YARN Job

Divyansh Dahiya

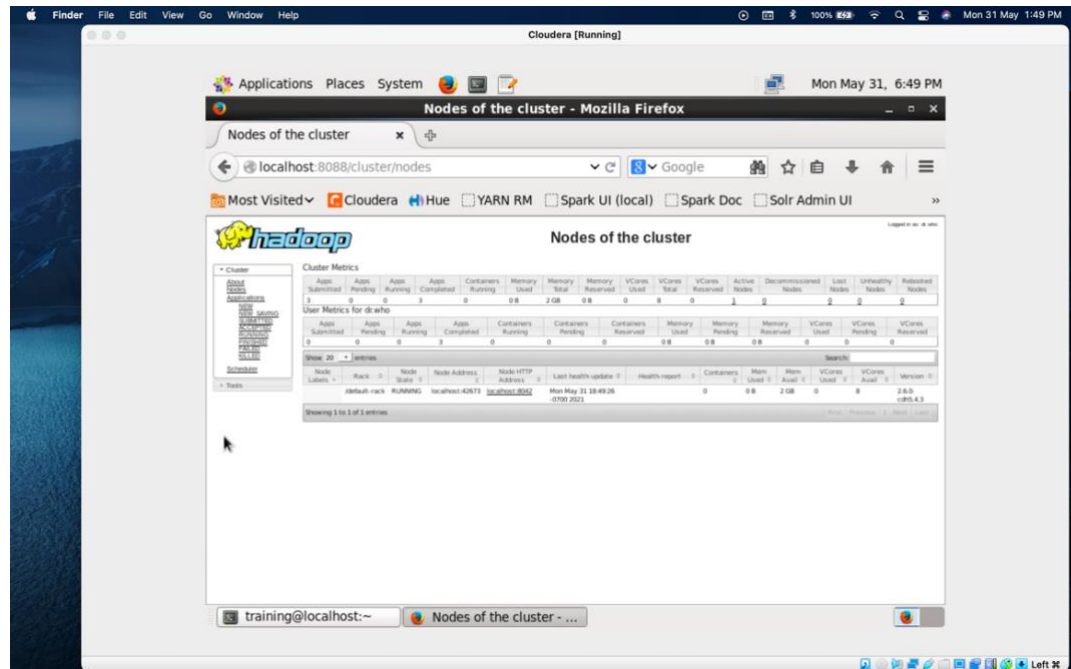LAB Chapter 04: YARN JOB

Run a YARN Job

In this lab, I did a hands-on with YARN cluster, to monitor the application using HUE File Browser and YARN web UI.

1. For viewing the YARN web UI I opened the Firefox browser and open the following site: https://localhost:8088
   Below is the screenshot of the same. There are some jobs which can be seen in the dashboard, that is because I ran the sqoop lab before doing the yarn lab.
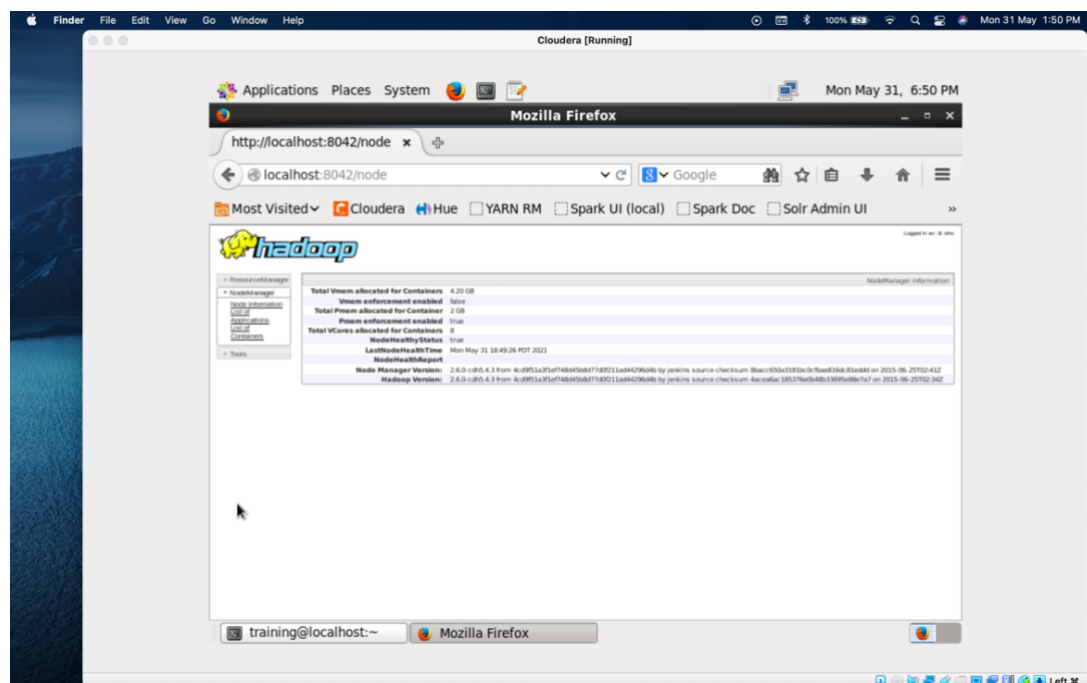
