

FLUME

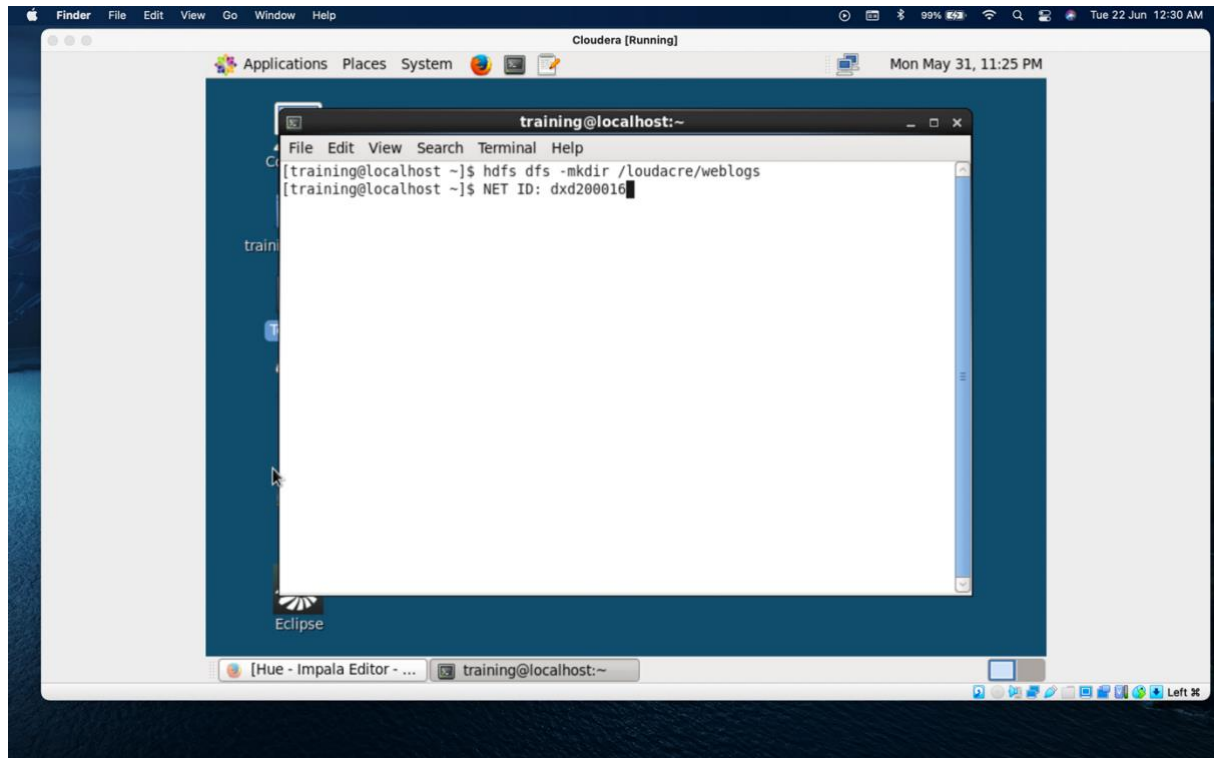
Divyansh Dahiya

LAB Chapter 09: FLUME

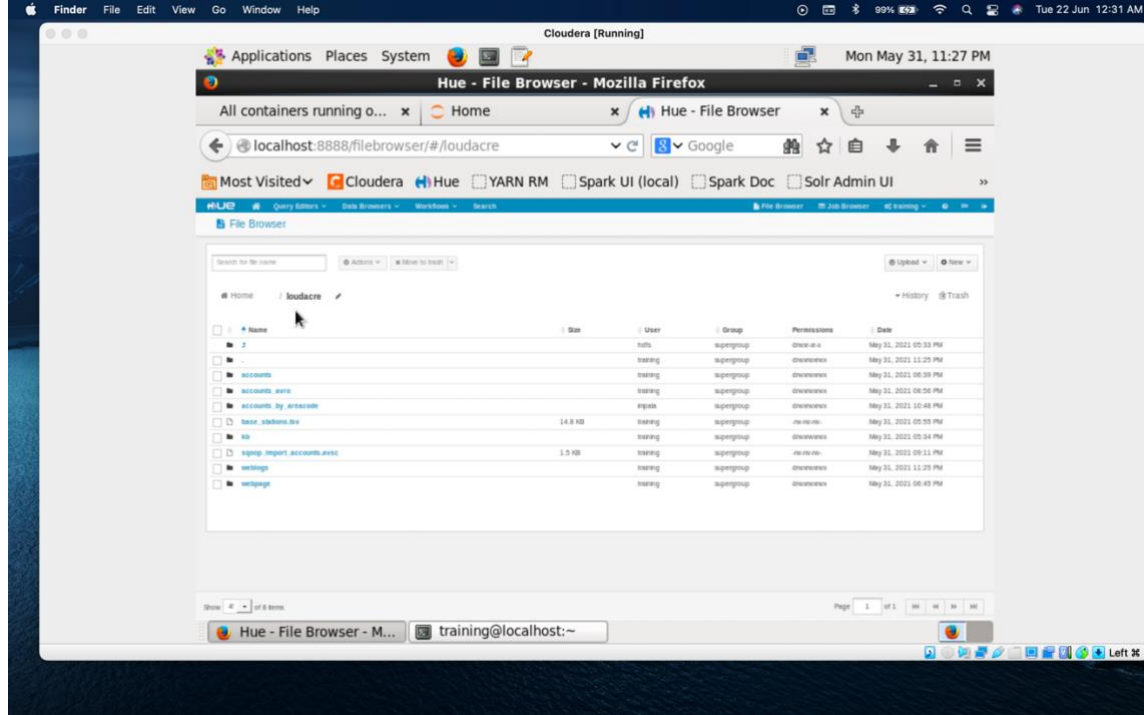
Collect Web Server Logs with Flume

In this lab, I will configure Flume to ingest web log data from a local directory to HDFS.

