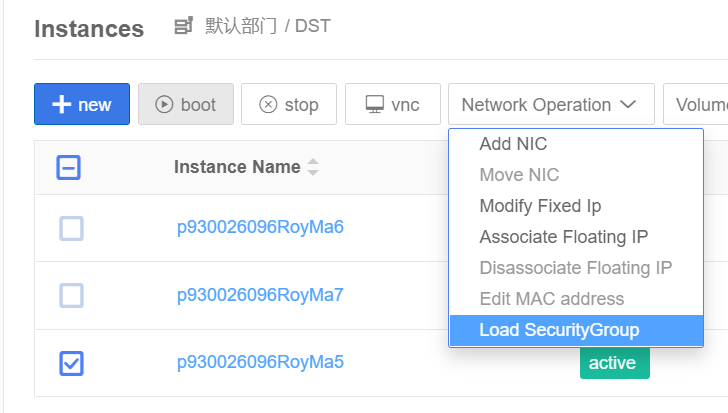
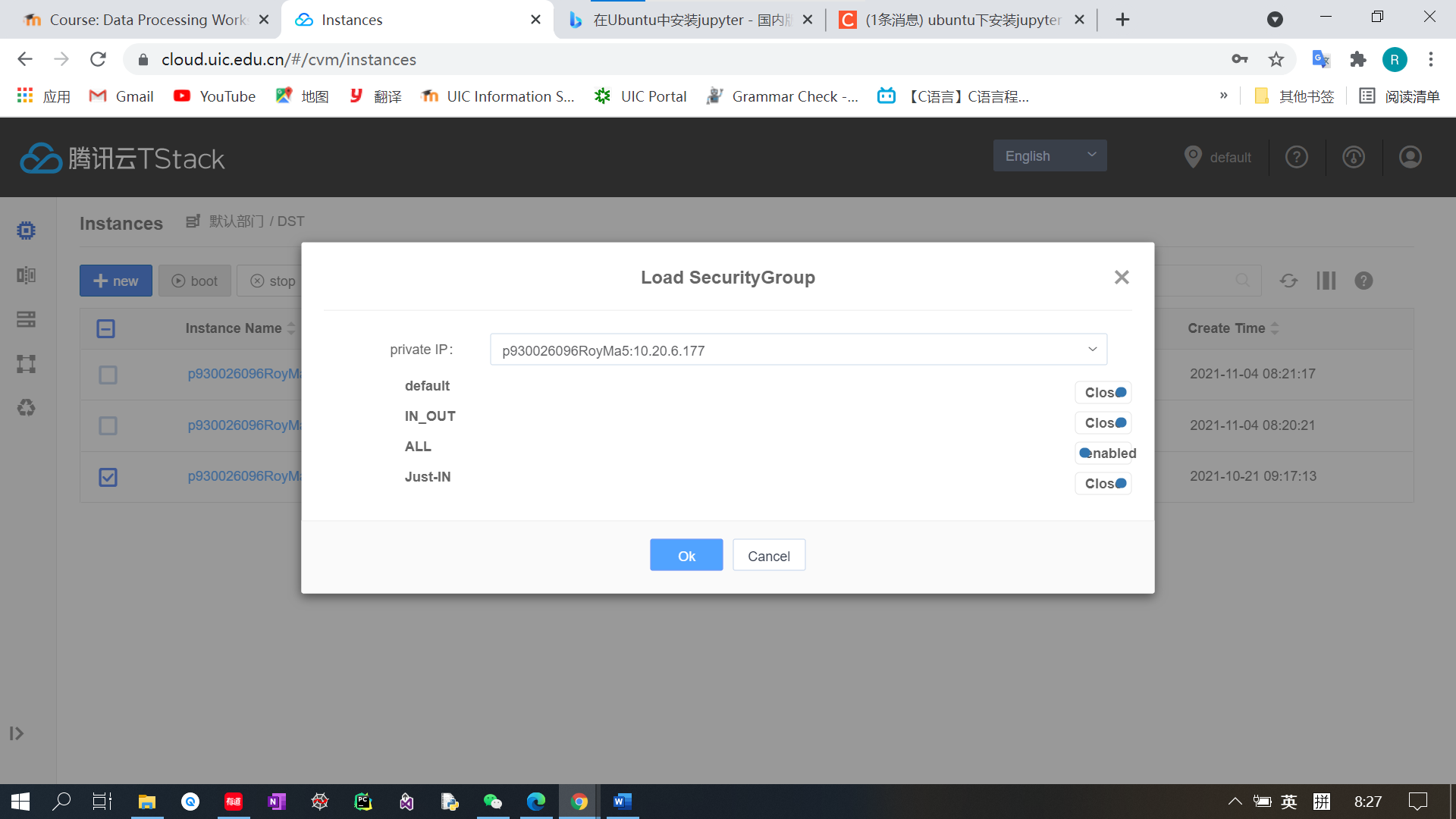
Intermediate