2. Then I clicked on the nodes on the left side of the screen, to view the nodes of the cluster.
   Below is the screenshot of the same.

3. Then I clicked on the Node HTTP address to open the Node Manager UI for that node. I was able to view the statistics, like amount of available memory and currently running applications.
   Below is the screenshot for the same.

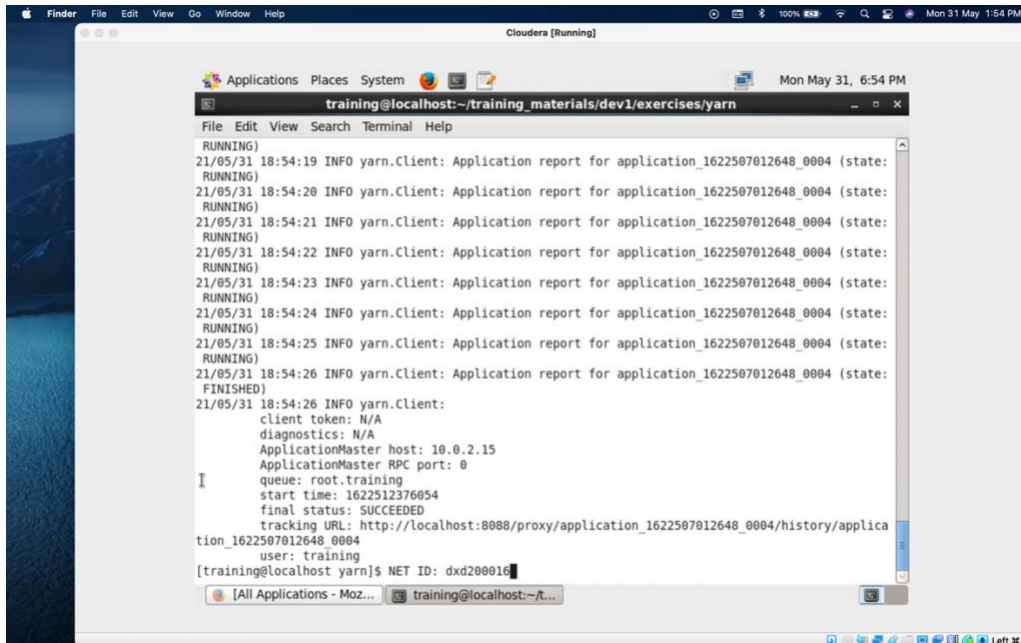4. Then I click on the RM HOME on the left hand of the screen to go back to the Resource Manager.