1. First, I ran the following command to create a directory in the HDFS to hold the data files ingested by Flume.
Command: `hdfs dfs -mkdir /loudacre/weblogs`

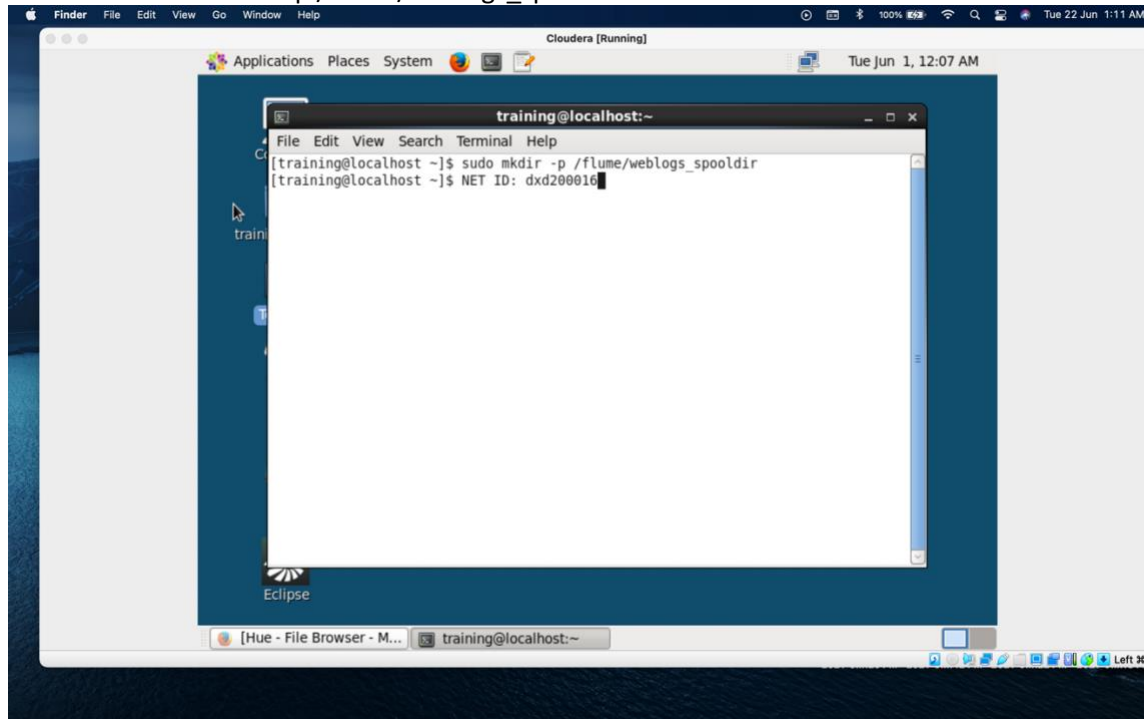


2. Below is the screenshot, confirming weblogs directory is created inside the HDFS.

