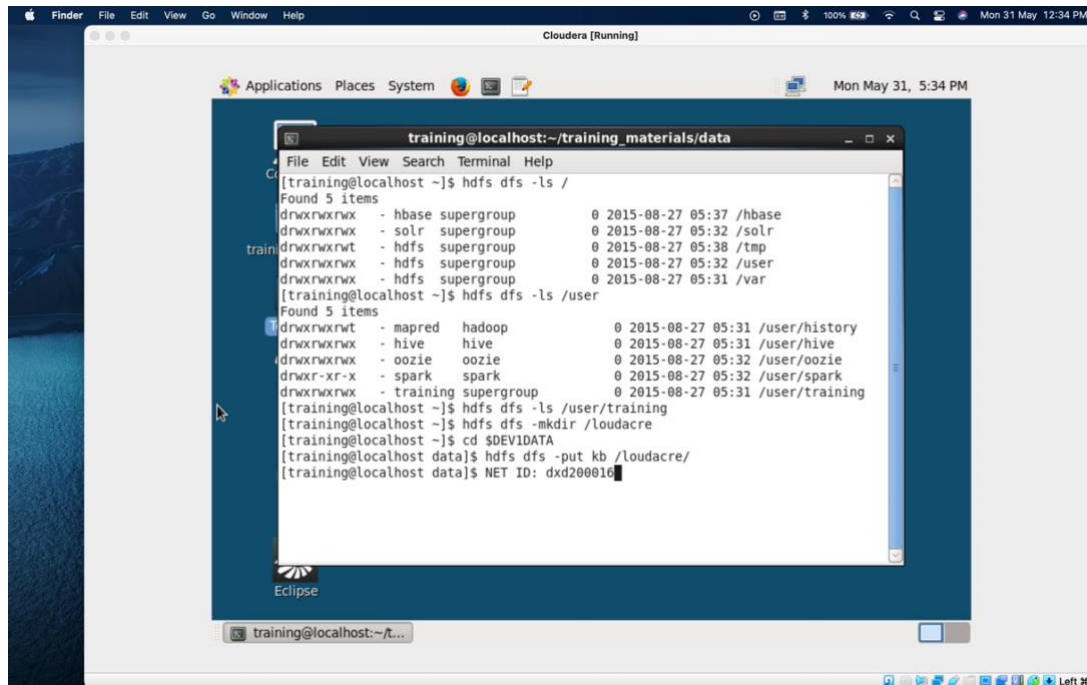
- Then I ran a command to change the directory of the command line terminal.



The screenshot shows the same terminal window, but the title bar now reads "training@localhost:~/training\_materials/data". The terminal displays the following commands and output:

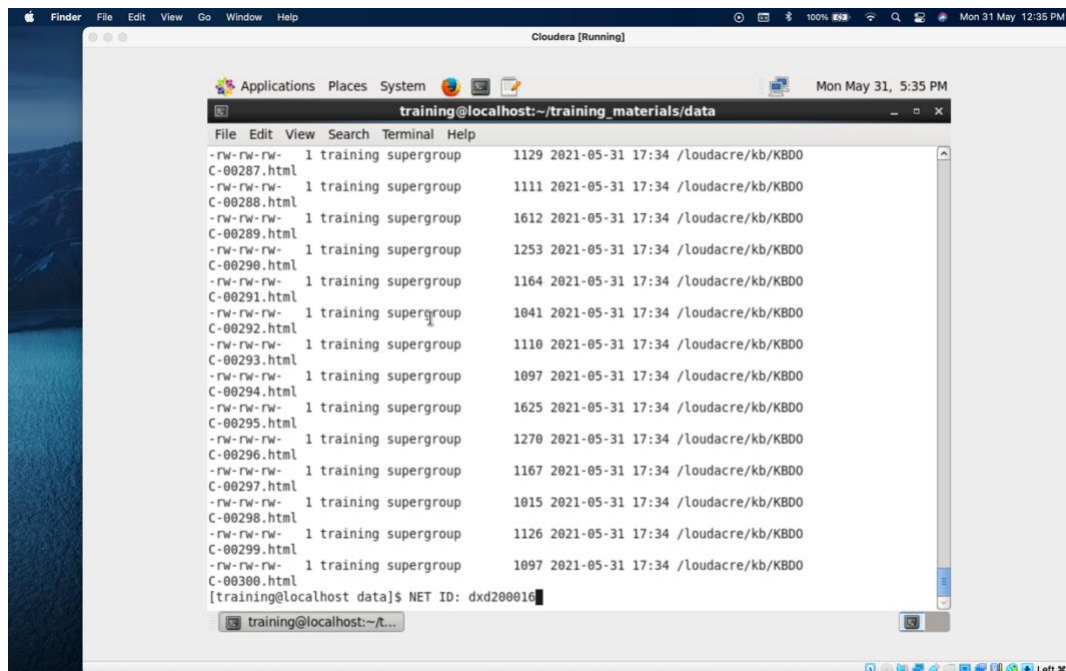
```
[training@localhost ~]$ hdfs dfs -ls /
Found 5 items
drwxrwxrwx - hbase supergroup      0 2015-08-27 05:37 /hbase
drwxrwxrwx - solr supergroup        0 2015-08-27 05:32 /solr
drwxrwxrwt - hdfs supergroup        0 2015-08-27 05:38 /tmp
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:32 /user
drwxrwxrwx - hdfs supergroup        0 2015-08-27 05:31 /var
[training@localhost ~]$ hdfs dfs -ls /user
Found 5 items
drwxrwxrwt - mapred  hadoop      0 2015-08-27 05:31 /user/history
drwxrwxrwx - hive   hive        0 2015-08-27 05:31 /user/hive
drwxrwxrwx - oozie  oozie       0 2015-08-27 05:32 /user/oozie
drwxr-xr-x - spark  spark       0 2015-08-27 05:32 /user/spark
drwxrwxrwx - training supergroup 0 2015-08-27 05:31 /user/training
[training@localhost ~]$ hdfs dfs -ls /user/training
[training@localhost ~]$ hdfs dfs -mkdir /loudacre
[training@localhost ~]$ cd $DEV1DATA
[training@localhost data]$ NET ID: dxd200010
```

6. Now, to upload the contents of the folder KB, I ran the following command: `hdfs dfs -ls /loudacre/kb`  
Below is the screenshot, that the command ran successfully and I was able to upload the contents of KB.



```
training@localhost:~/training_materials/data
[training@localhost ~]$ hdfs dfs -ls /
Found 5 items
drwxr-xr-x - hbase supergroup      0 2015-08-27 05:37 /hbase
drwxr-xr-x - solr supergroup        0 2015-08-27 05:32 /solr
drwxr-xr-x - hdfs supergroup        0 2015-08-27 05:38 /tmp
drwxr-xr-x - hdfs supergroup        0 2015-08-27 05:32 /user
drwxr-xr-x - hdfs supergroup        0 2015-08-27 05:31 /var
[training@localhost ~]$ hdfs dfs -ls /user
Found 5 items
drwxr-xr-x - mapred hadoop          0 2015-08-27 05:31 /user/history
drwxr-xr-x - hive hive              0 2015-08-27 05:31 /user/hive
drwxr-xr-x - oozie oozie            0 2015-08-27 05:32 /user/oozie
drwxr-xr-x - spark spark           0 2015-08-27 05:32 /user/spark
drwxr-xr-x - training supergroup    0 2015-08-27 05:31 /user/training
[training@localhost ~]$ hdfs dfs -ls /user/training
[training@localhost ~]$ hdfs dfs -mkdir /loudacre
[training@localhost ~]$ cd $DEV1DATA
[training@localhost data]$ hdfs dfs -put kb /loudacre/
[training@localhost data]$ NET ID: dxd200016
```