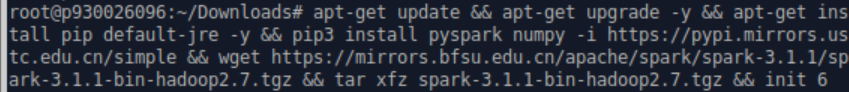
Reset the security group of all the workers and master.



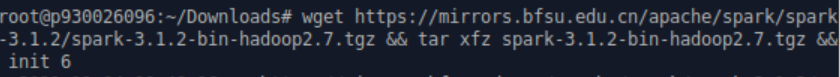
Set the security group to “ALL”



Enter the downloads directory.



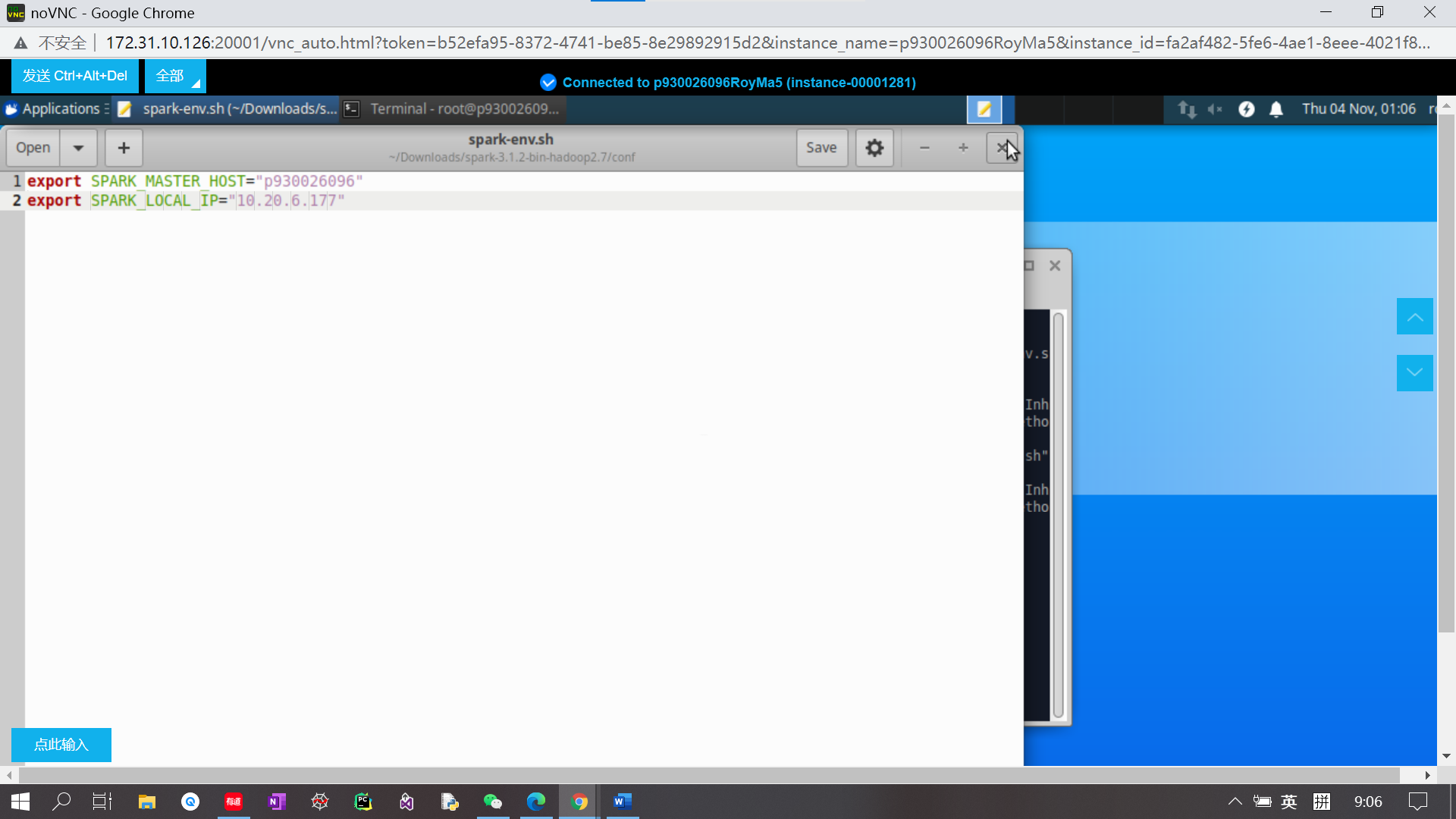
While the web page of spark-3.1.1 is 404(not found)