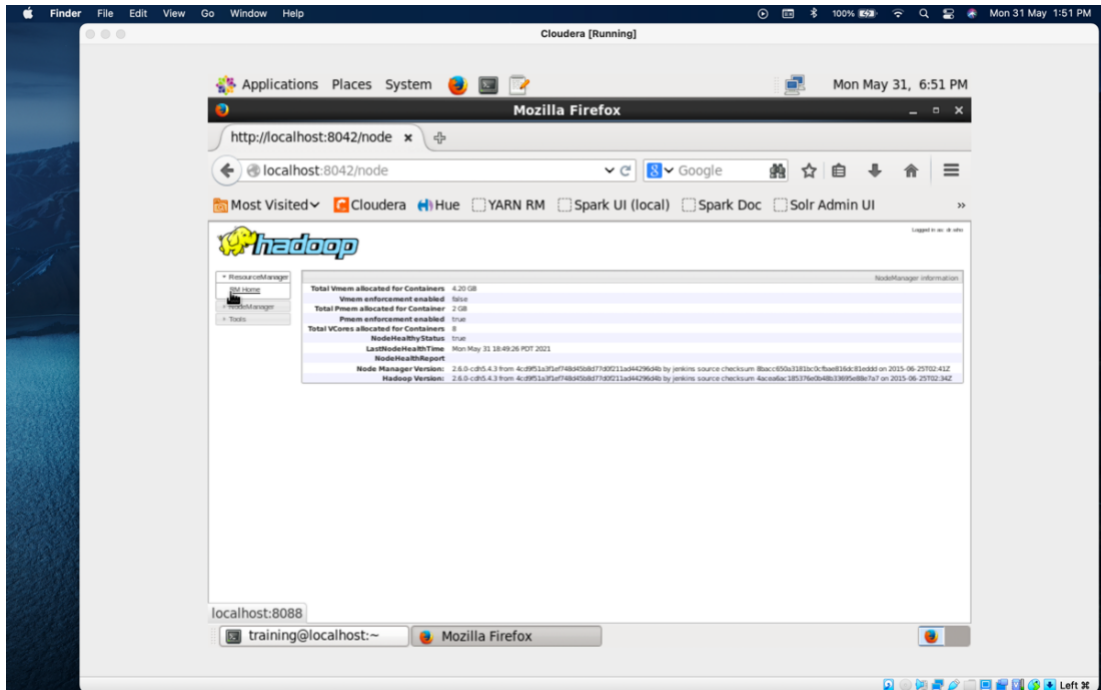
5. Then, for submitting the application to the YARN cluster I open the command line and change the directory to: cd $DEV1/exercises/yarn
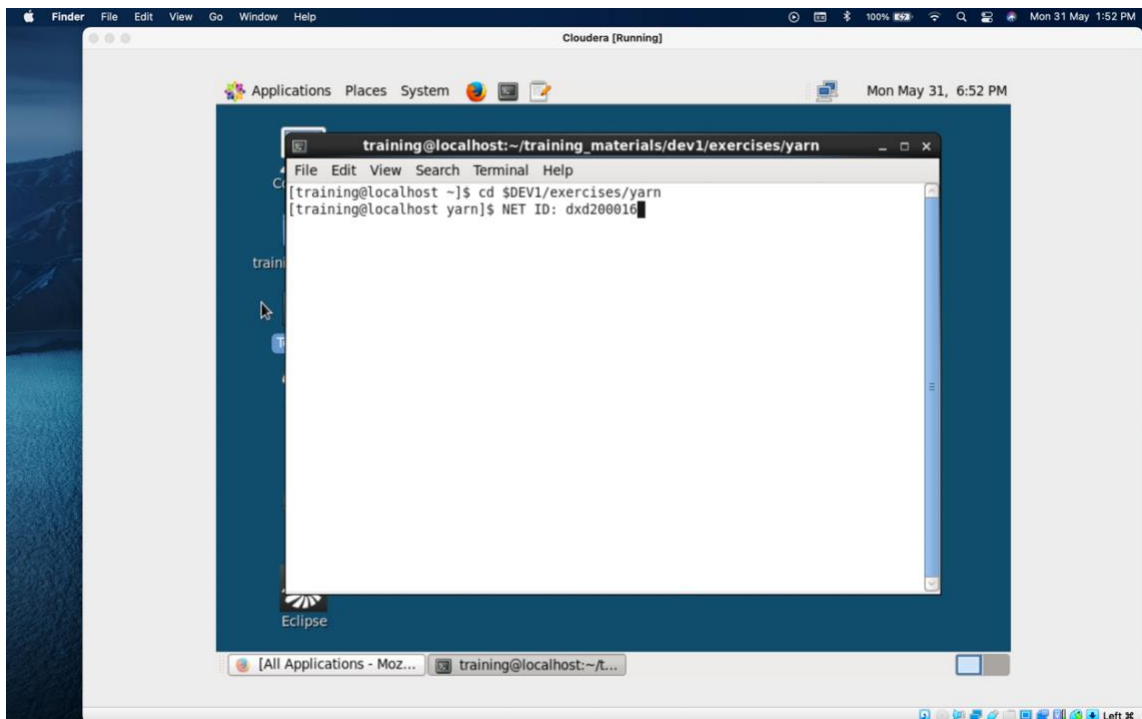
6. Then I ran the wordcount.py program on the YARN cluster in order to count the frequency of the words in the knowledge dataset.