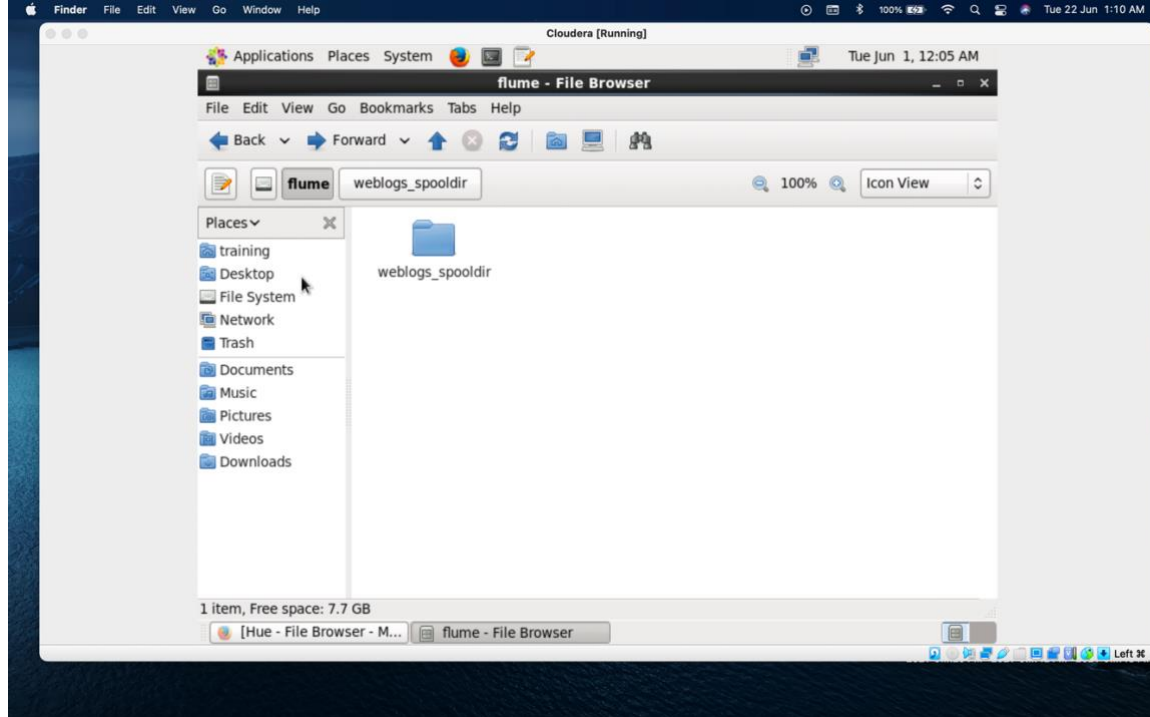


3. Then, I ran the following command to create the spool directory into which the web log simulator will store data files for Flume to ingest.

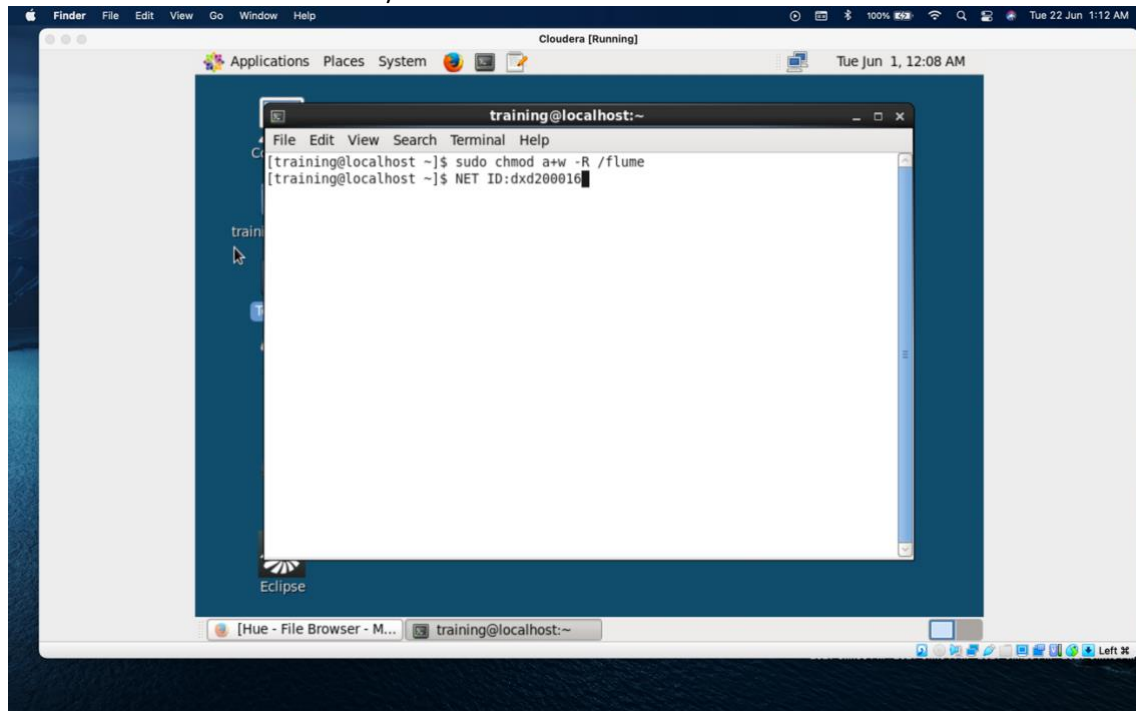
Command: `sudo mkdir -p /flume/weblogs_spool`



