**Group 3: practice 4**

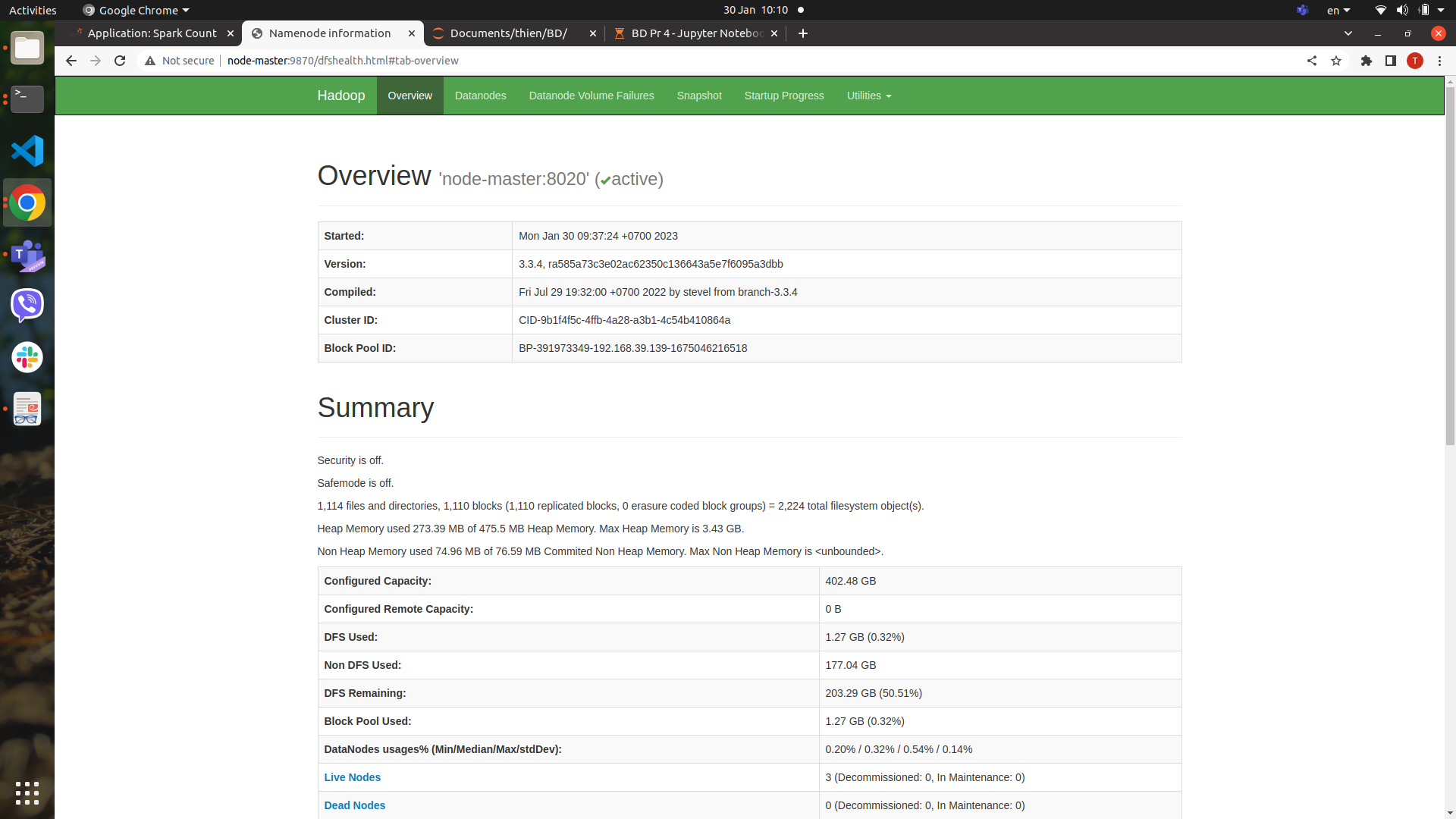
Nguyễn Huy Hoàng – 20194433

