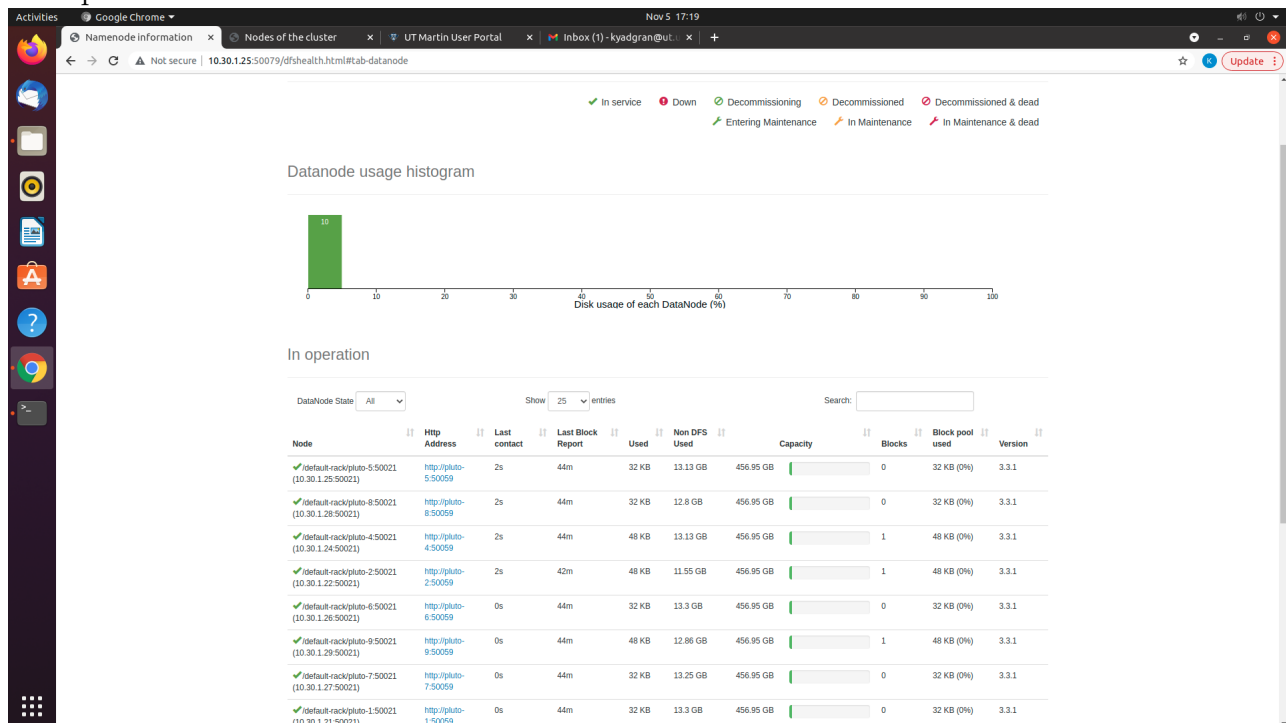


Discussion about HDFS setup and Dataset

Kyannah Grant

November 5, 2021

This assignment was very tedious and hard. It was hard to find ports that were not in use. Ports are very hard to deal with, especially when you don't know how to set one up. I took a long time trying to find out where my node managers were held. Yarn took so long to set up as well. Despite the setup being like HDFS, I still spent a lot of time setting it up. All I feel is relief now that it is set up and running, but I will be very sad if I do something I am not supposed to and I break something. Here are my screenshots of my HDFS and yarn working as well as showing that data is uploaded inside HDFS.



The top screenshot shows the 'Browse Directory' page in the Hadoop web interface. The URL is `1030.125.50079/explorer.html#/user/kyadgran`. The page displays a single file named `yam-site.xml` with a size of 128 MB, last modified on Nov 05 16:50. The file is owned by `kyadgran` and belongs to the `supergroup`. The page also shows a search bar and a 'Go!' button.

The bottom screenshot shows the 'Nodes of the cluster' page, displaying a table of 10 nodes. The table has columns for Name, State, Role, Hostname, Rack, and various metrics. The nodes are all in a 'RUNNING' state and have a role of 'default-rack'.

Name	State	Role	Hostname	Rack	...
iddefault-rack	RUNNING	pluto-3.8042	pluto-3.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-7.8042	pluto-7.8142	Fri Nov 05 16:43:24 -0500 2021	0
iddefault-rack	RUNNING	pluto-10.8042	pluto-10.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-9.8042	pluto-9.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-4.8042	pluto-4.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-1.8042	pluto-1.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-2.8042	pluto-2.8142	Fri Nov 05 16:43:24 -0500 2021	0
iddefault-rack	RUNNING	pluto-6.8042	pluto-6.8142	Fri Nov 05 16:43:23 -0500 2021	0
iddefault-rack	RUNNING	pluto-8.8042	pluto-8.8142	Fri Nov 05 16:43:23 -0500 2021	0

My dataset used for my project will consist of sports articles. The user can use the system to find similar articles the user was reading. They should be able to put the article in and the system will output articles that are similar about the subject. This is my thought for now. It may change in the future if things I think of something in the future, but I like this thought right now though.