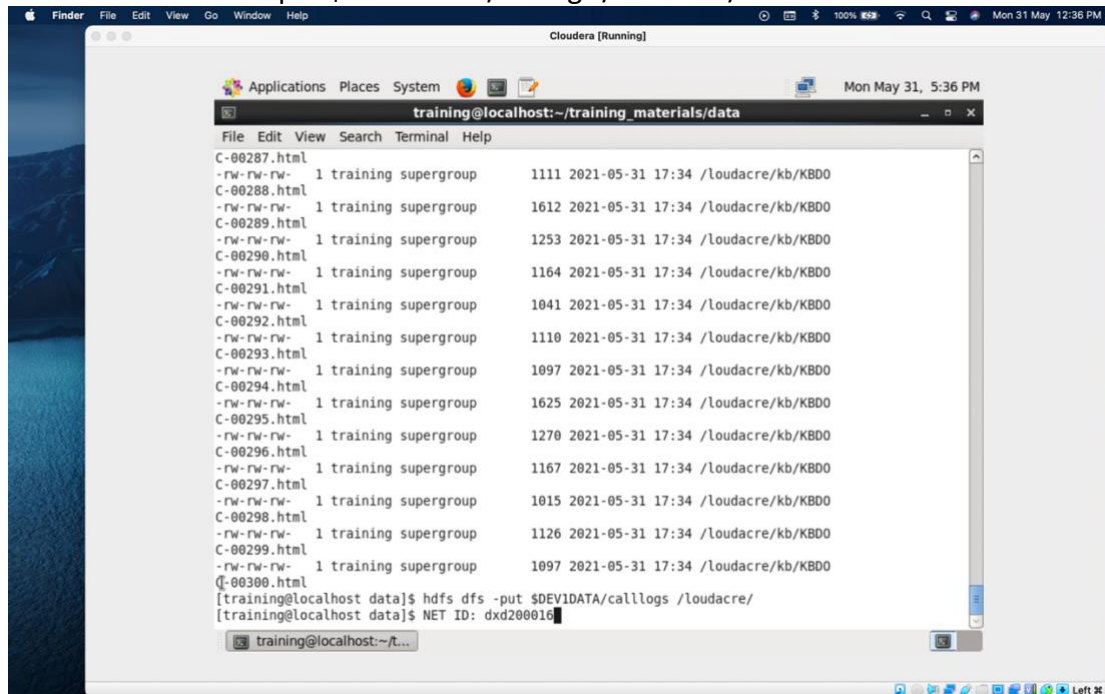
7. After that, to verify if my previous upload command ran successfully, I ran the following command: `hdfs dfs -ls /loudacre/kb`  
Below is the screenshot of all the contents in the loudacre/kb in HDFS



```
training@localhost:~/training_materials/data
-rw-rw-rw- 1 training supergroup 1129 2021-05-31 17:34 /loudacre/kb/KB00
C-00287.html
-rw-rw-rw- 1 training supergroup 1111 2021-05-31 17:34 /loudacre/kb/KB00
C-00288.html
-rw-rw-rw- 1 training supergroup 1612 2021-05-31 17:34 /loudacre/kb/KB00
C-00289.html
-rw-rw-rw- 1 training supergroup 1253 2021-05-31 17:34 /loudacre/kb/KB00
C-00290.html
-rw-rw-rw- 1 training supergroup 1164 2021-05-31 17:34 /loudacre/kb/KB00
C-00291.html
-rw-rw-rw- 1 training supergroup 1041 2021-05-31 17:34 /loudacre/kb/KB00
C-00292.html
-rw-rw-rw- 1 training supergroup 1110 2021-05-31 17:34 /loudacre/kb/KB00
C-00293.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KB00
C-00294.html
-rw-rw-rw- 1 training supergroup 1625 2021-05-31 17:34 /loudacre/kb/KB00
C-00295.html
-rw-rw-rw- 1 training supergroup 1270 2021-05-31 17:34 /loudacre/kb/KB00
C-00296.html
-rw-rw-rw- 1 training supergroup 1167 2021-05-31 17:34 /loudacre/kb/KB00
C-00297.html
-rw-rw-rw- 1 training supergroup 1015 2021-05-31 17:34 /loudacre/kb/KB00
C-00298.html
-rw-rw-rw- 1 training supergroup 1126 2021-05-31 17:34 /loudacre/kb/KB00
C-00299.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KB00
C-00300.html
[training@localhost data]$ NET ID: dxd200016
```

8. Also, I practiced uploading some data and removing them as they were no longer needed.  
Below is the screenshot of the same:

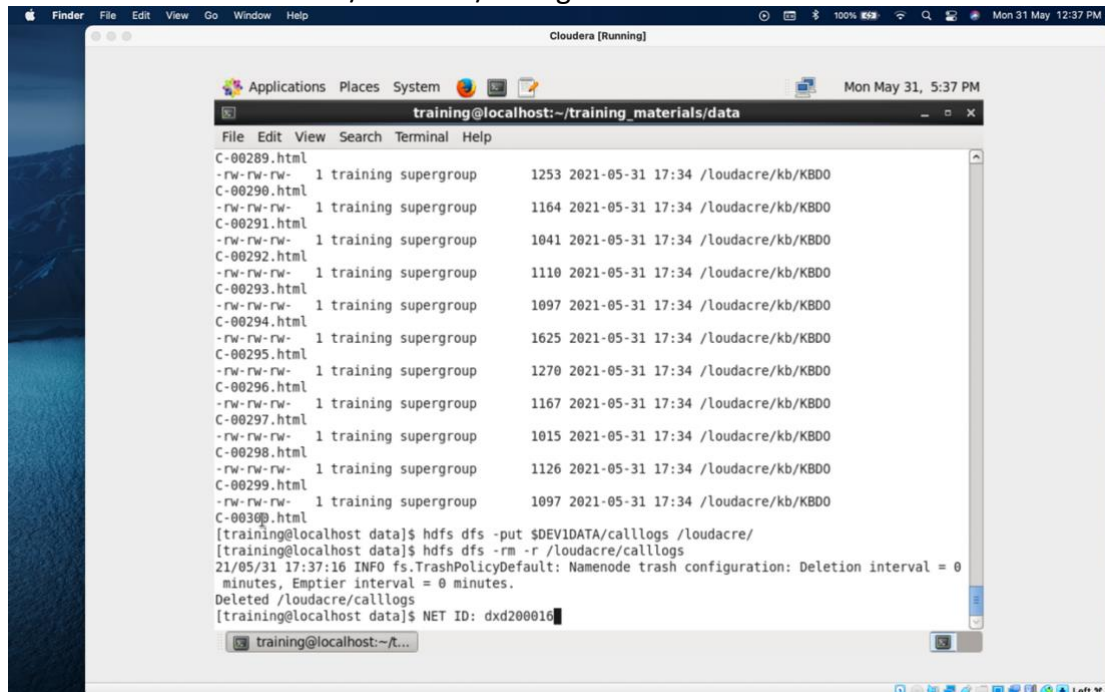
Command: `hdfs dfs -put $DEV1DATA/callogs /loudacre/`



The screenshot shows a Cloudera terminal window titled "training@localhost:~/training\_materials/data". The terminal displays a list of files in the current directory, including C-00287.html through C-00300.html. Below the file list, the command `hdfs dfs -put $DEV1DATA/callogs /loudacre/` is entered and executed. The output shows the command was successful, and the terminal prompt returns to `[training@localhost data]$`.

```
training@localhost:~/training_materials/data
File Edit View Search Terminal Help
C-00287.html
-rw-rw-rw- 1 training supergroup 1111 2021-05-31 17:34 /loudacre/kb/KBD0
C-00288.html
-rw-rw-rw- 1 training supergroup 1612 2021-05-31 17:34 /loudacre/kb/KBD0
C-00289.html
-rw-rw-rw- 1 training supergroup 1253 2021-05-31 17:34 /loudacre/kb/KBD0
C-00290.html
-rw-rw-rw- 1 training supergroup 1164 2021-05-31 17:34 /loudacre/kb/KBD0
C-00291.html
-rw-rw-rw- 1 training supergroup 1041 2021-05-31 17:34 /loudacre/kb/KBD0
C-00292.html
-rw-rw-rw- 1 training supergroup 1110 2021-05-31 17:34 /loudacre/kb/KBD0
C-00293.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KBD0
C-00294.html
-rw-rw-rw- 1 training supergroup 1625 2021-05-31 17:34 /loudacre/kb/KBD0
C-00295.html
-rw-rw-rw- 1 training supergroup 1270 2021-05-31 17:34 /loudacre/kb/KBD0
C-00296.html
-rw-rw-rw- 1 training supergroup 1167 2021-05-31 17:34 /loudacre/kb/KBD0
C-00297.html
-rw-rw-rw- 1 training supergroup 1015 2021-05-31 17:34 /loudacre/kb/KBD0
C-00298.html
-rw-rw-rw- 1 training supergroup 1126 2021-05-31 17:34 /loudacre/kb/KBD0
C-00299.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KBD0
C-00300.html
[training@localhost data]$ hdfs dfs -put $DEV1DATA/callogs /loudacre/
[training@localhost data]$ NET ID: dxd200016
```