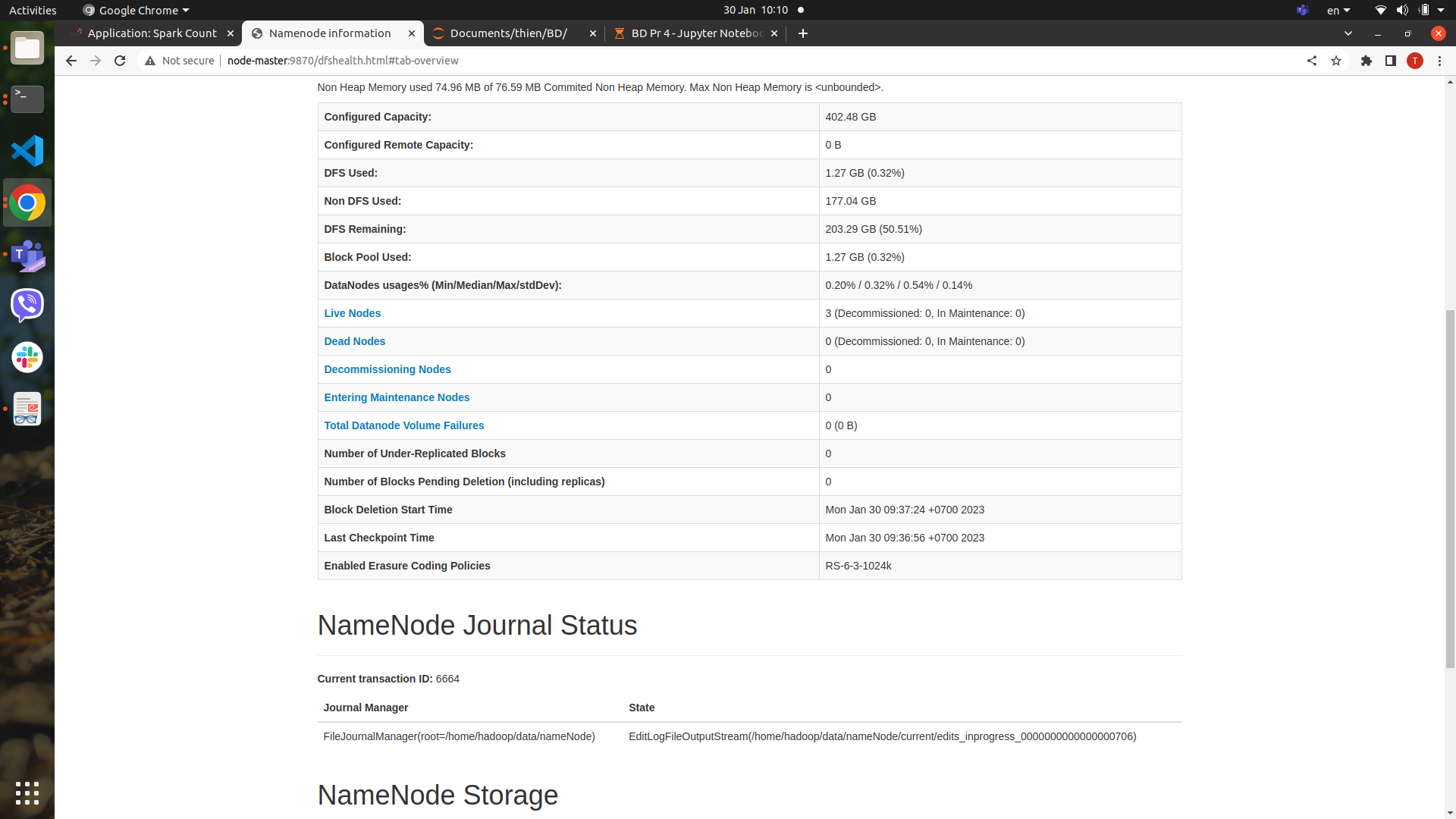
Trần Quốc Khánh – 20190085

