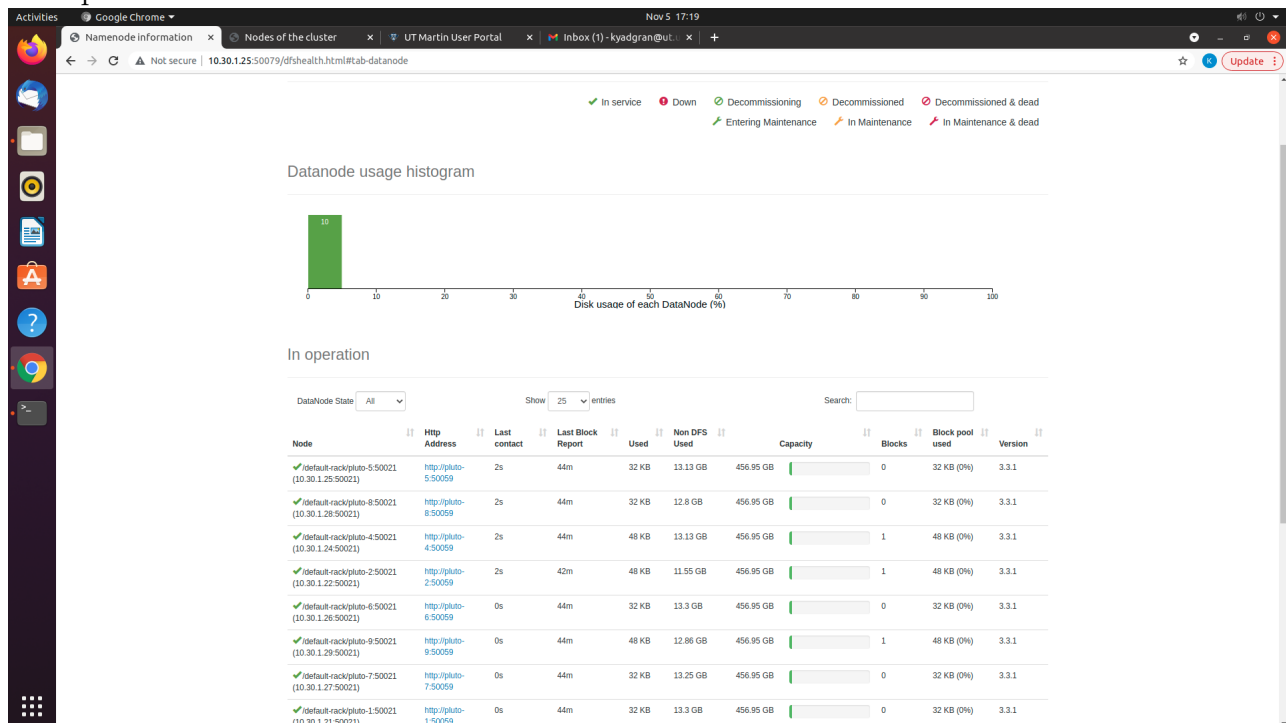# Discussion about HDFS setup and Dataset

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.

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 though 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.