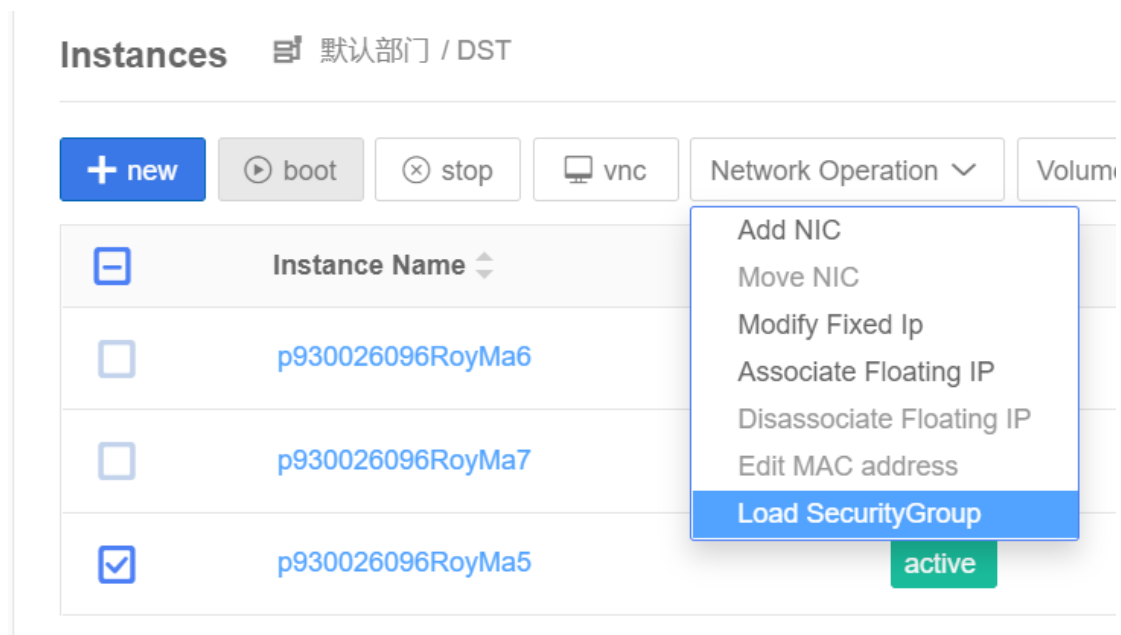
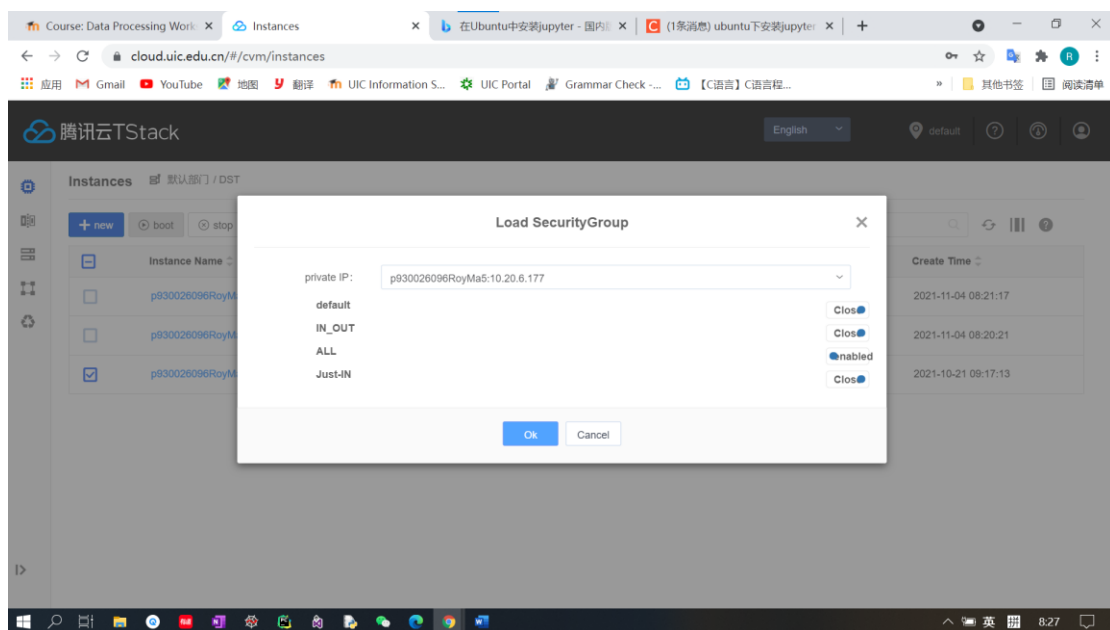


Intermediate



Reset the security group of all the workers and master.