Command: `hdfs dfs -rm -r /loudacre/callogs`

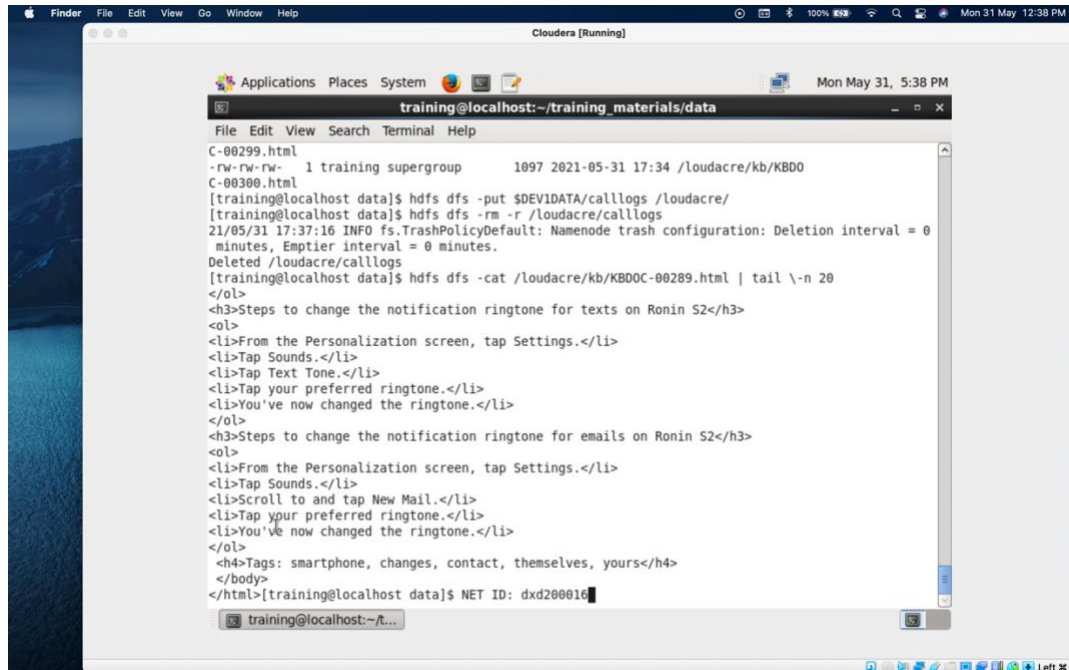


The screenshot shows the same Cloudera terminal window. The file list is now shorter, starting from C-00289.html. The command `hdfs dfs -rm -r /loudacre/callogs` is entered and executed. The output shows the command was successful, and the terminal prompt returns to `[training@localhost data]$`.

```
training@localhost:~/training_materials/data
File Edit View Search Terminal Help
C-00289.html
-rw-rw-rw- 1 training supergroup 1253 2021-05-31 17:34 /loudacre/kb/KBD0
C-00290.html
-rw-rw-rw- 1 training supergroup 1164 2021-05-31 17:34 /loudacre/kb/KBD0
C-00291.html
-rw-rw-rw- 1 training supergroup 1041 2021-05-31 17:34 /loudacre/kb/KBD0
C-00292.html
-rw-rw-rw- 1 training supergroup 1110 2021-05-31 17:34 /loudacre/kb/KBD0
C-00293.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KBD0
C-00294.html
-rw-rw-rw- 1 training supergroup 1625 2021-05-31 17:34 /loudacre/kb/KBD0
C-00295.html
-rw-rw-rw- 1 training supergroup 1270 2021-05-31 17:34 /loudacre/kb/KBD0
C-00296.html
-rw-rw-rw- 1 training supergroup 1167 2021-05-31 17:34 /loudacre/kb/KBD0
C-00297.html
-rw-rw-rw- 1 training supergroup 1015 2021-05-31 17:34 /loudacre/kb/KBD0
C-00298.html
-rw-rw-rw- 1 training supergroup 1126 2021-05-31 17:34 /loudacre/kb/KBD0
C-00299.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KBD0
C-00300.html
[training@localhost data]$ hdfs dfs -put $DEV1DATA/callogs /loudacre/
[training@localhost data]$ hdfs dfs -rm -r /loudacre/callogs
21/05/31 17:37:16 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0
minutes, Emptier interval = 0 minutes.
Deleted /loudacre/callogs
[training@localhost data]$ NET ID: dxd200016
```



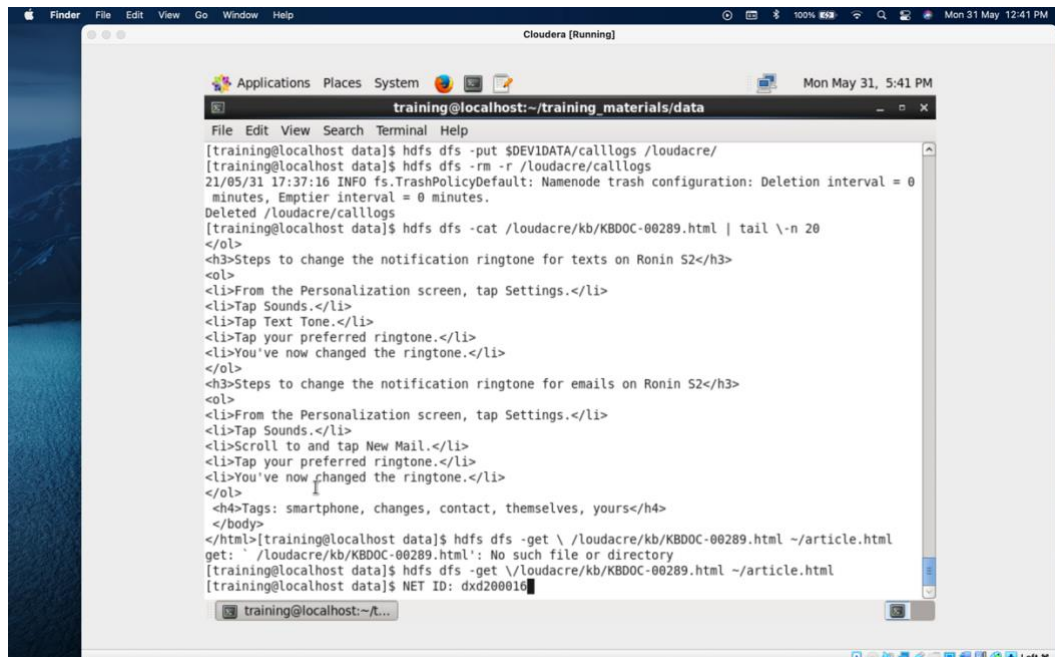
9. Now, I tried to view the contents of the file in the HDFS  
Command: `hdfs dfs -cat /loudacre/kb/KBDOC-00289.html | tail \-n 20`  
Below is screenshot of the contents of KBDOC-00289.html file.



The screenshot shows a terminal window titled "training@localhost:~/training\_materials/data". The terminal displays the output of the command `hdfs dfs -cat /loudacre/kb/KBDOC-00289.html | tail \-n 20`. The output is an HTML document snippet showing steps to change notification ringtones for texts and emails on a Ronin S2 smartphone. The document includes headings, lists, and tags. The terminal also shows the file's metadata and the user's command history.

```
training@localhost:~/training_materials/data
File Edit View Search Terminal Help
C-00299.html
-rw-rw-rw- 1 training supergroup 1097 2021-05-31 17:34 /loudacre/kb/KBDOC
C-00300.html
[training@localhost data]$ hdfs dfs -put $DEV1DATA/calllogs /loudacre/
[training@localhost data]$ hdfs dfs -rm -r /loudacre/calllogs
21/05/31 17:37:16 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0
minutes, Emptier interval = 0 minutes.
Deleted /loudacre/calllogs
[training@localhost data]$ hdfs dfs -cat /loudacre/kb/KBDOC-00289.html | tail \-n 20
</ol>
<h3>Steps to change the notification ringtone for texts on Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Tap Text Tone.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h3>Steps to change the notification ringtone for emails on Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Scroll to and tap New Mail.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h4>Tags: smartphone, changes, contact, themselves, yours</h4>
</body>
</html>[training@localhost data]$ NET ID: dxd200016
```