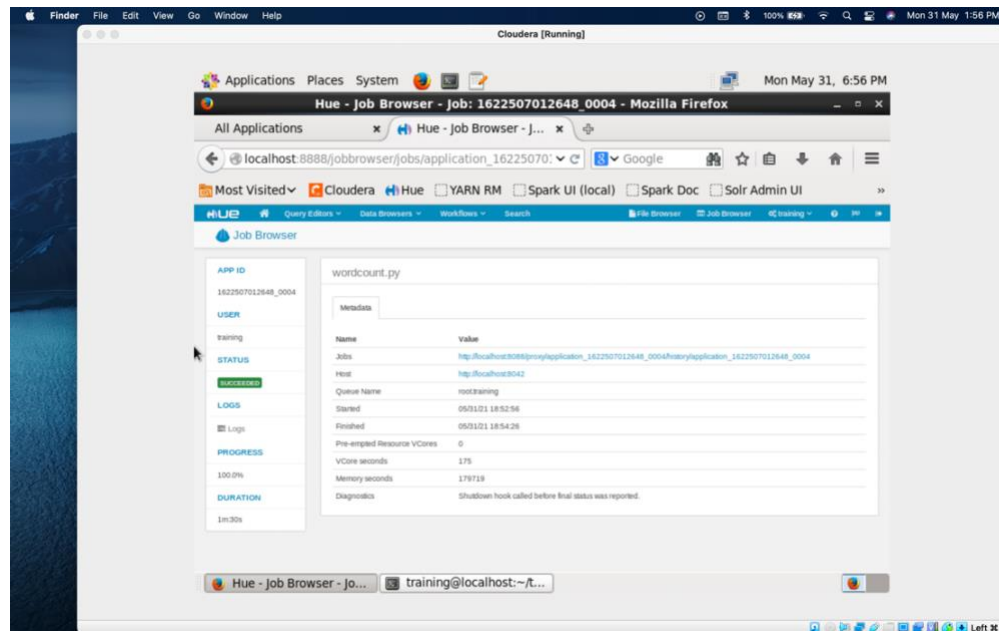Command: spark-submit --master yarn-cluster \ wordcount.py /loudacre/kb/*

7. Then I open the HUE File Browser and click on the Job Browser to view all the jobs.
I was able to all the jobs that were completed successfully and shows the status as Succeeded.