Instead, we download 3.1.2



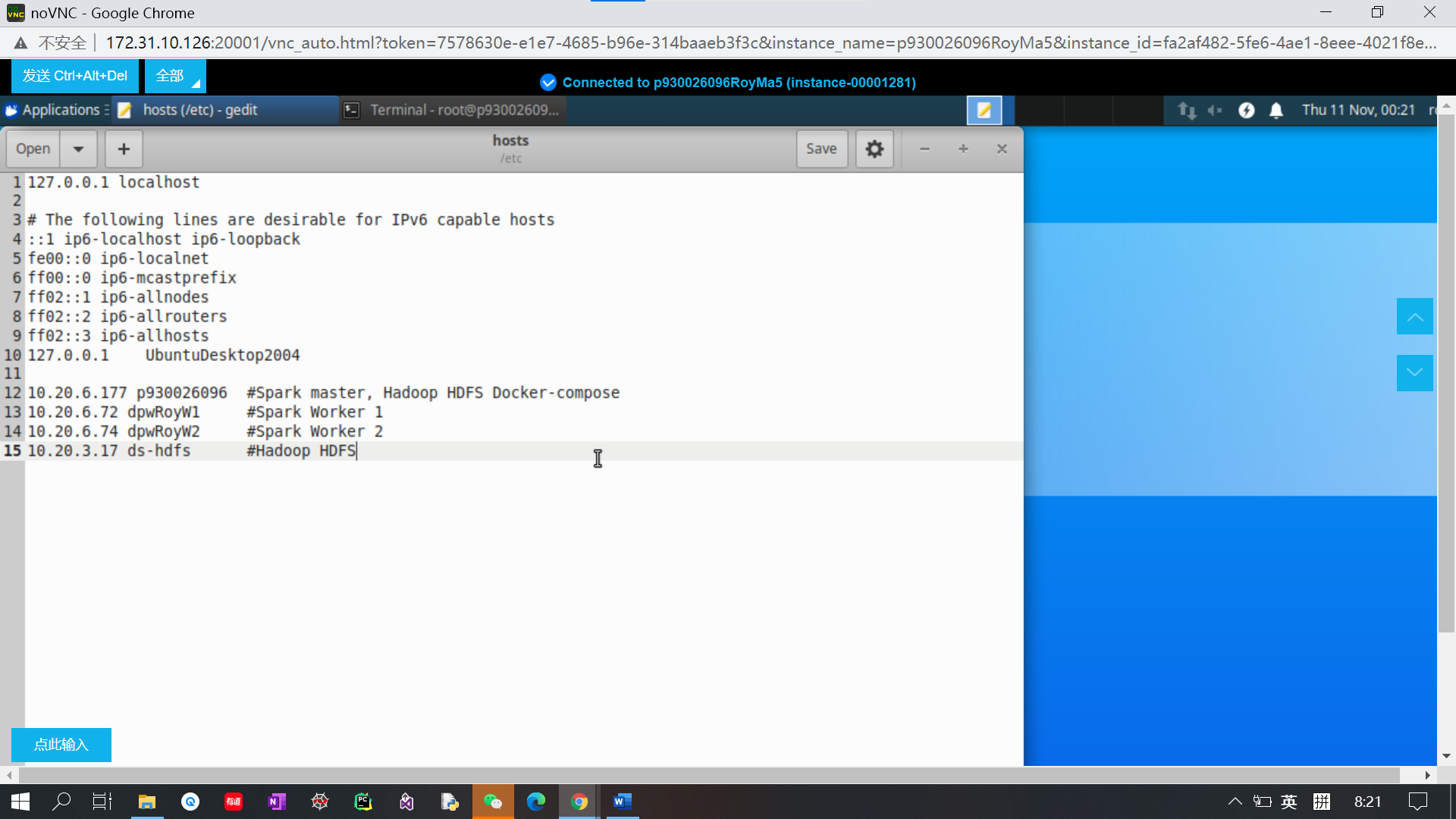
Edit the configuration .sh file of spark-env.