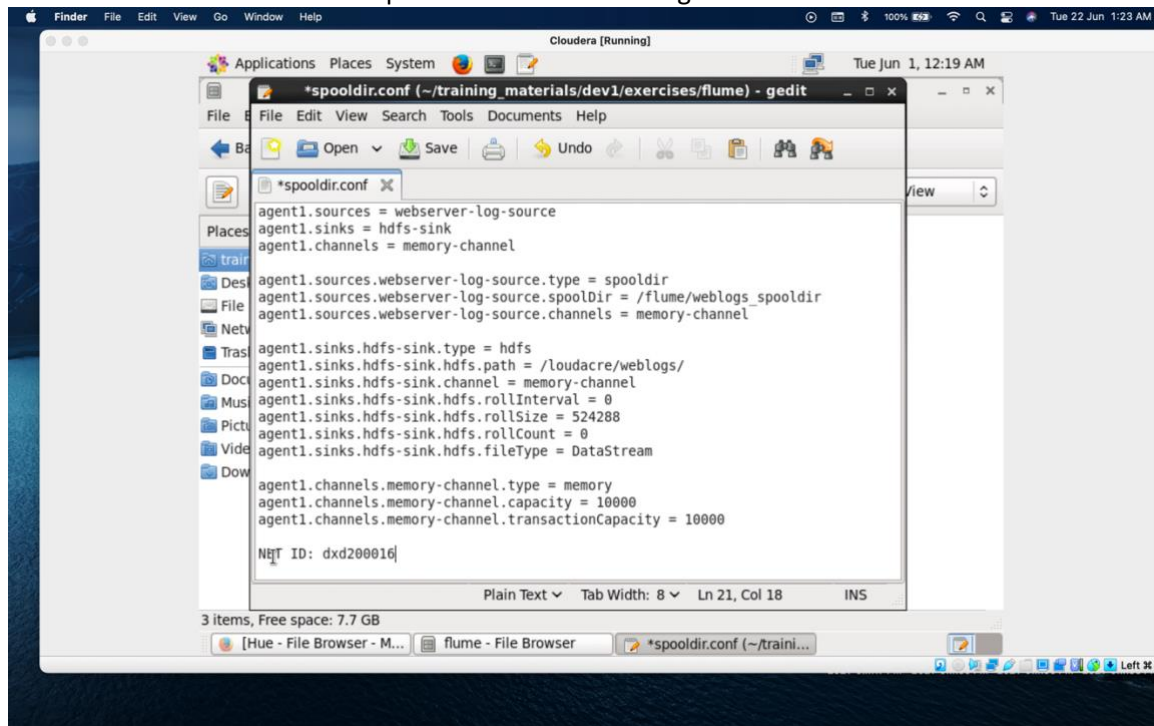
4. Below is the screenshot of the weblogs_spooldir directory created in the previous step.



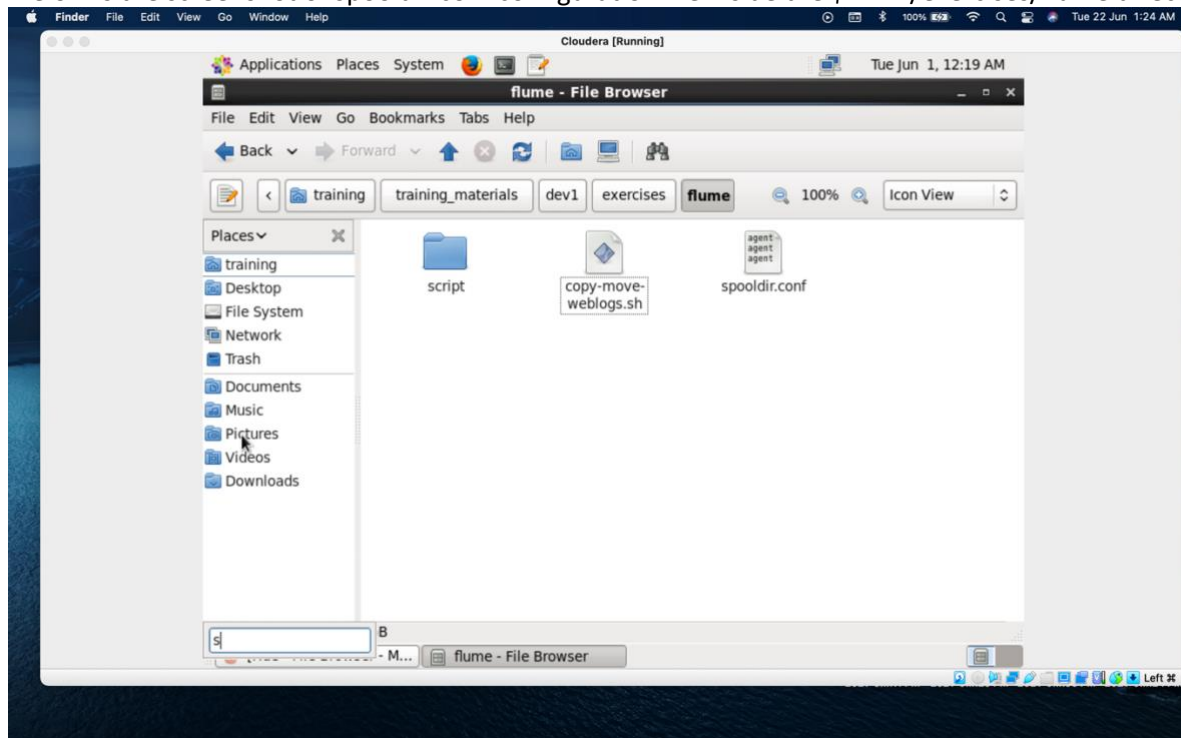
5. Then, I ran the following command to give permission to all users to write to the /flume/weblogs_spooldir directory.
Command: `sudo chmod a+w -R /flume`



6. Then, I created Flume configuration file spooldir.conf inside the \$DEV1/exercises/flume directory. Flume agent is configured through java properties file and uses hierarchical references. Below is the screenshot of the spooldir.conf Flume configuration file.