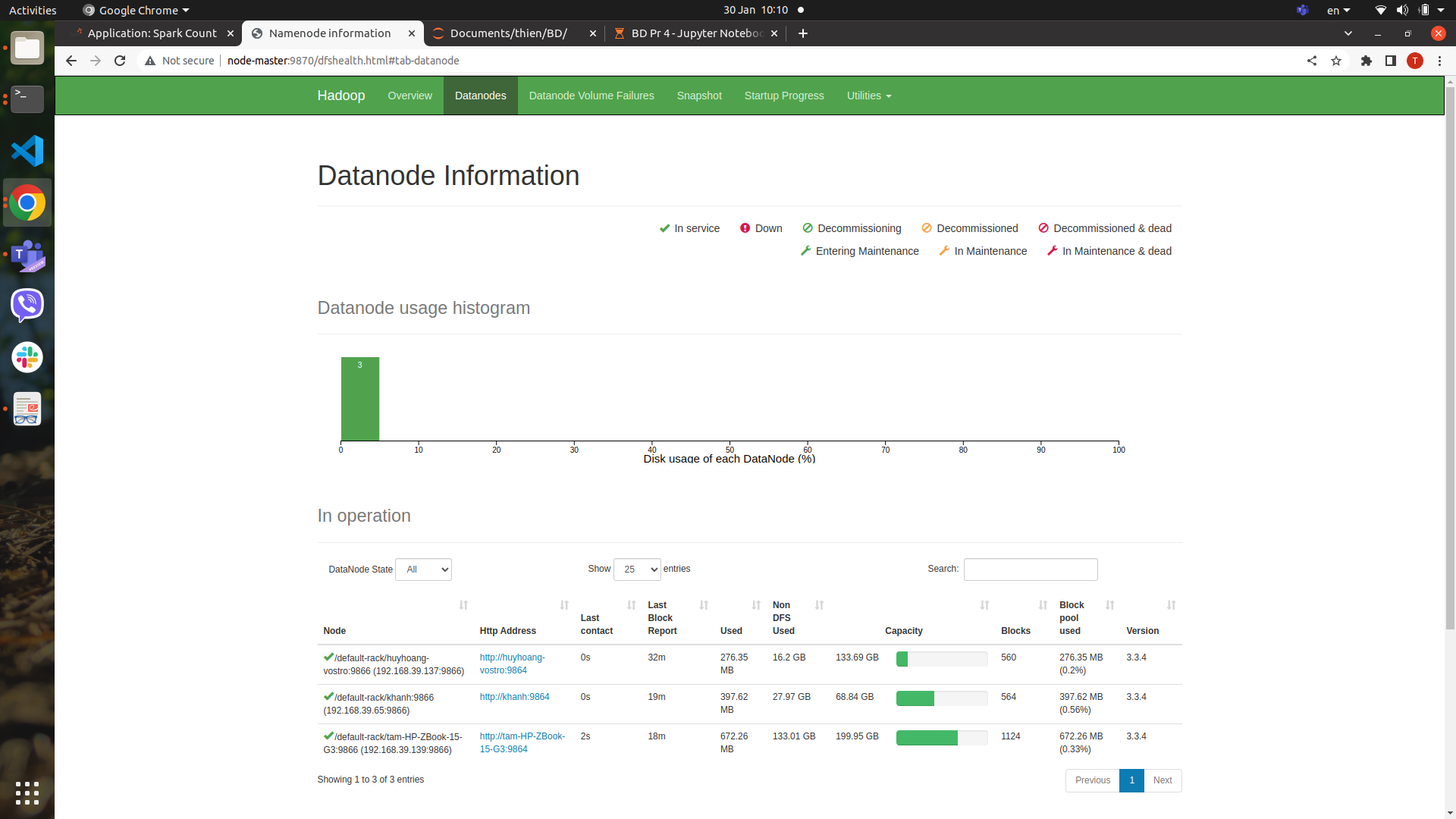
Hoàng Thiện Tâm – 20194450

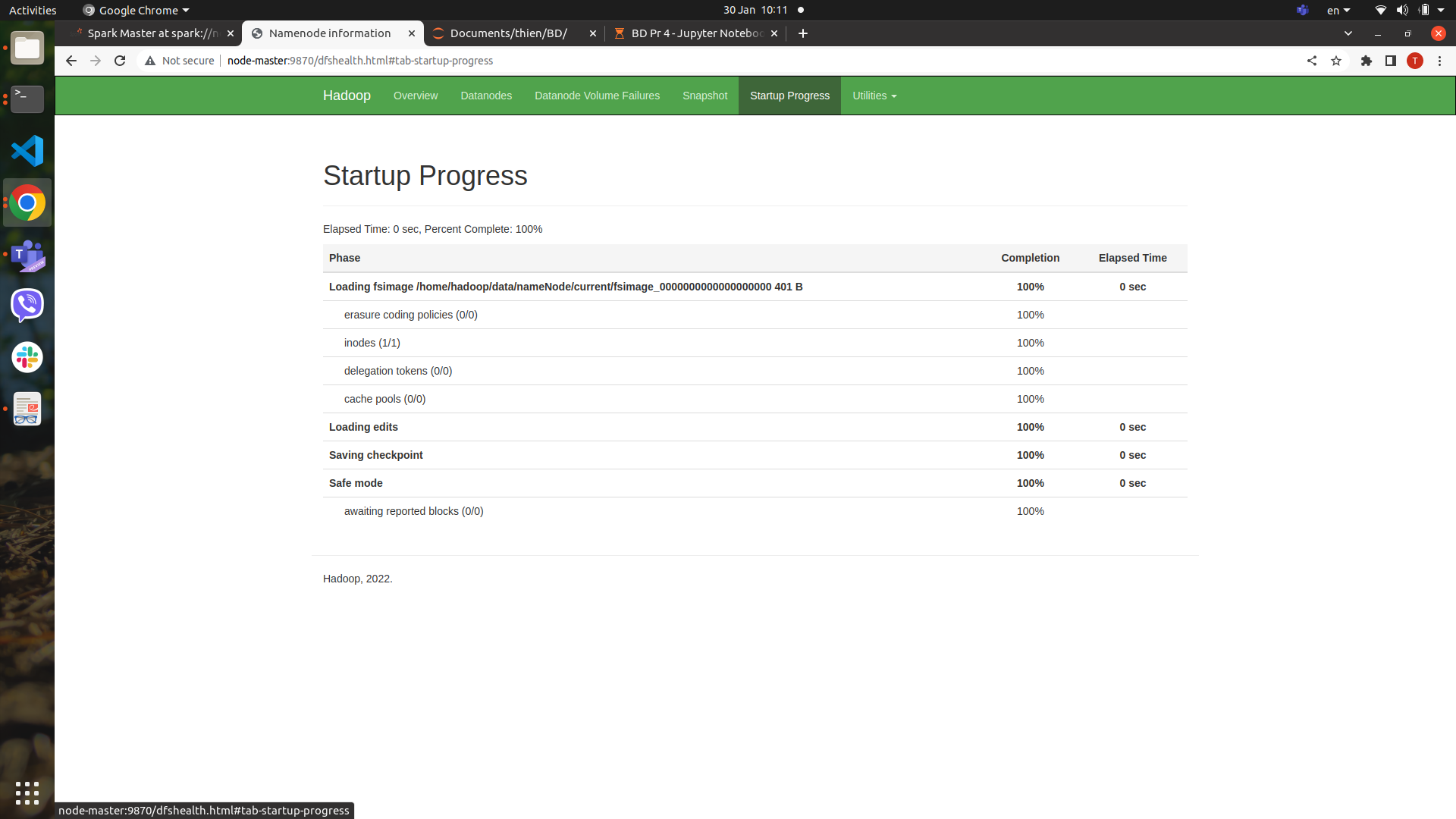