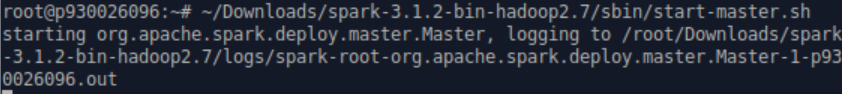
Add master host and local IP to the file.



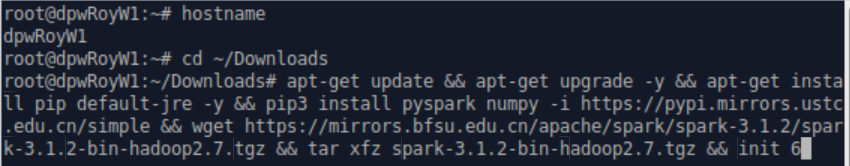
Edit file “/etc/hosts”



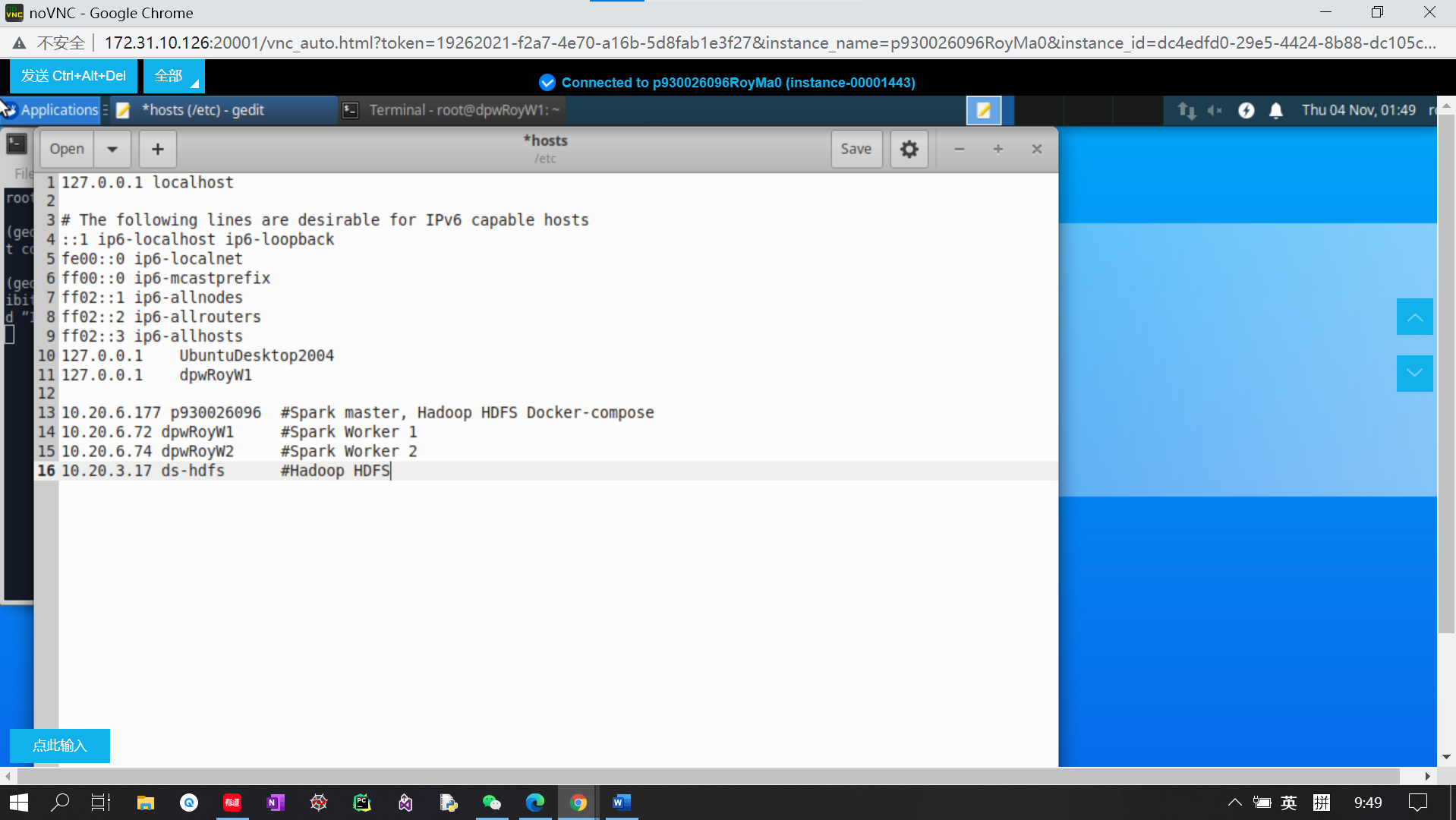
Add the master and workers’ name and local IP into the file. After saving the file, the configuration of the master is done.