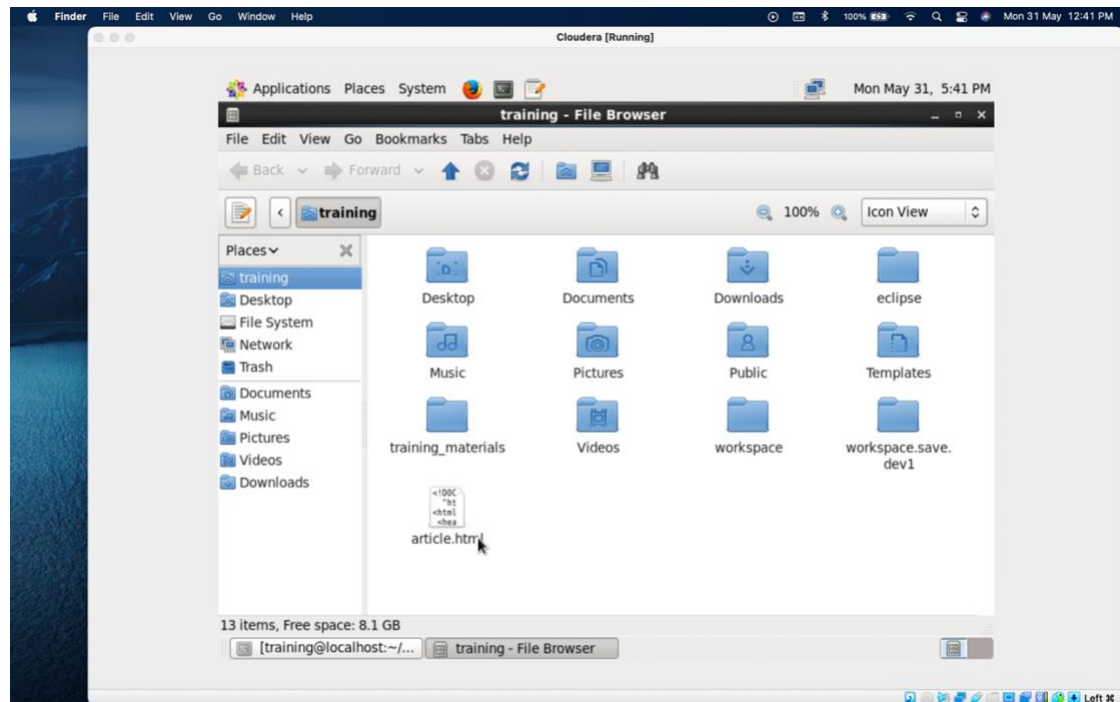
10. Now I copied the data from the hdfs file into a file called article.html in the local system.  
Command: `hdfs dfs -get \ /loudacre/kb/KBDOC-00289.html ~/article.html`



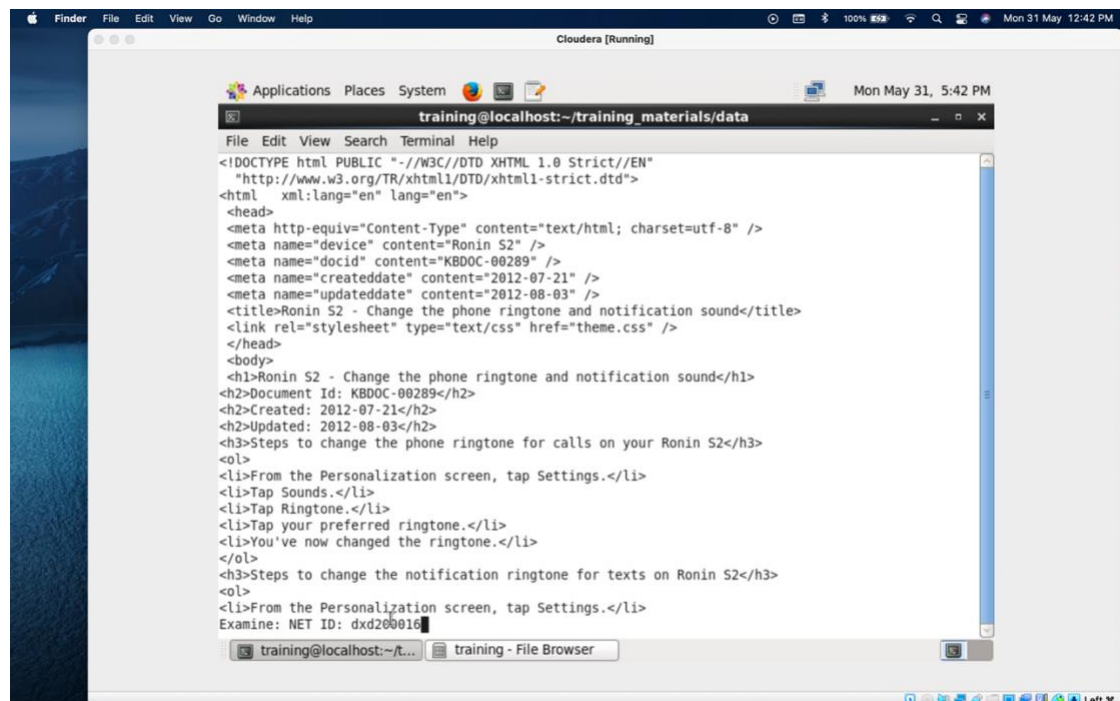
The screenshot shows a terminal window titled "training@localhost:~/training\_materials/data". The terminal displays the command `hdfs dfs -get \ /loudacre/kb/KBDOC-00289.html ~/article.html` and its output. The output shows the file being copied from HDFS to the local system. The terminal also shows the file's metadata and the user's command history.

```
training@localhost:~/training_materials/data
File Edit View Search Terminal Help
[training@localhost data]$ hdfs dfs -put $DEV1DATA/calllogs /loudacre/
[training@localhost data]$ hdfs dfs -rm -r /loudacre/calllogs
21/05/31 17:37:16 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0
minutes, Emptier interval = 0 minutes.
Deleted /loudacre/calllogs
[training@localhost data]$ hdfs dfs -cat /loudacre/kb/KBDOC-00289.html | tail \-n 20
</ol>
<h3>Steps to change the notification ringtone for texts on Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Tap Text Tone.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h3>Steps to change the notification ringtone for emails on Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Scroll to and tap New Mail.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h4>Tags: smartphone, changes, contact, themselves, yours</h4>
</body>
</html>[training@localhost data]$ hdfs dfs -get \ /loudacre/kb/KBDOC-00289.html ~/article.html
get: ' /loudacre/kb/KBDOC-00289.html': No such file or directory
[training@localhost data]$ hdfs dfs -get \loudacre/kb/KBDOC-00289.html ~/article.html
[training@localhost data]$ NET ID: dxd200016
```

The file was successfully stored in the local system.

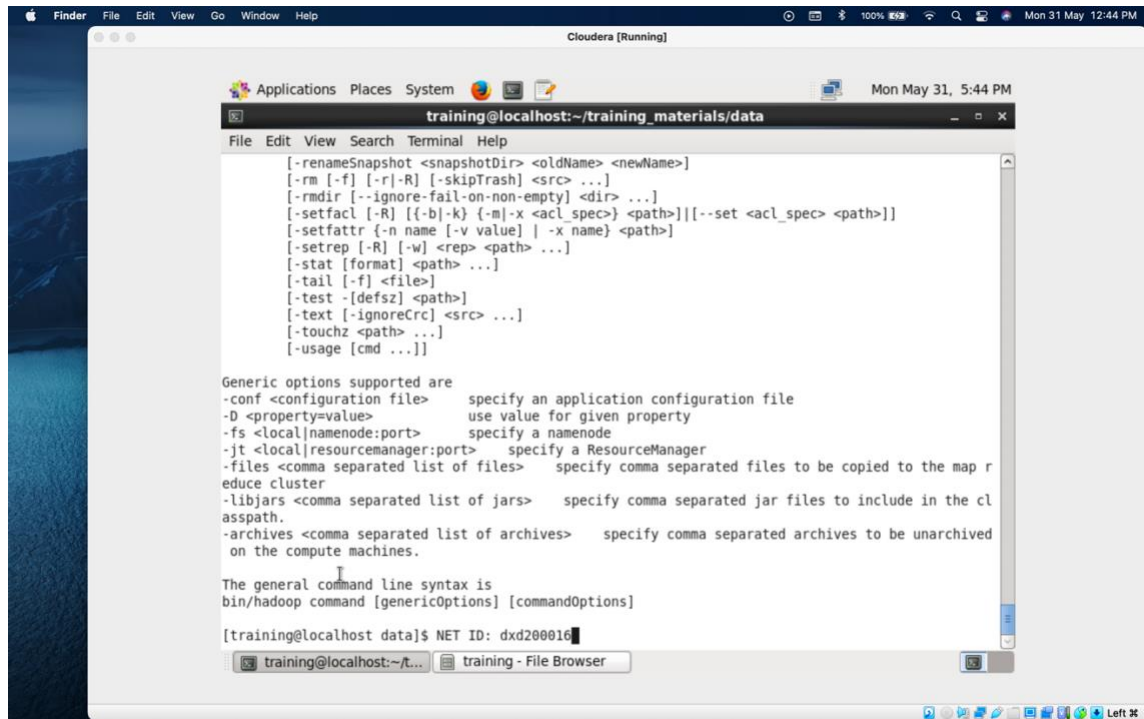


Then I open the file (article.html) stored locally using the command line.  
Command: less ~/article.html