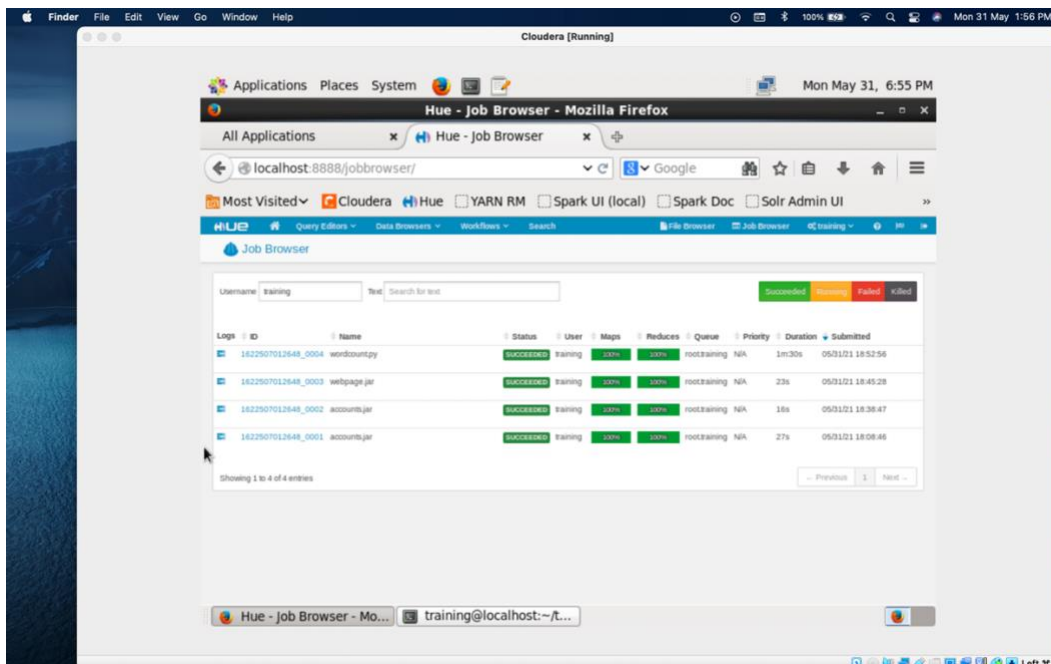
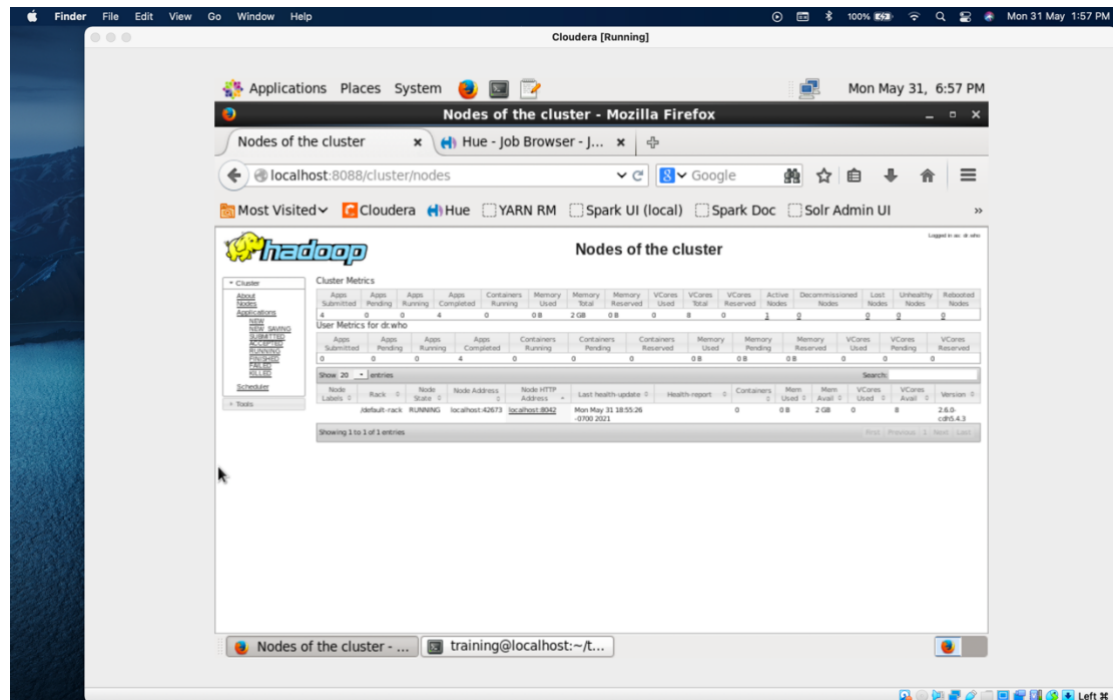8. Then I click on the application ID to view the details of the application, such as Started, Finished, Queue Name, etc.