Then we start master.



For the first worker (dpwRoyW1), we install jre, pyspark numpy, spark-3.1.2-hadoop2.7.

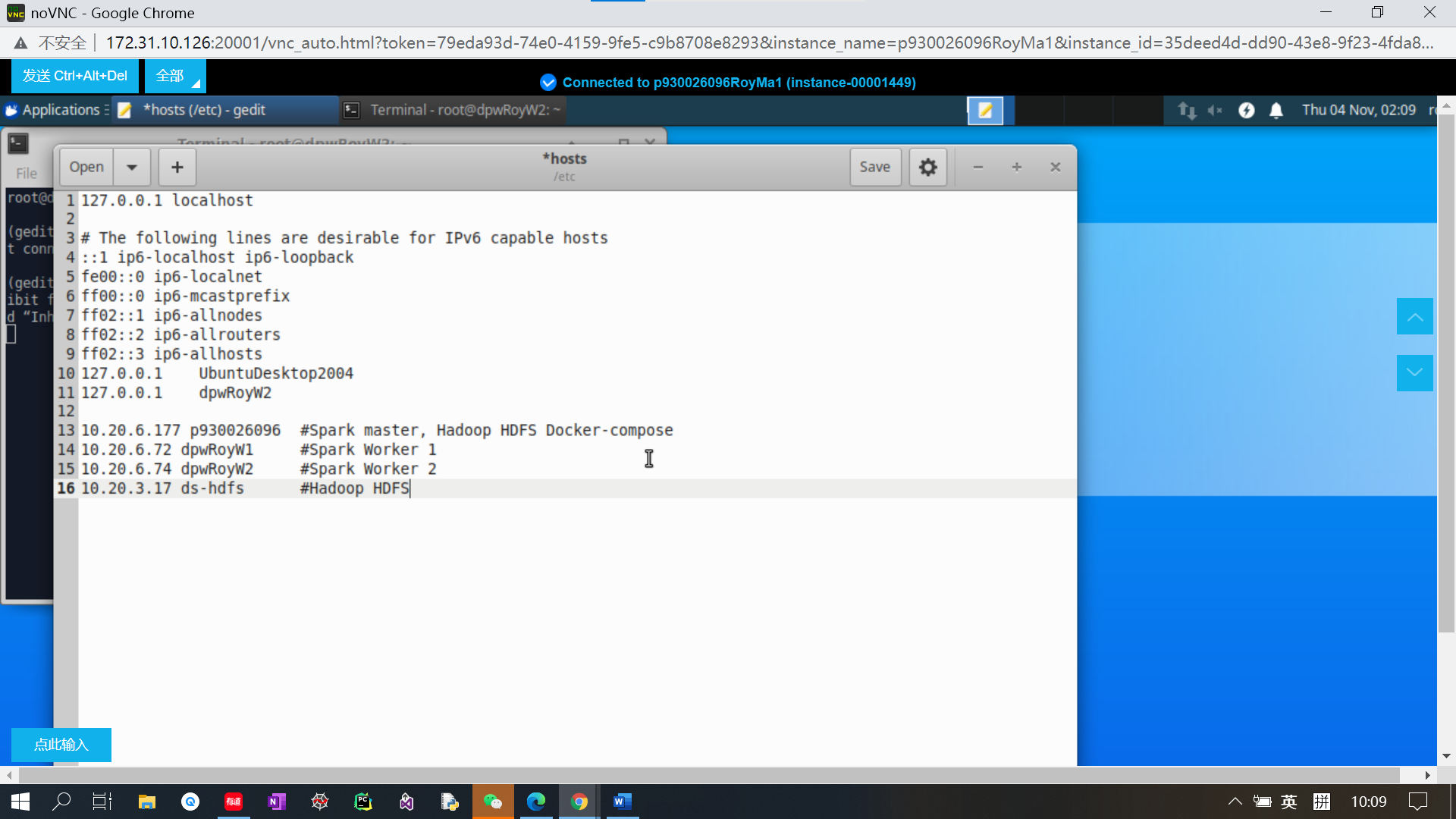


Add workers and master’s hostname and local