Set the security group to "ALL"

```
root@p930026096:~# cd ~/Downloads
```

Enter the downloads directory.

```
root@p930026096:~/Downloads# apt-get update && apt-get upgrade -y && apt-get install pip default-jre -y && pip3 install pyspark numpy -i https://pypi.mirrors.us
tc.edu.cn/simple && wget https://mirrors.bfsu.edu.cn/apache/spark/spark-3.1.1/spark-3.1.1-bin-hadoop2.7.tgz && tar xzf spark-3.1.1-bin-hadoop2.7.tgz && init 6
```

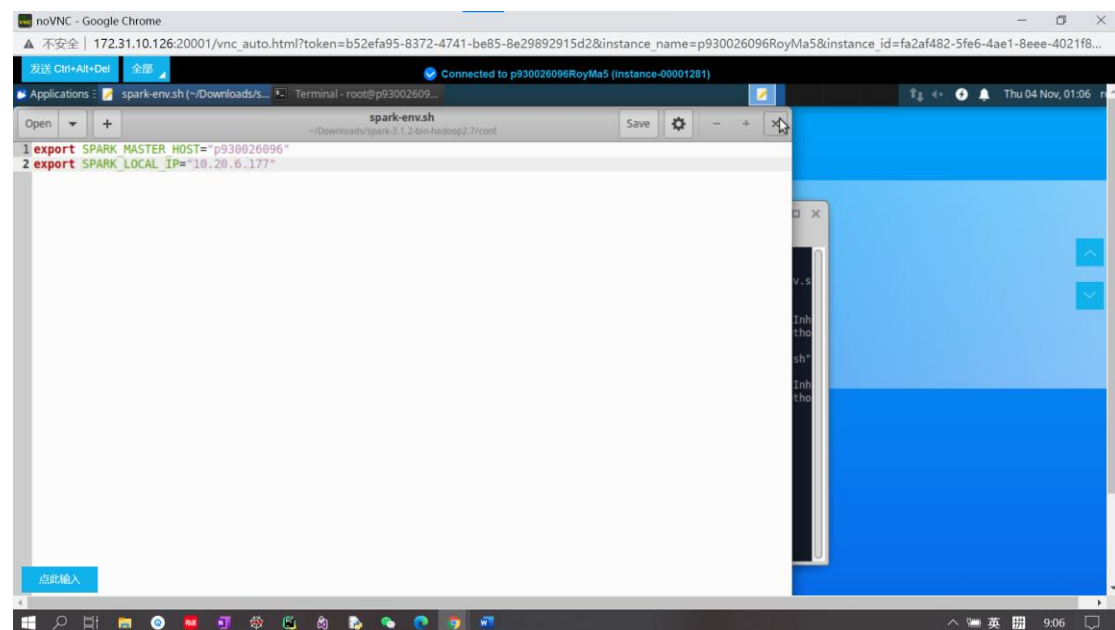
While the web page of spark-3.1.1 is 404(not found)

```
root@p930026096:~/Downloads# wget https://mirrors.bfsu.edu.cn/apache/spark/spark-3.1.2/spark-3.1.2-bin-hadoop2.7.tgz && tar xzf spark-3.1.2-bin-hadoop2.7.tgz && init 6
```

Instead, we download 3.1.2

```
root@p930026096:~# gedit Downloads/spark-3.1.2-bin-hadoop2.7/conf/spark-env.sh
```