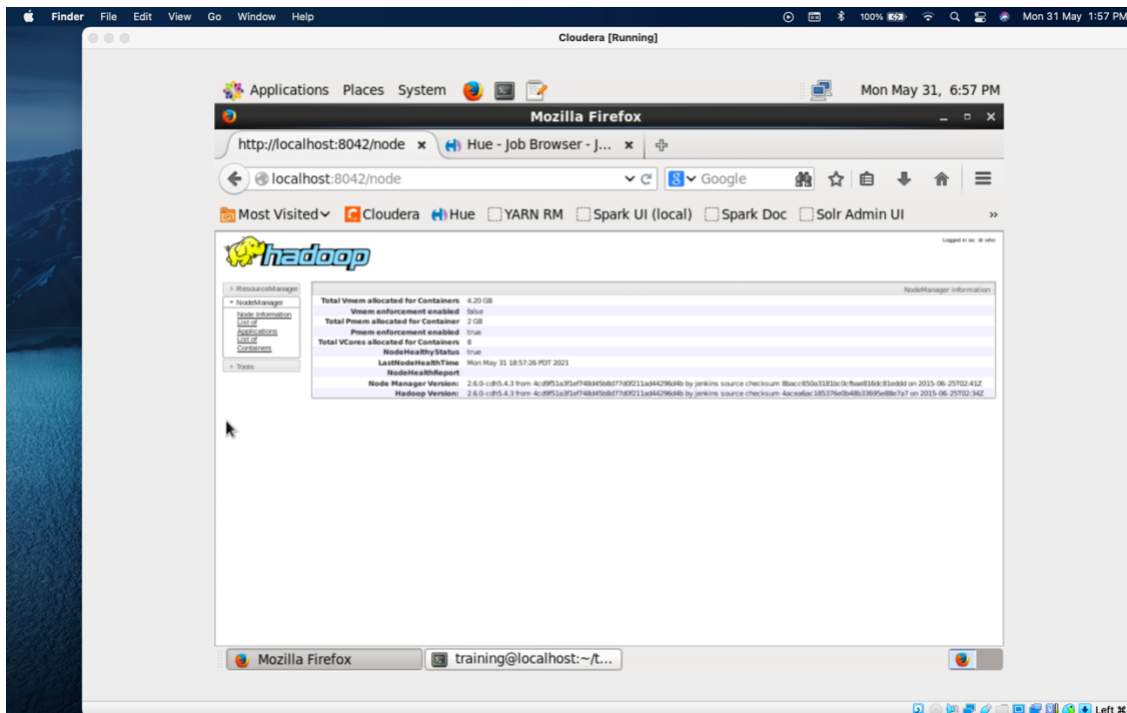

9. To get a more detailed view, I again open the YARN UI and I was able to see all the applications I ran or started before that time.