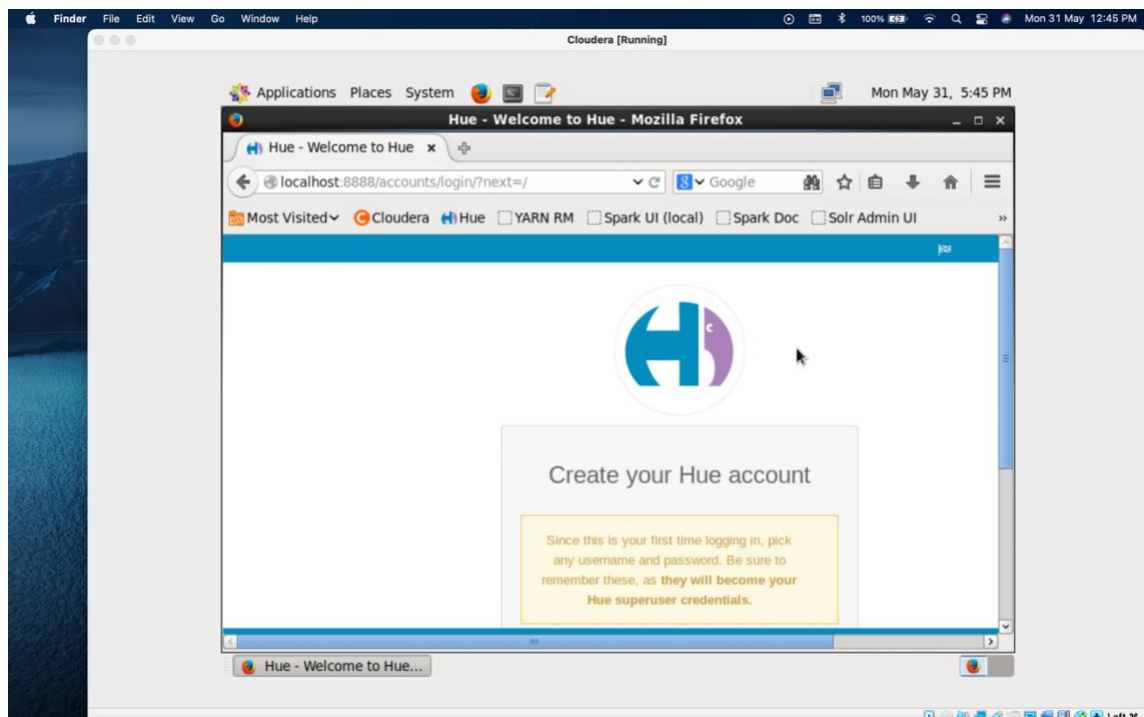




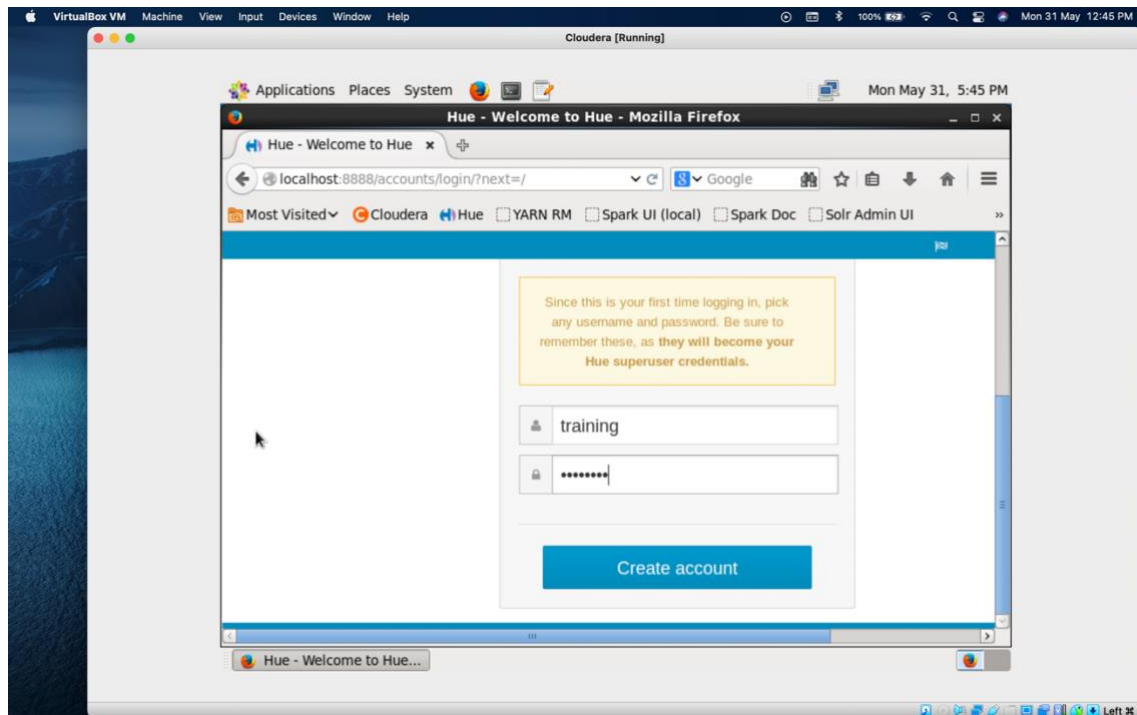
11. Finally, I ran a command to view all important commands while working with HDFS on command line.  
Command: `hdfs dfs`



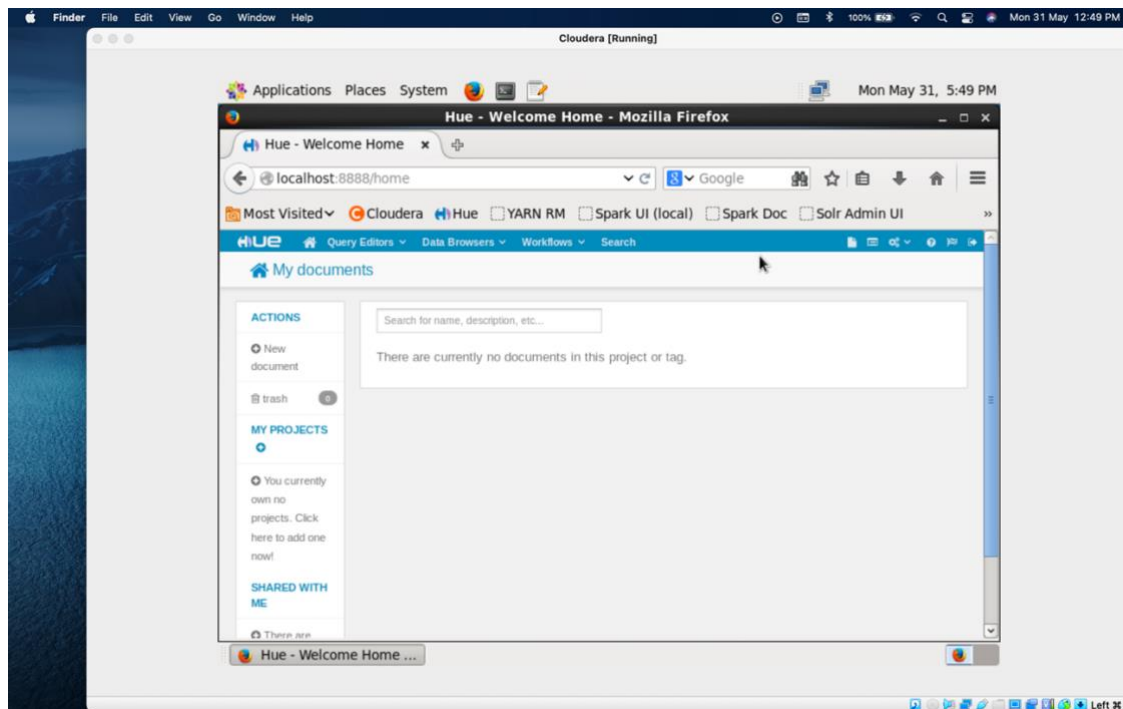
12. Now, I went to the Firefox browser to view open the HUE file browser interface.  
Site: <https://localhost:8888>