Phạm Như Thuấn - 20194456

* We set up the 3 nodes cluster:

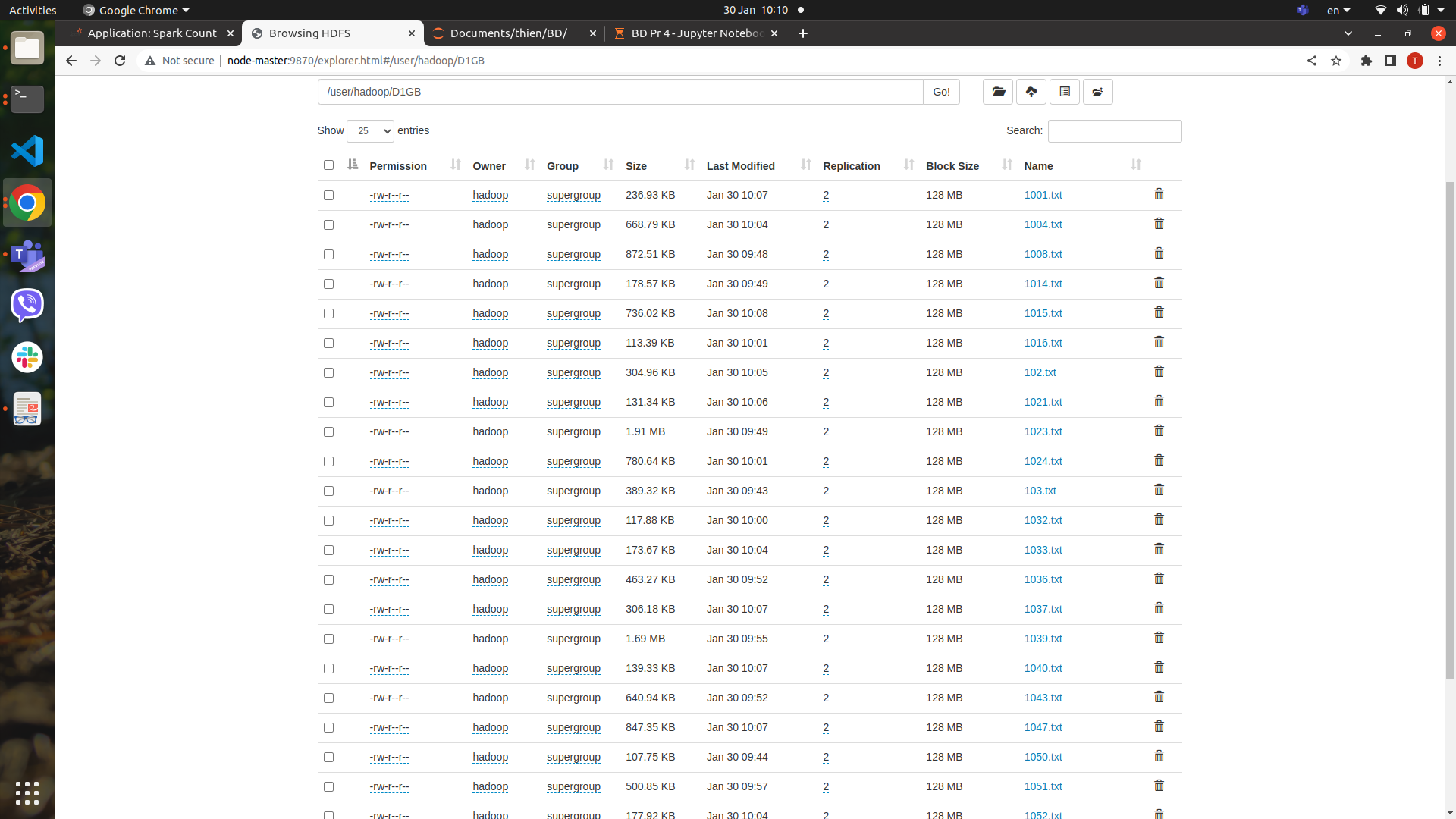




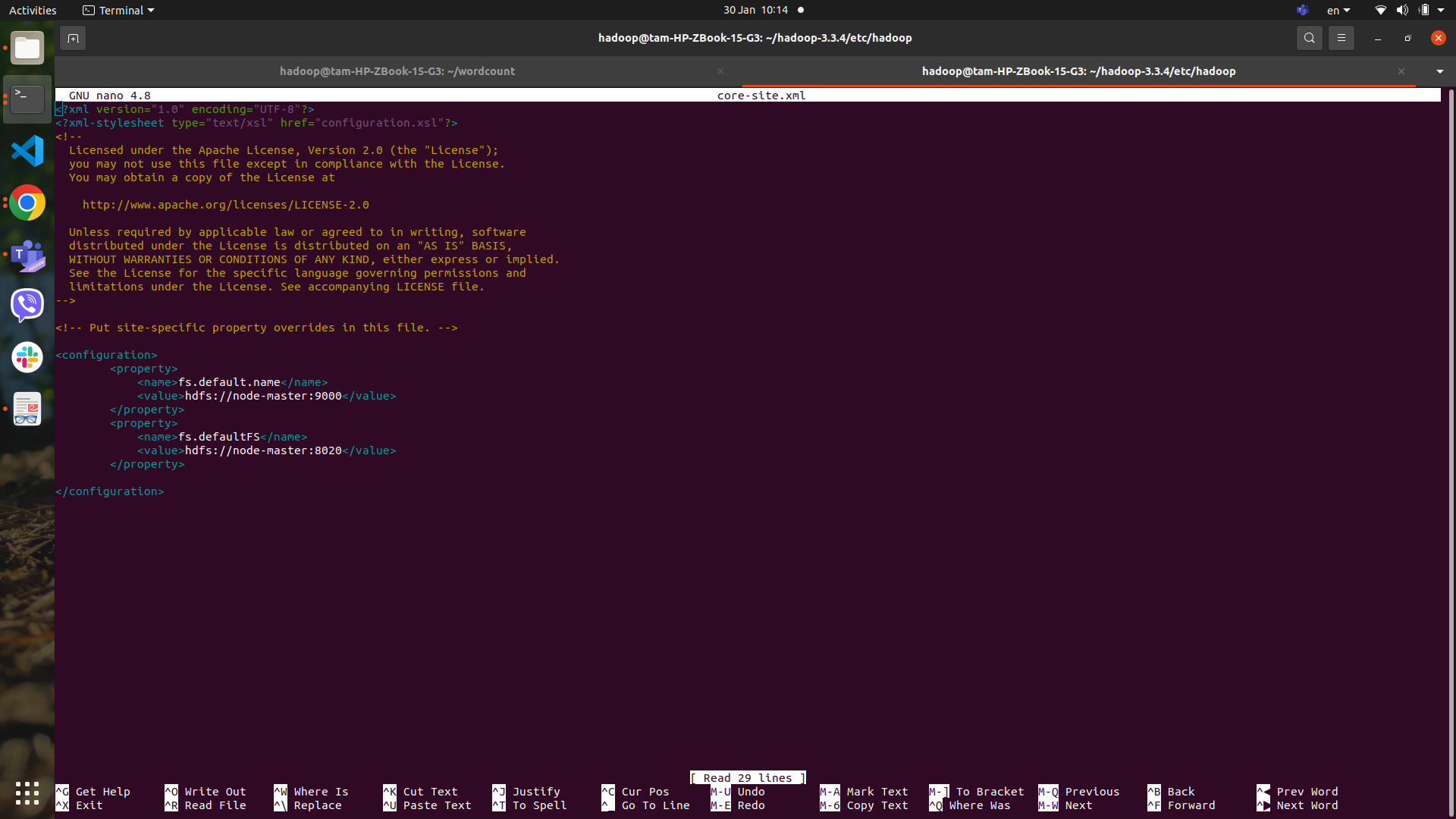