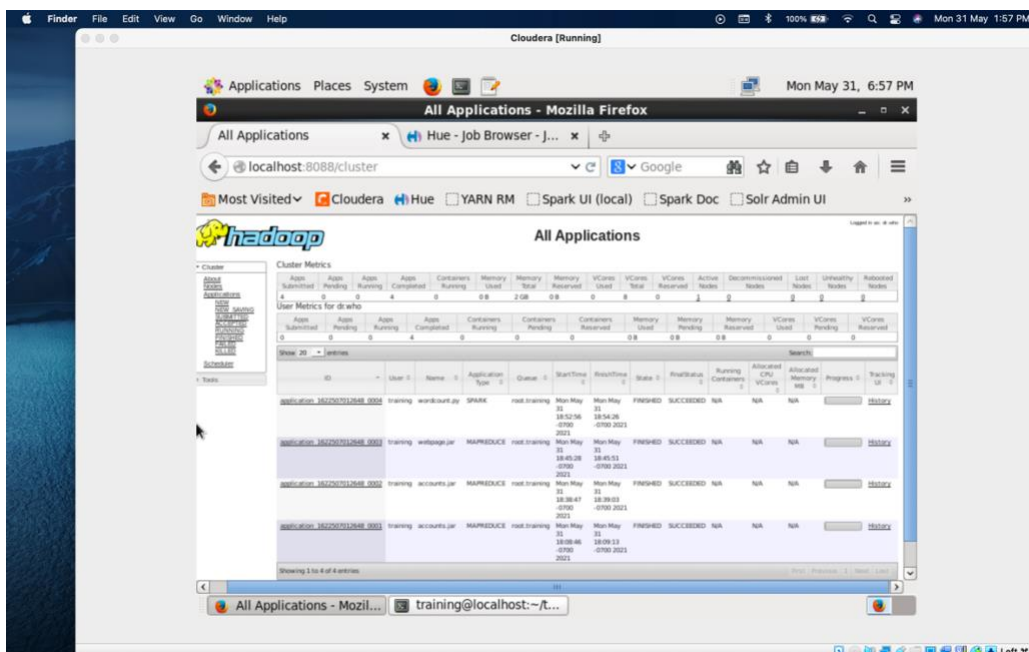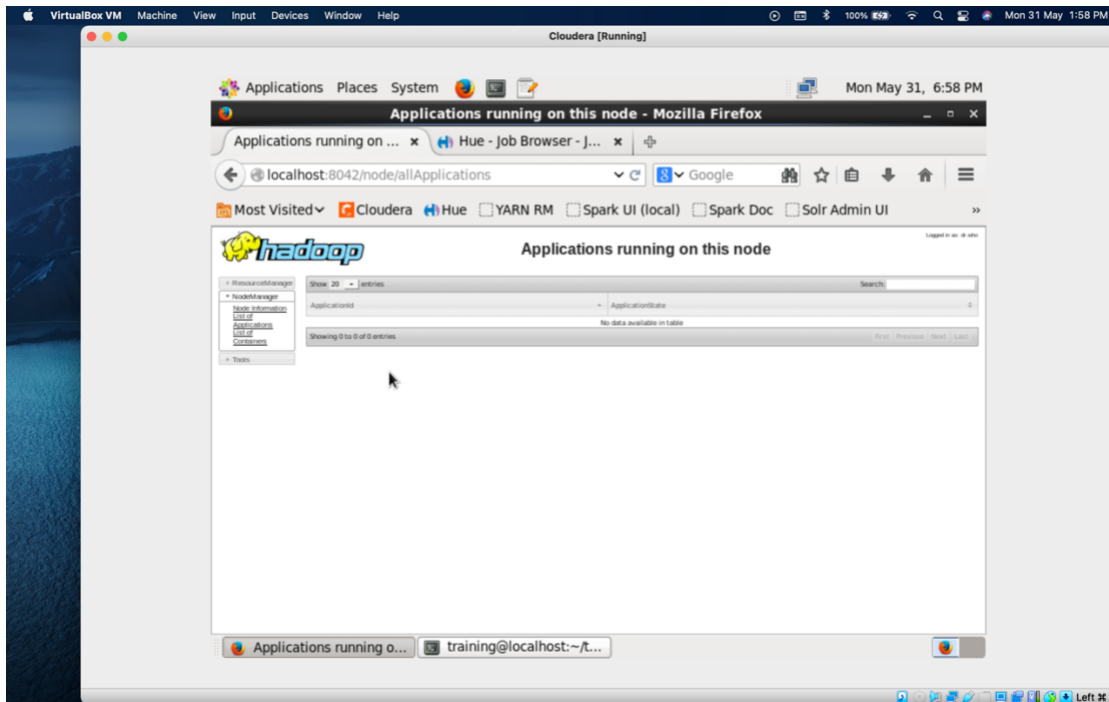
10. After that I again clicked on the nodes on the left side to view the nodes of the cluster.

11. Then I click on the Node HTTP Address to see the statistics.

12. On the left side of the screen, I click on List of Applications. As there were no active running application at the current moment, so the list was empty.

13. Again, I clicked on the list of containers to check if there are any active containers running on the node.