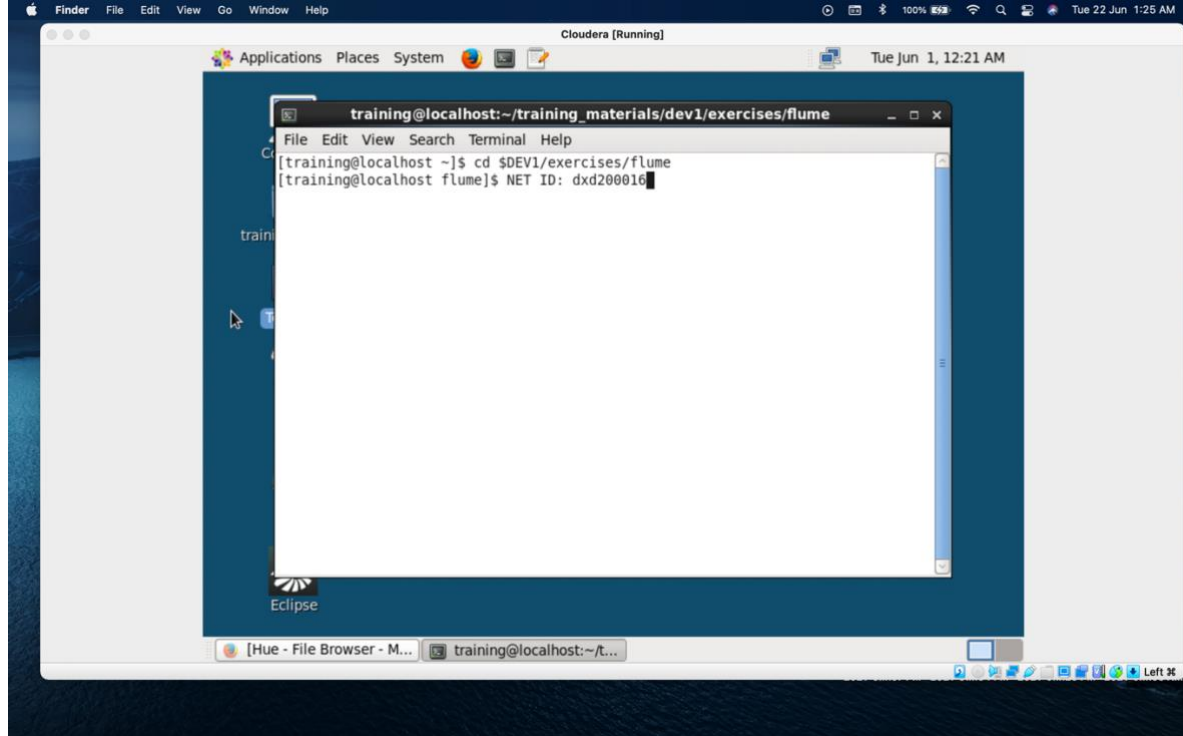


7. Below is the screenshot of spooldir.conf configuration file inside the \$DEV1/exercises/flume directory.



8. Then, I opened the terminal and ran the following command to change the directory.

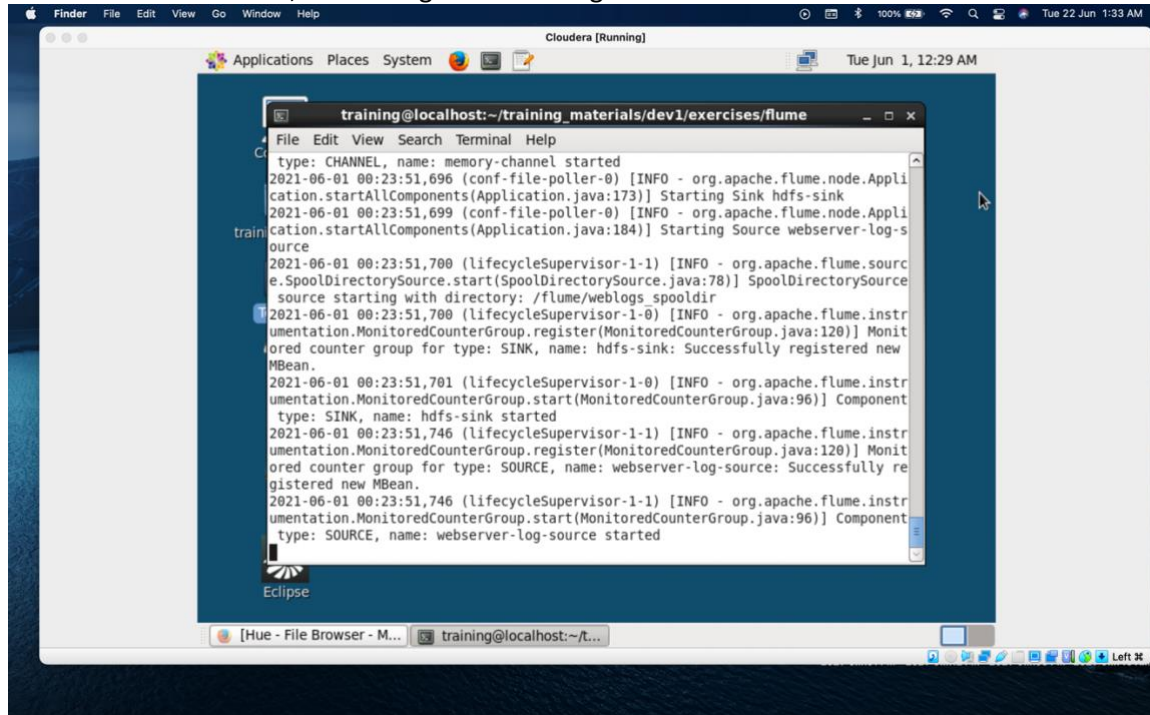
Command: `cd $DEV1/exercises/flume`



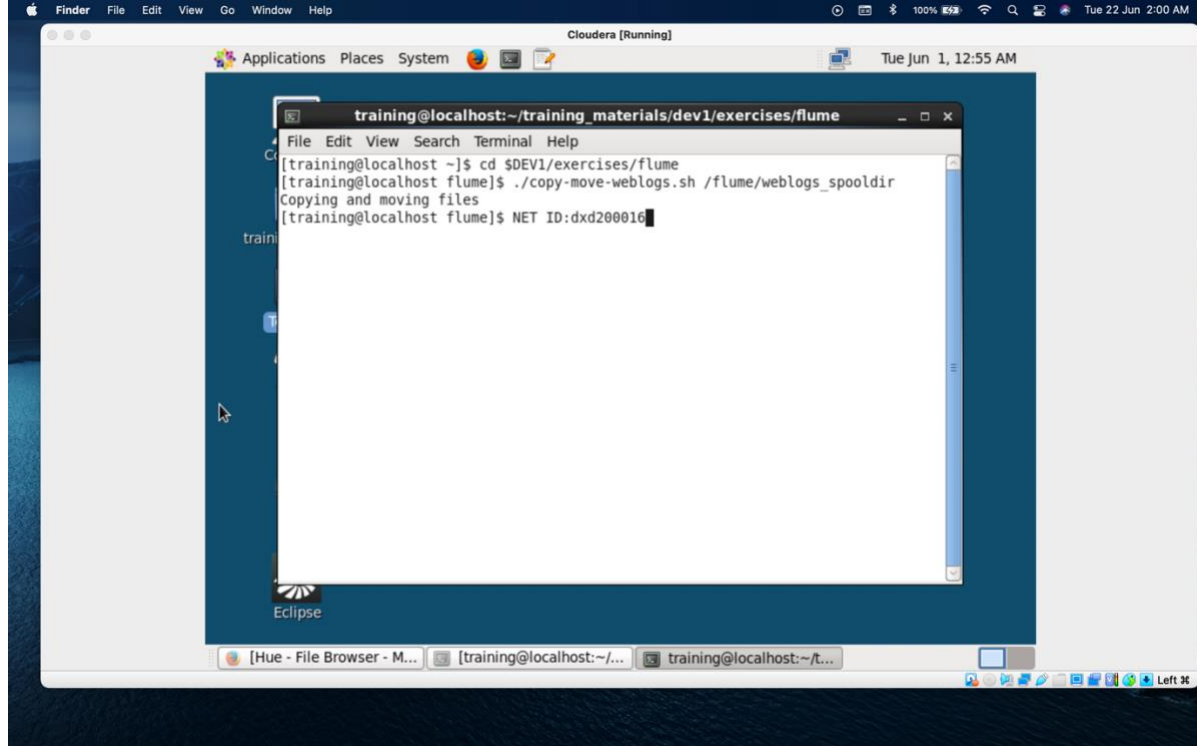