IP in the worker’s conf .sh file.



Start worker1 by the command.

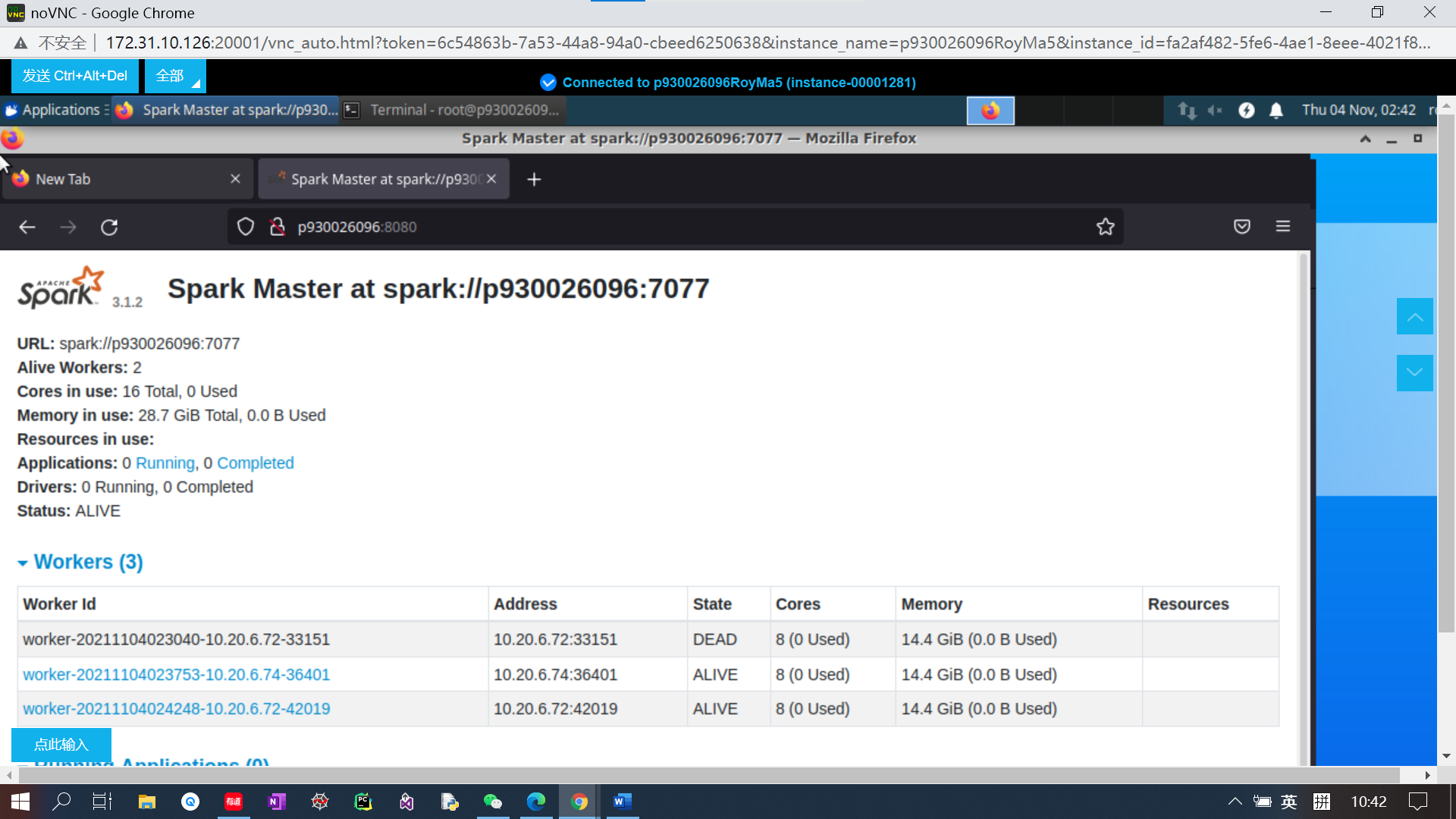
The step of worker2 is almost the same with the worker1 first do the installation.