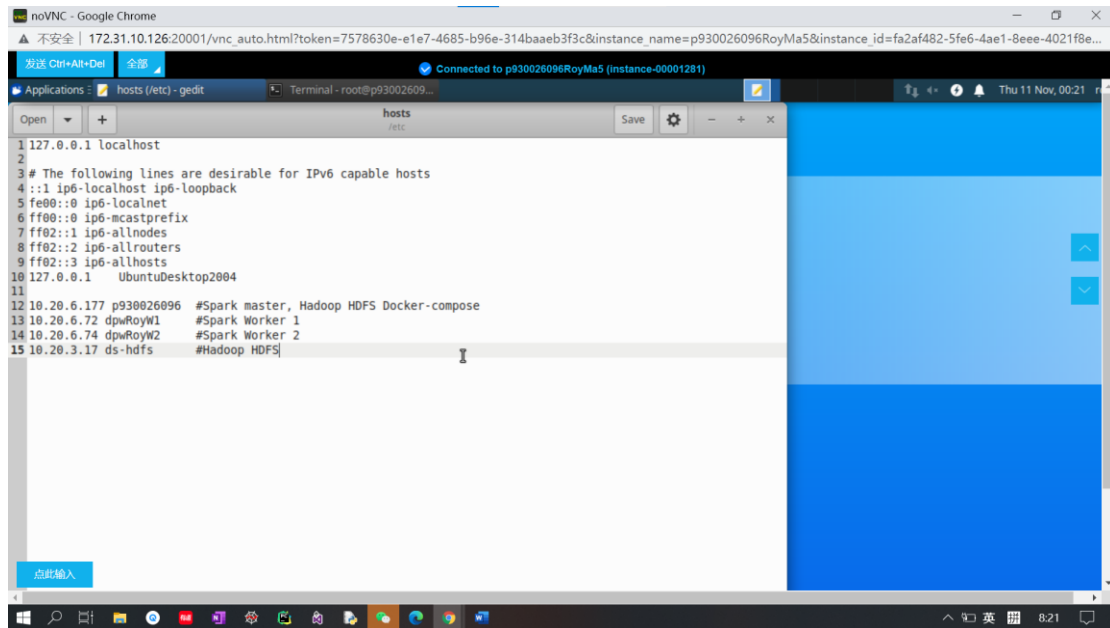
Edit the configuration .sh file of spark-env.



Add master host and local IP to the file.

```
root@p930026096:~# gedit /etc/hosts
```

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.

```

root@p930026096:~# ~/Downloads/spark-3.1.2-bin-hadoop2.7/sbin/start-master.sh
starting org.apache.spark.deploy.master.Master, logging to /root/Downloads/spark-3.1.2-bin-hadoop2.7/logs/spark-root-org.apache.spark.deploy.master.Master-1-p930026096.out

```

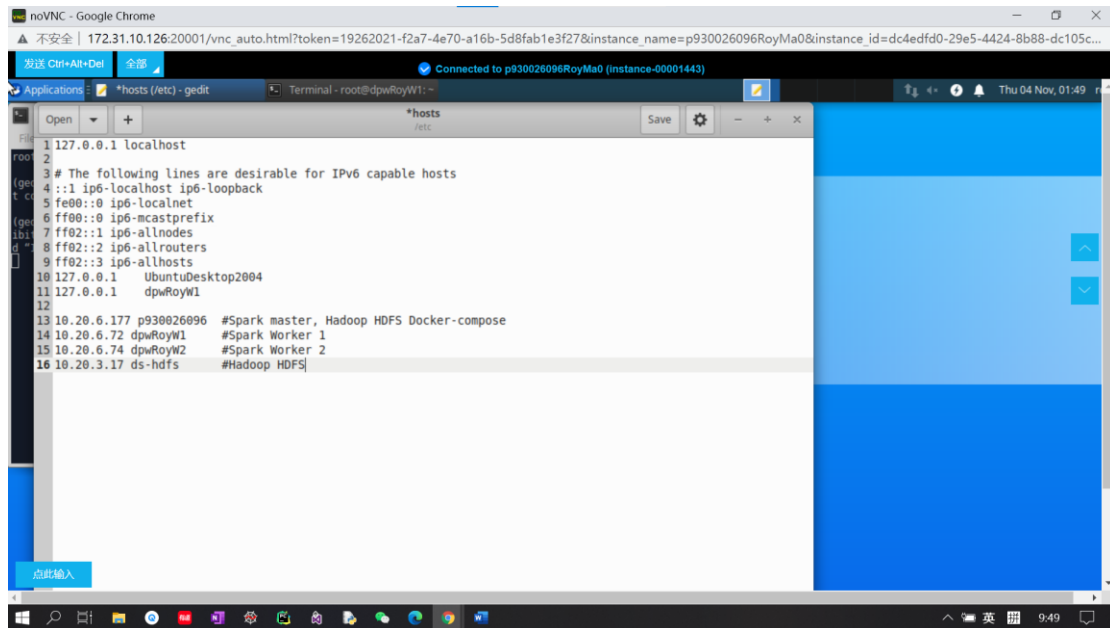
Then we start master.

```

root@dpwRoyW1:~# hostname
dpwRoyW1
root@dpwRoyW1:~# cd ~/Downloads
root@dpwRoyW1:~/Downloads# apt-get update && apt-get upgrade -y && apt-get install pip default-jre -y && pip3 install pyspark numpy -i https://pypi.mirrors.ustc.edu.cn/simple && wget https://mirrors.bfsu.edu.cn/apache/spark/spark-3.1.2/spark-3.1.2-bin-hadoop2.7.tgz && tar xzf spark-3.1.2-bin-hadoop2.7.tgz && init 6