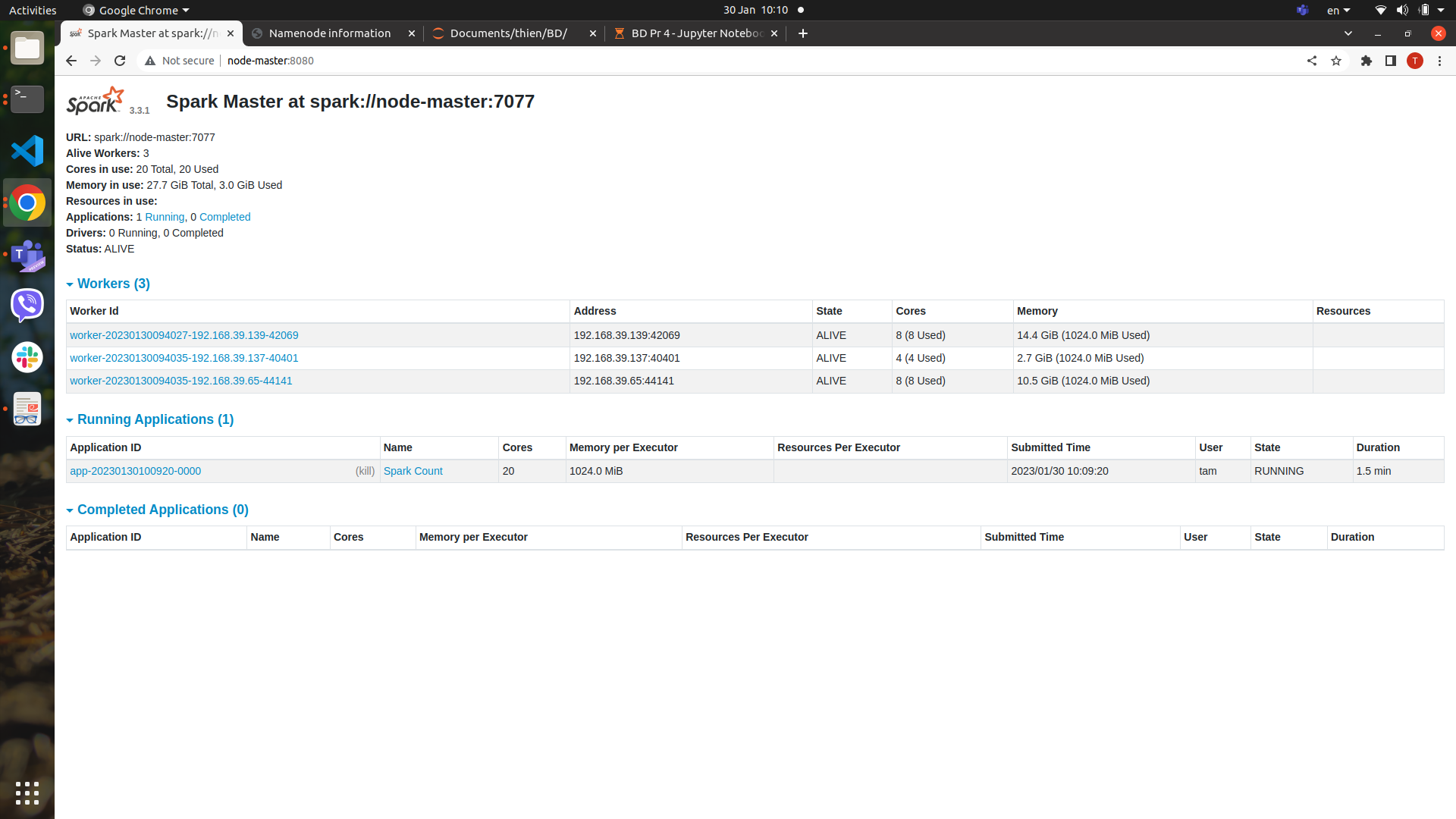
* Put 1 Gb data



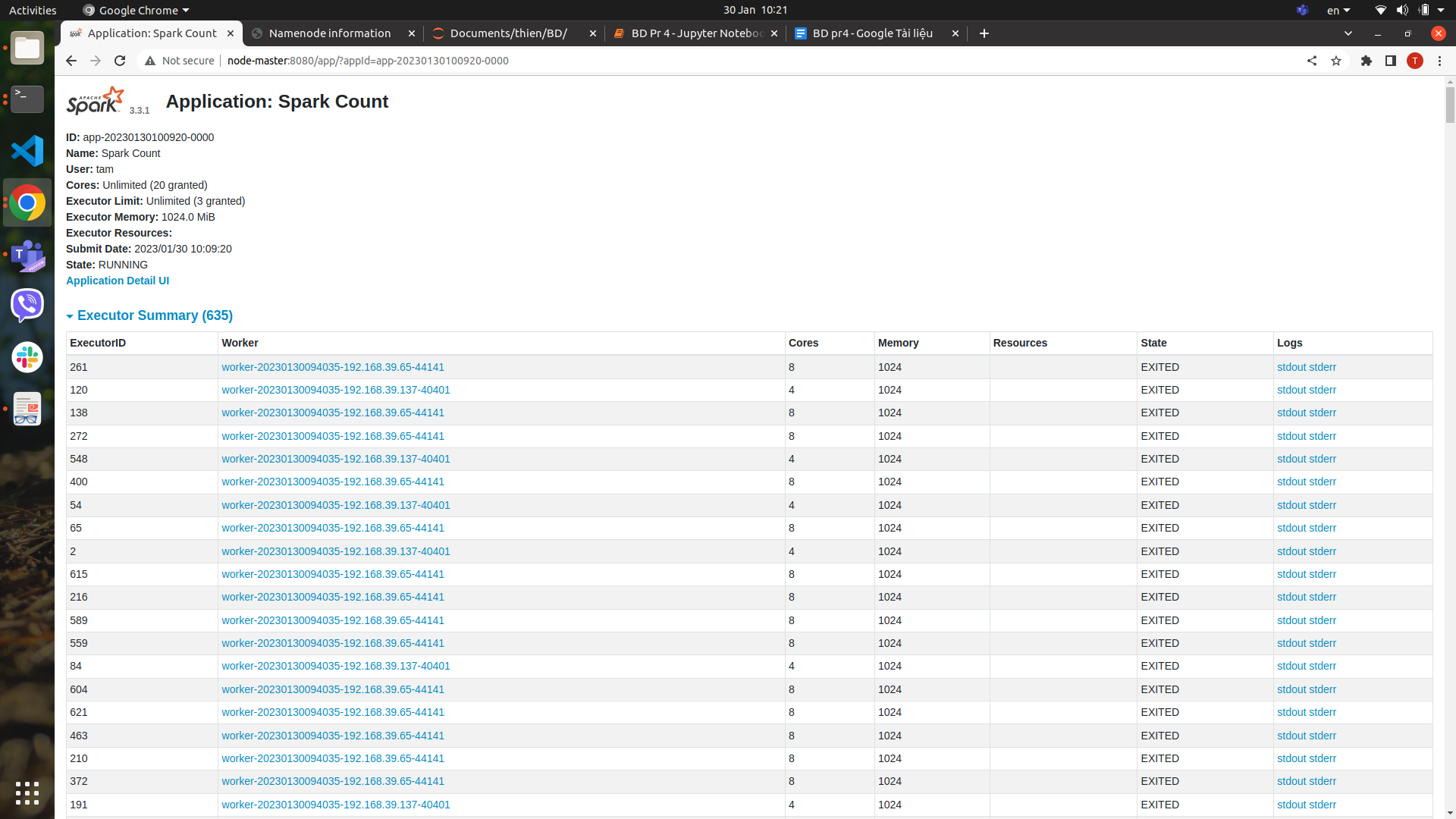
* Specify the url to access the data in hdfs



* Application on spark api



* Job detail of the application Wordcount



* Jupyter code and result