9. Then, I ran the following command and started the Flume agent using the configuration file.

Command: `flume-ng agent --conf /etc/flume-ng/conf \ --conf-file solution/spooldir.conf \ --name agent1 - Dflume.root.logger=INFO,console`

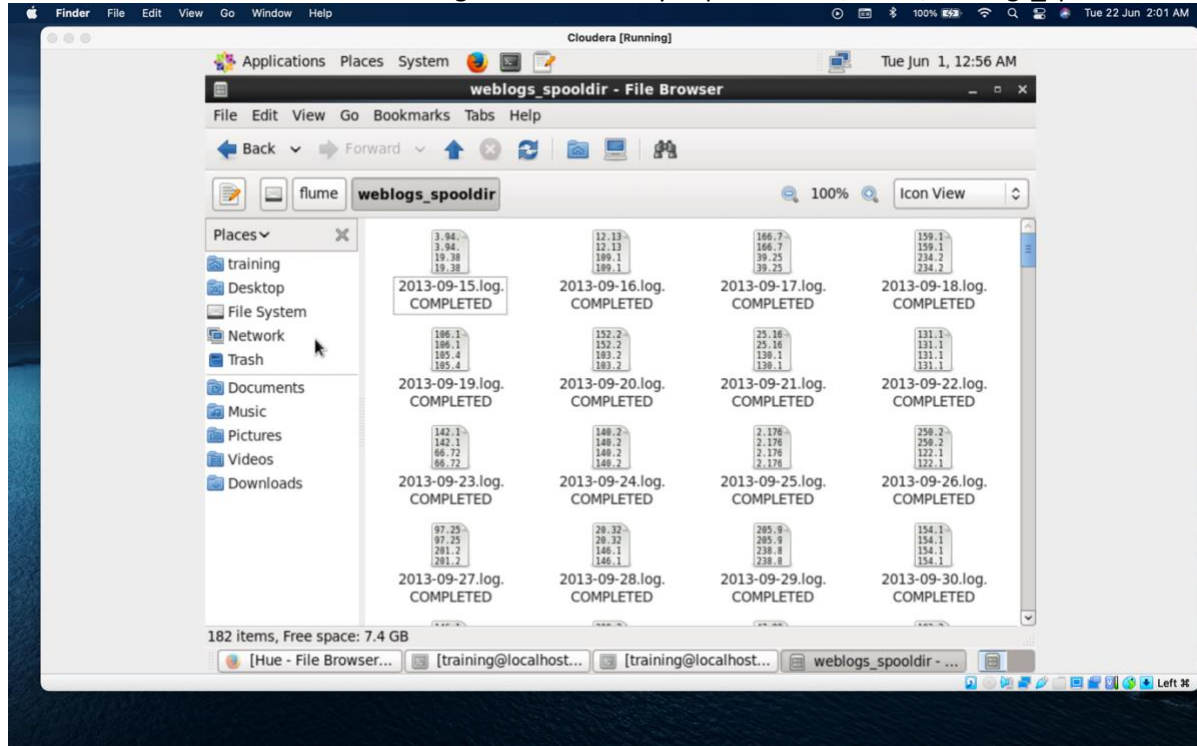
Below is the screenshot, confirming that Flume agent is started.