```

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.

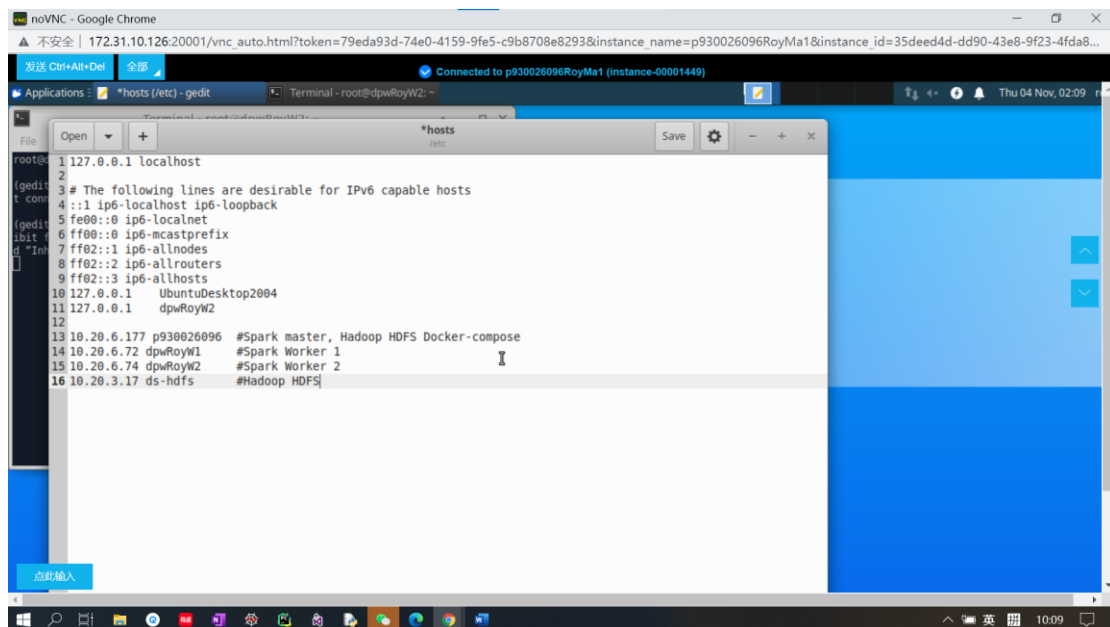
```

root@dpwRoyW1:~# ~/Downloads/spark-3.1.2-bin-hadoop2.7/sbin/start-worker.sh p930026096:7077

```

Start worker1 by the command.