Add the workers and master’s hostname and local IP of the configuration file of worker2.



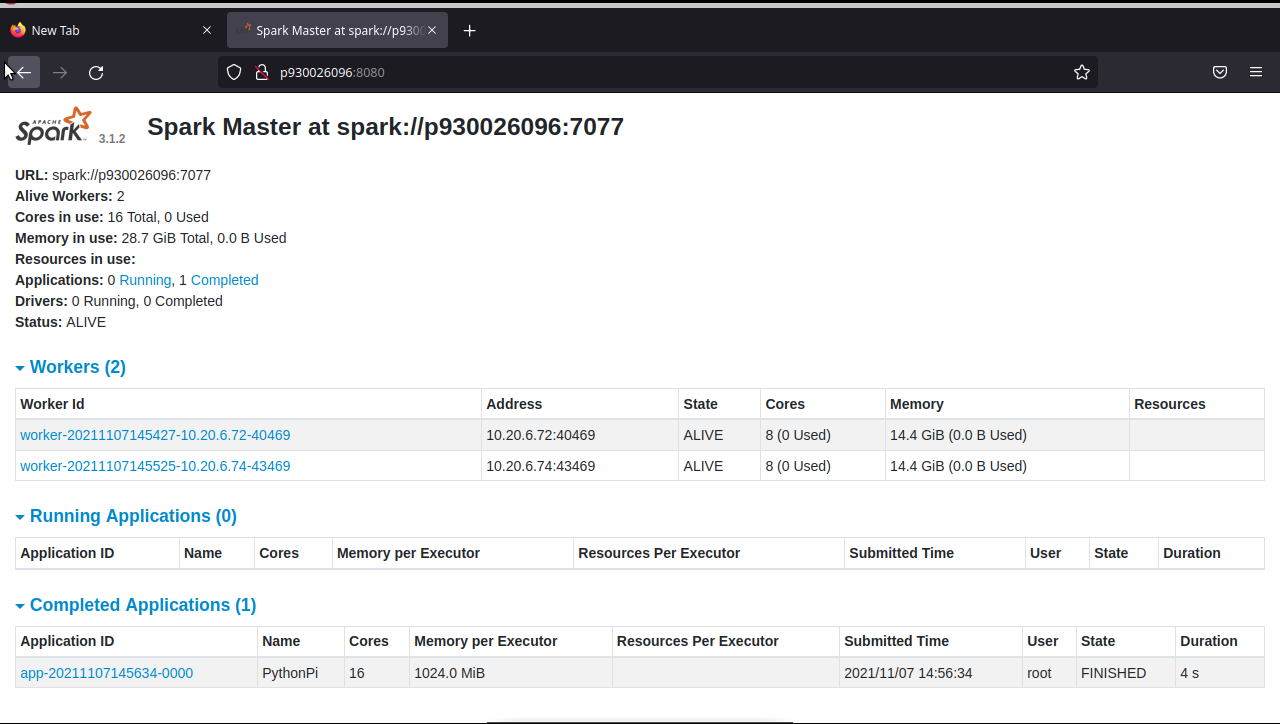
Start worker2 by the command.



Open the website “p930026096:7077” to check the current state of the spark cluster. Now there is a master and there are two alive workers as shown on the graph.



Execute the pi.py code by pyspark-submit



Check the website “p930026096:7077”, and there is an application under completed applications which is “PythonPI”.