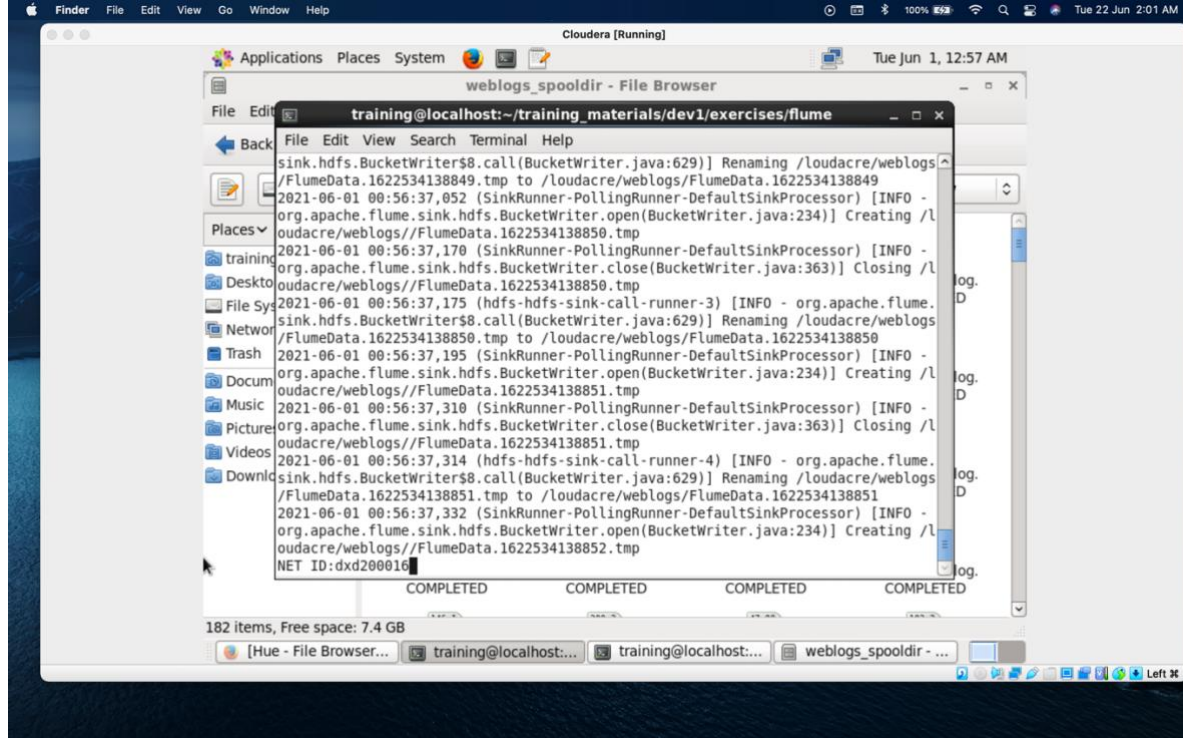
10. Then, I ran the following commands, to change the directory and ran the copy-move-weblogs.sh script to copy the web log files to the weblogs_spooldir.