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



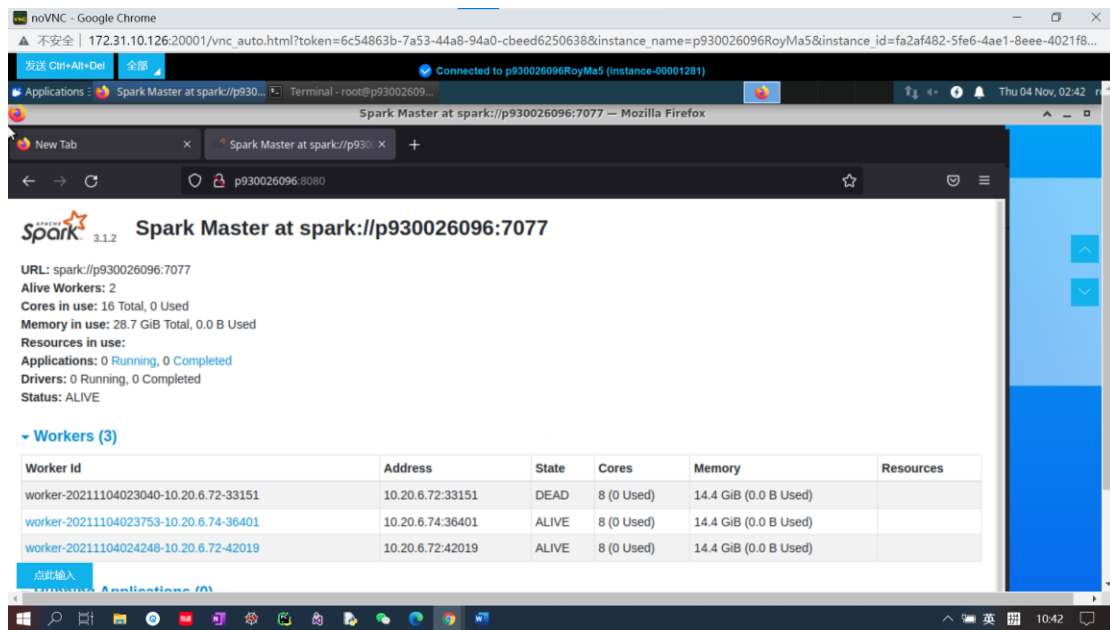
Add the workers and master's hostname and local IP of the configuration file of worker2.

```

root@dpwRoyW2:~# ~/Downloads/spark-3.1.2-bin-hadoop2.7/sbin/start-worker.sh p930026096:7077

```

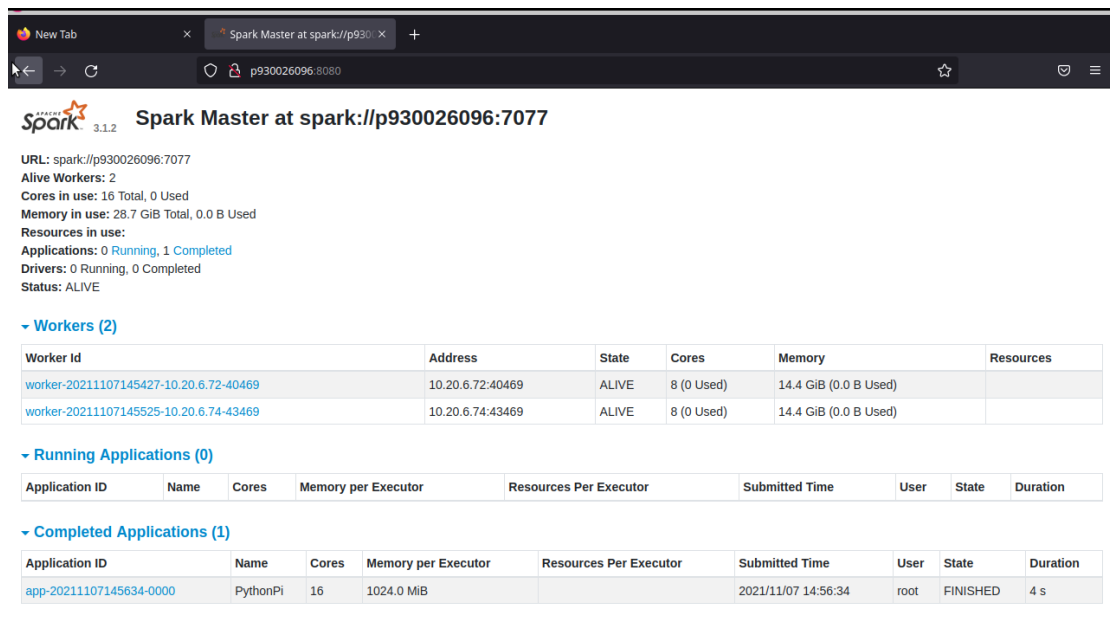
Start worker2 by the command.



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.

```
root@p930026096:~/Downloads/spark-3.1.2-bin-hadoop2.7# bin/spark-submit --master spark://p930026096:7077 examples/src/main/python/pi.py
```

Execute the pi.py code by pyspark-submit



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