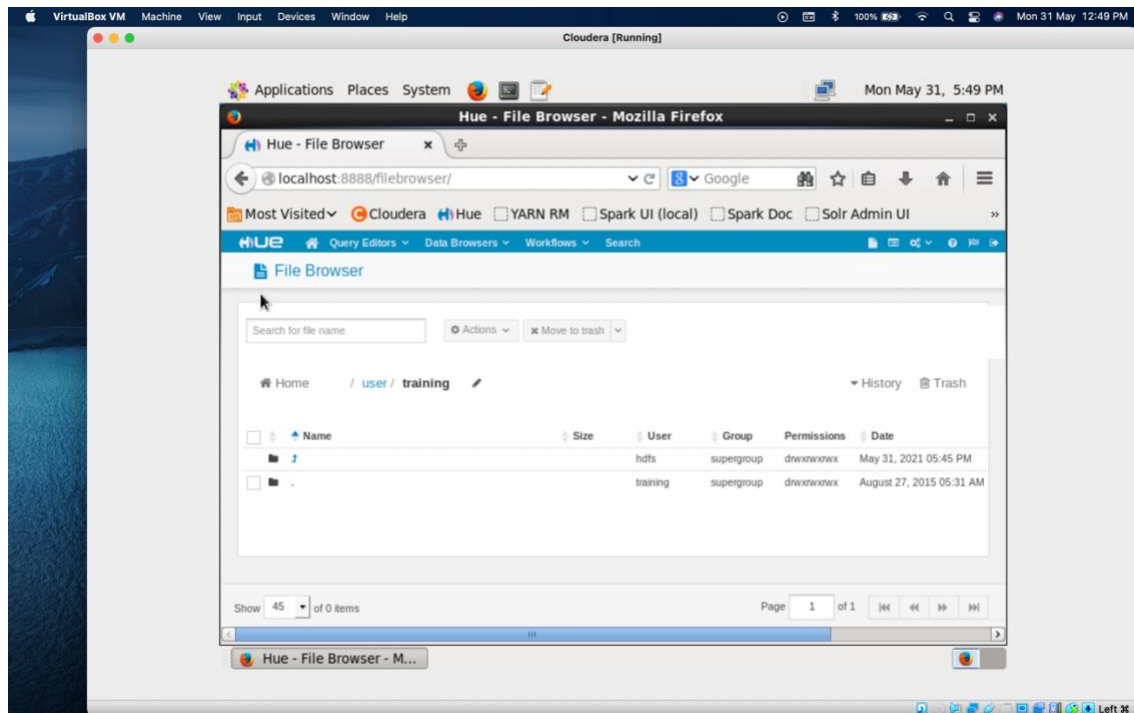
13. Now to login, I created an account first with username training.



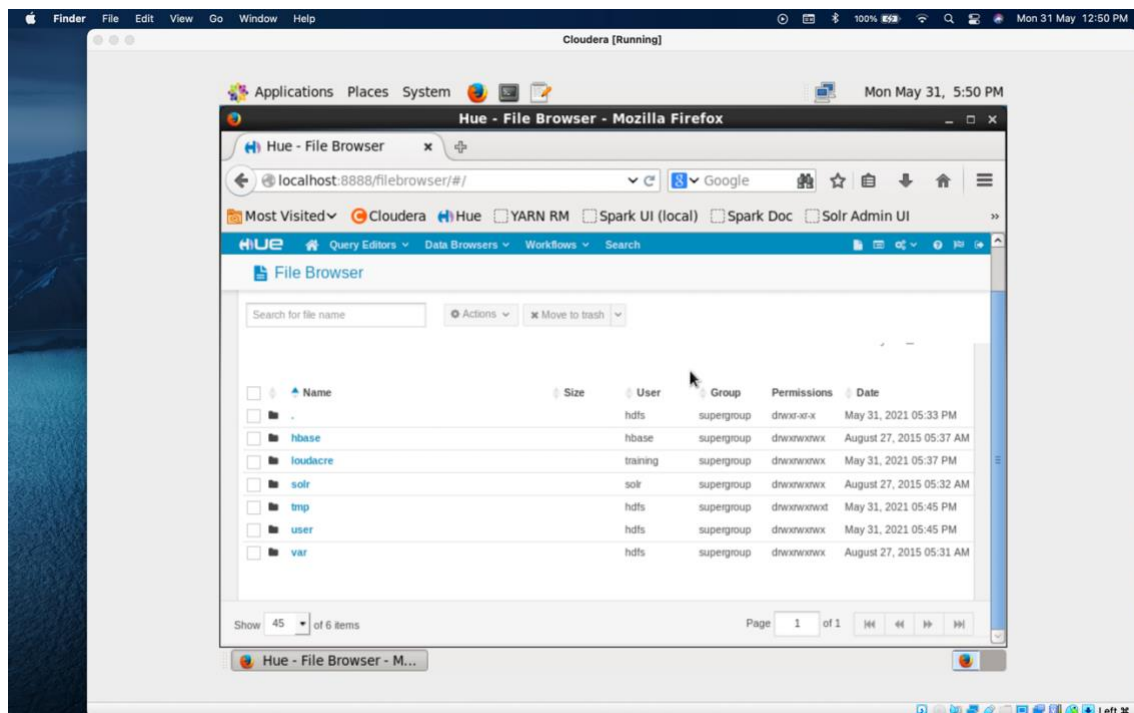
14. After logging in I was able to see the HUE dashboard.



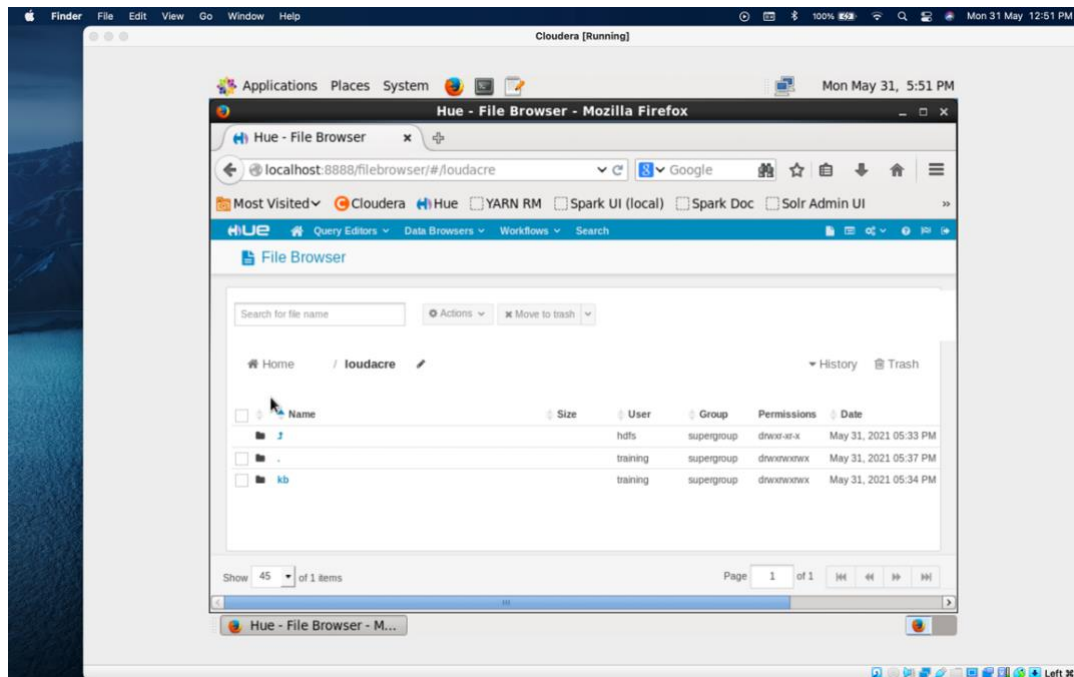
15. Then I clicked on the file browser on the menu on the top of dashboard to open the file browser.



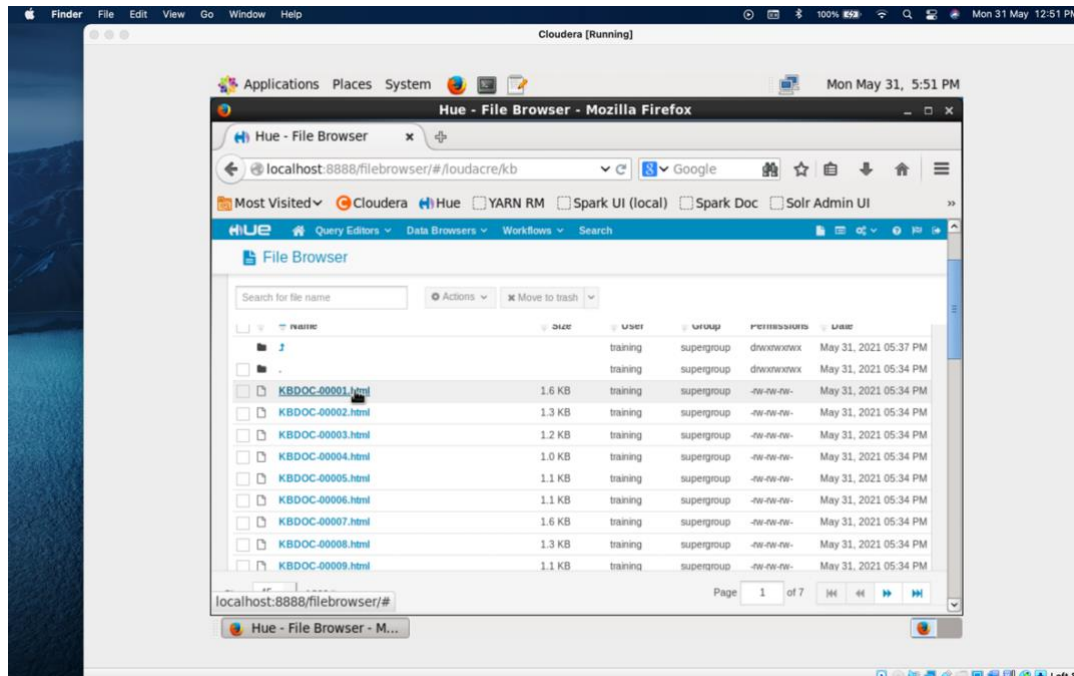
16. After that opened the HDFS root directory to view the contents inside.