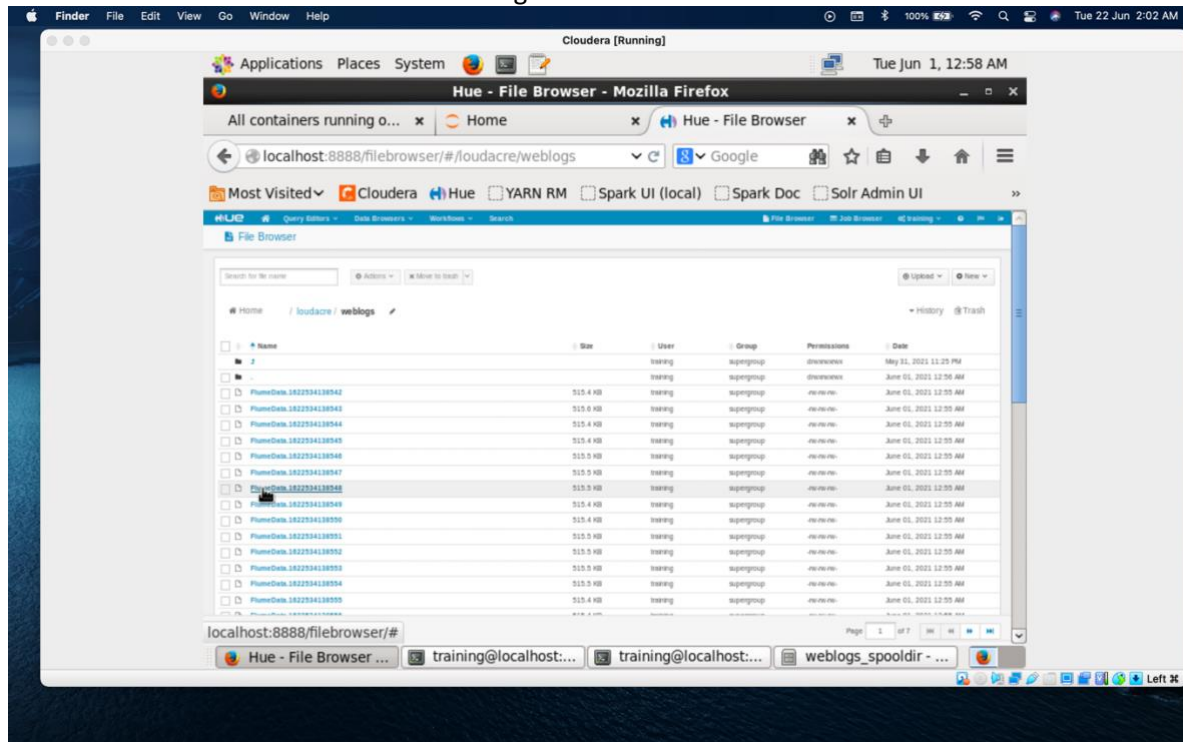
11. Below is the screenshot of the web log files successfully copied to the /flume/weblogs_spooldir.



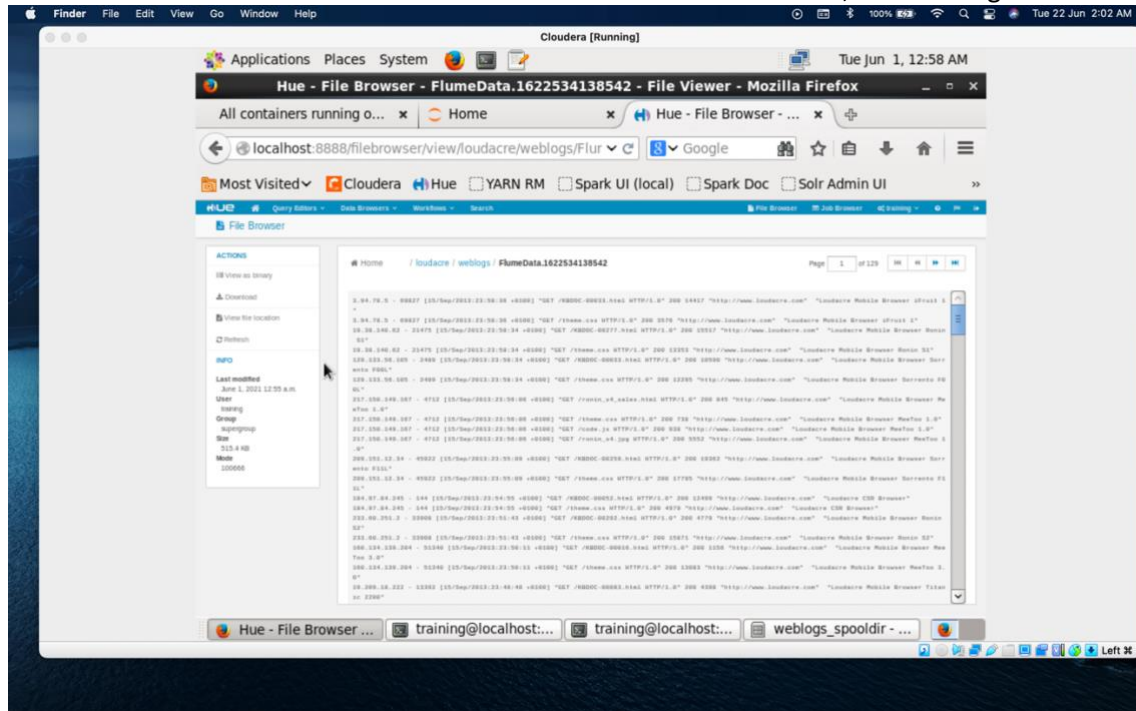
12. Below is the screenshot of the output of the logging information on the terminal on which Flume agent was running. It shows the information about the files Flume ingested into HDFS.



13. Then, I opened the Hue Web Interface and opened the `/loudacre/weblogs` directory where files were added. Below is the screenshot of files added using Flume.



14. Below is the screenshot of one of the files FlumeData.1622534138542, added using Flume.



15. When the operation was finished, the terminal was closed using the command: CTRL + C