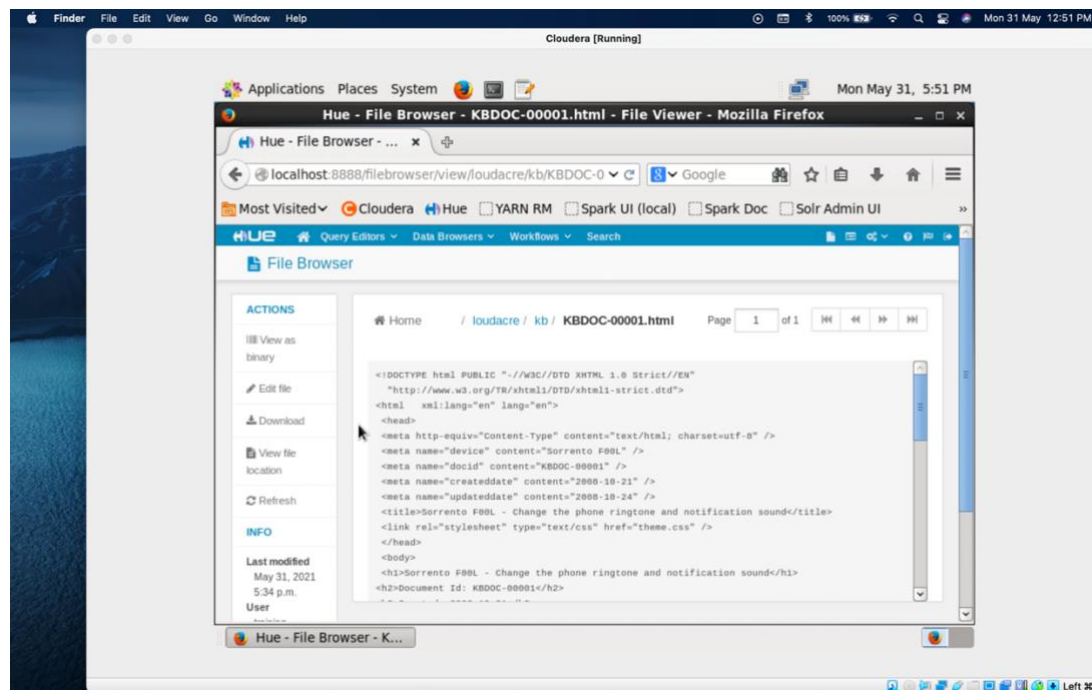
17. Then I click on the loudacre folder to view the contents.



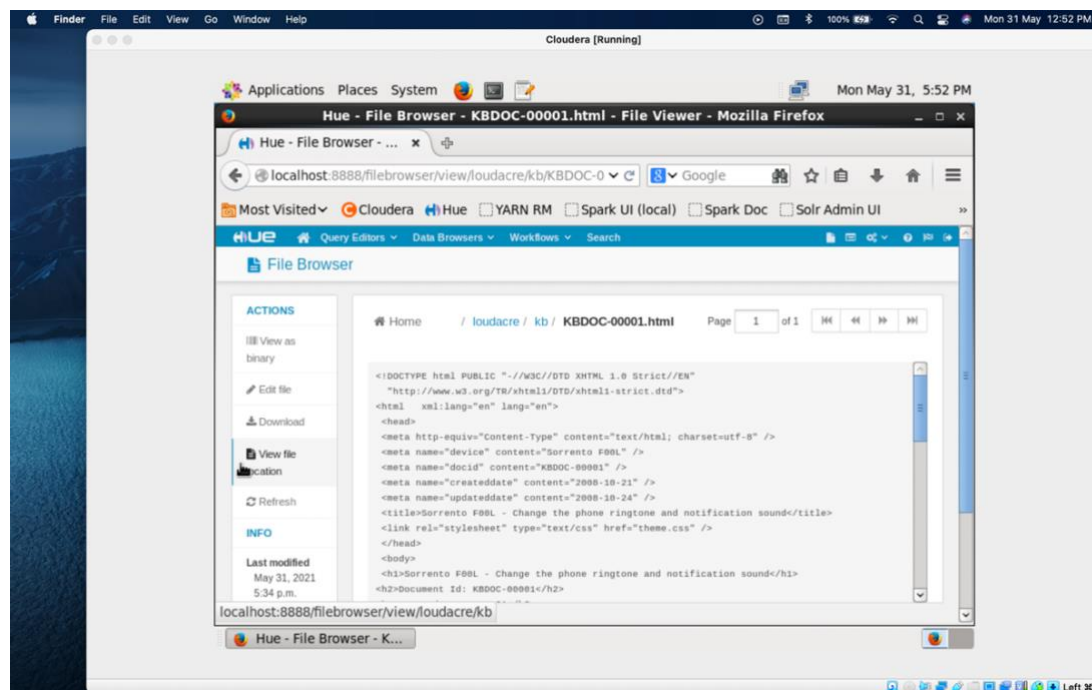
18. After that I opened the kb folder to view the files uploaded in the same.



19. After that I click on the KBDOC-0001.html to view the contents of the file.  
I was able to view all the contents inside the file.

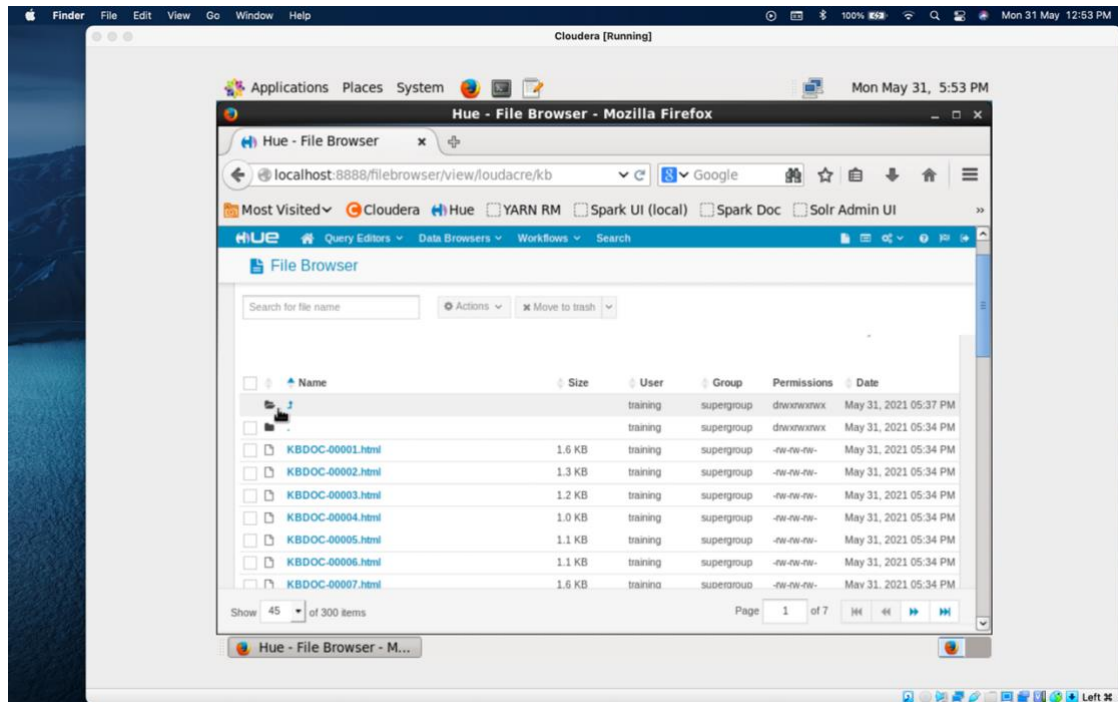


20. After that to go back to the file location I clicked on the View File Location.

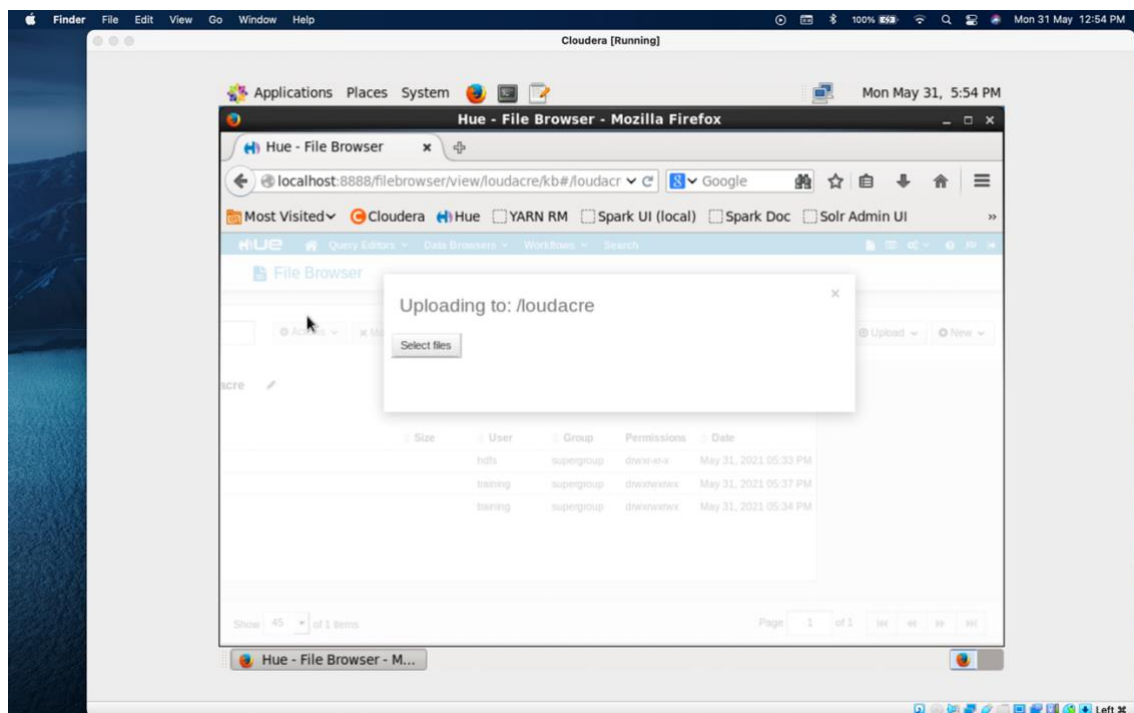




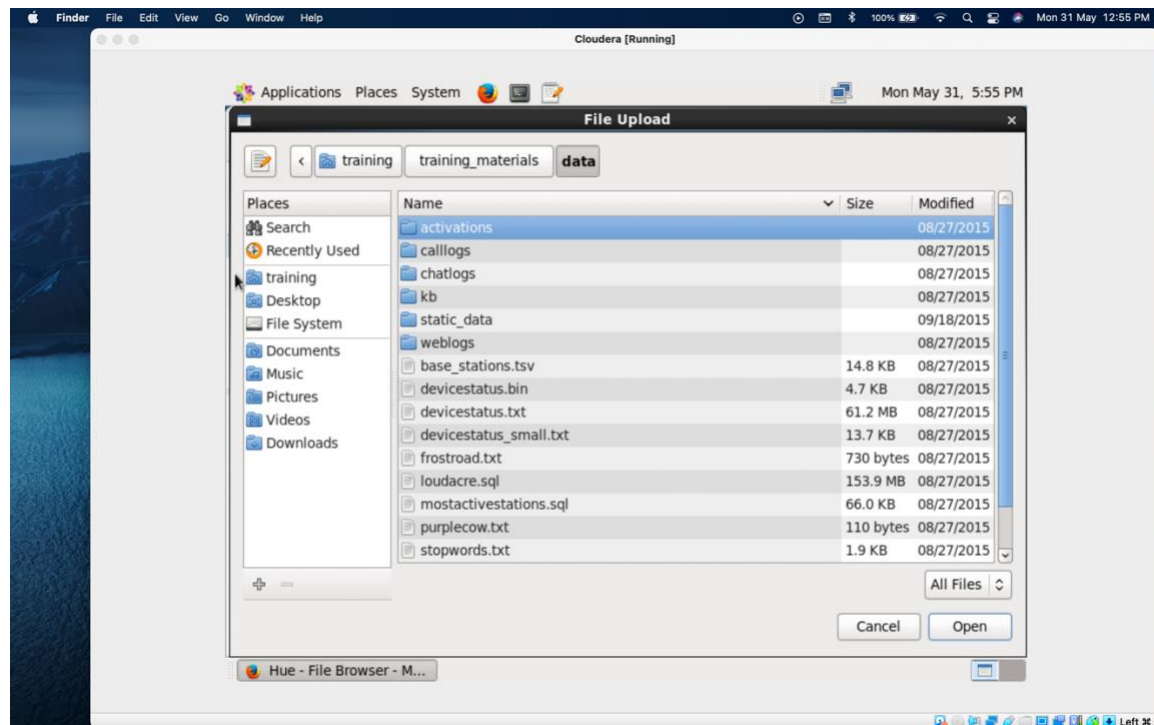
21. Then I click on the return arrow to go back to the base directory



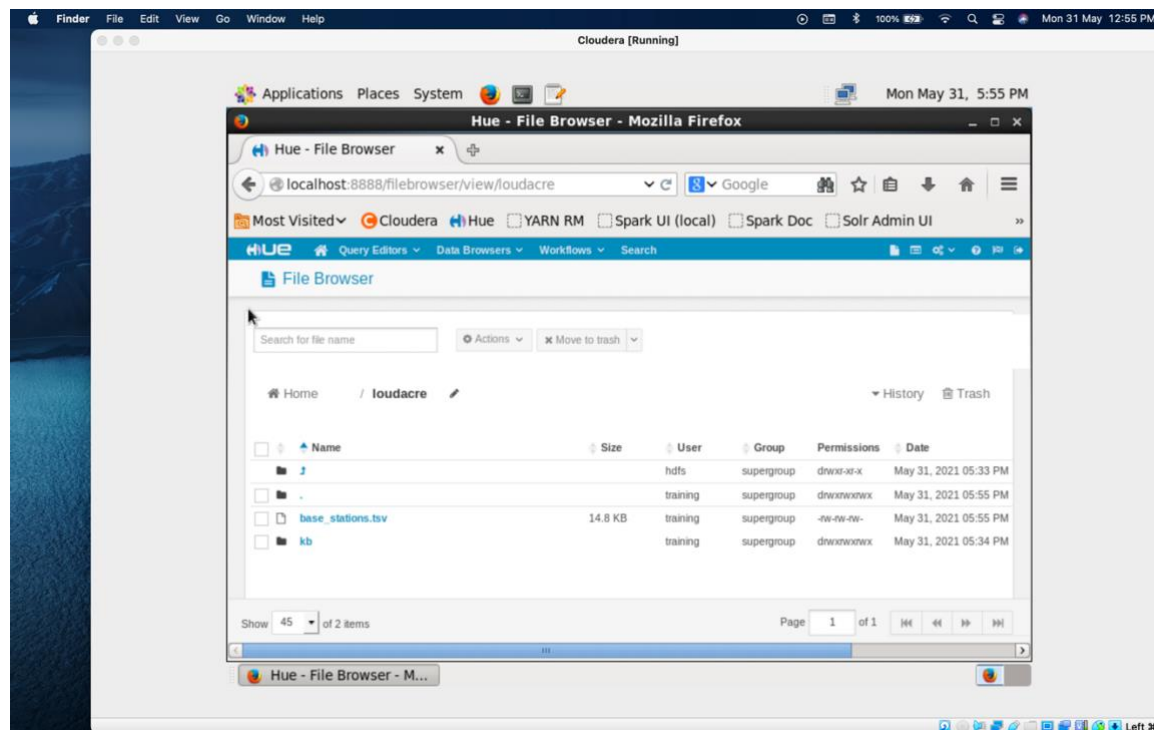
22. After that, to upload more files in the loudacre folder I clicked on the upload button.



23. Then I selected the file called base\_stations,tsv and click on the Open button.



24. The file was successfully uploaded and can be seen in the below screenshot.



25. Then I selected the file, and clicked on the actions button, to view all the action I can perform on the selected file.

There were mainly 5 actions that I was able to preform, as seen in the below screenshot.

