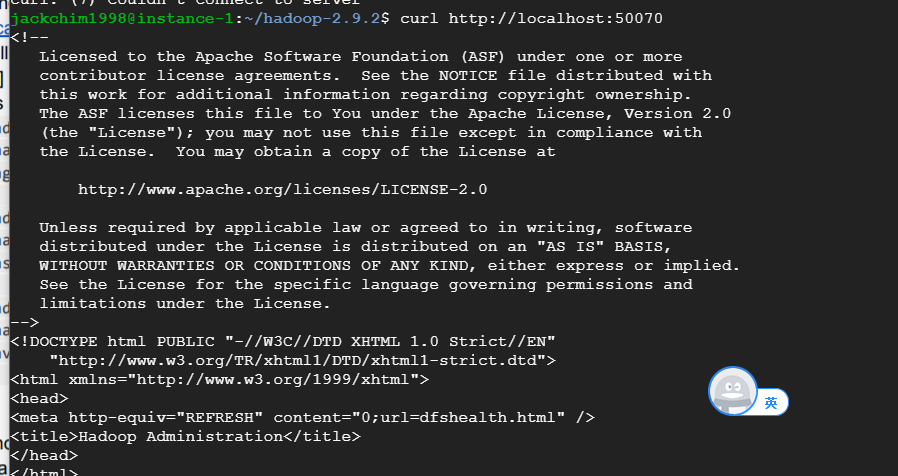
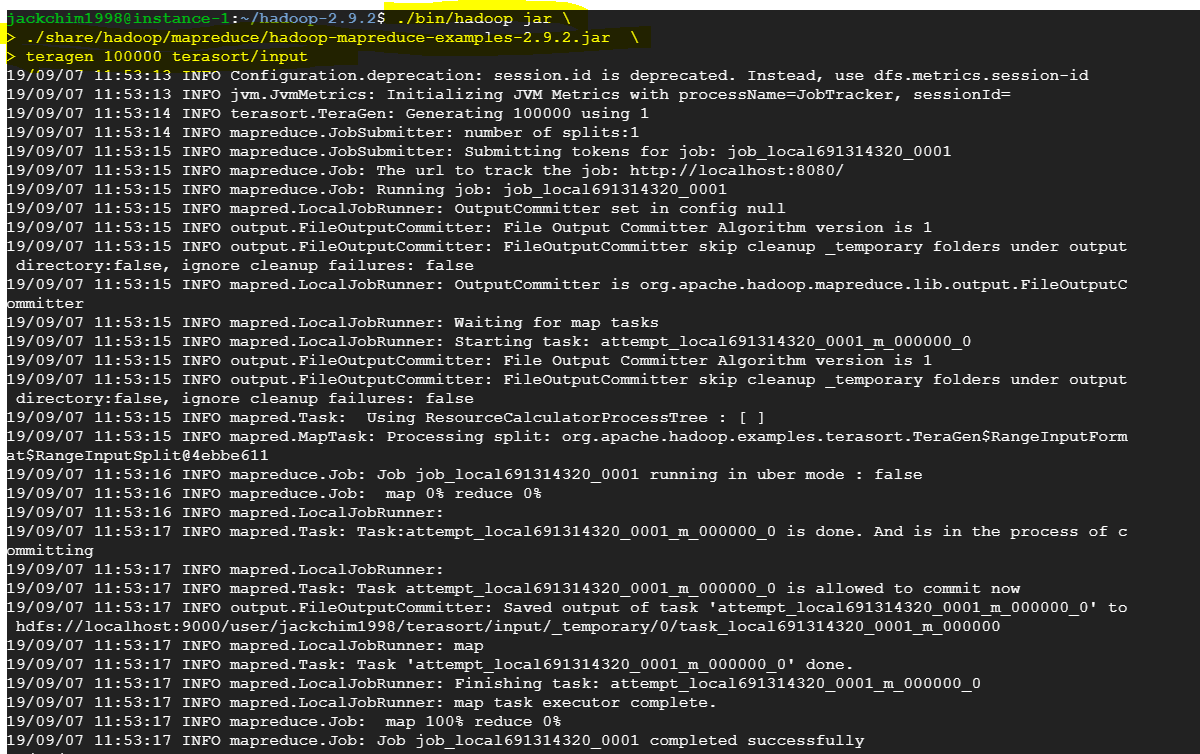
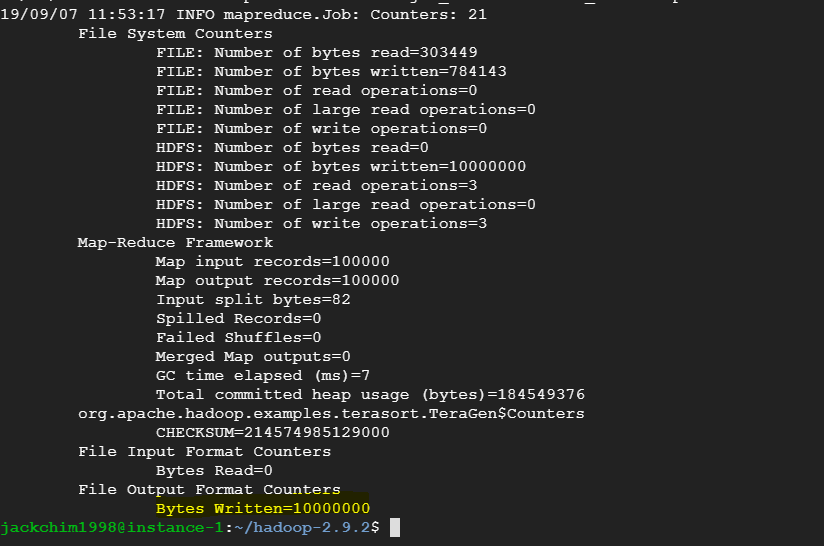
1. [20 marks] Single-node Hadoop Setup

After installation, we browse the port 50070 of localhost successfully.

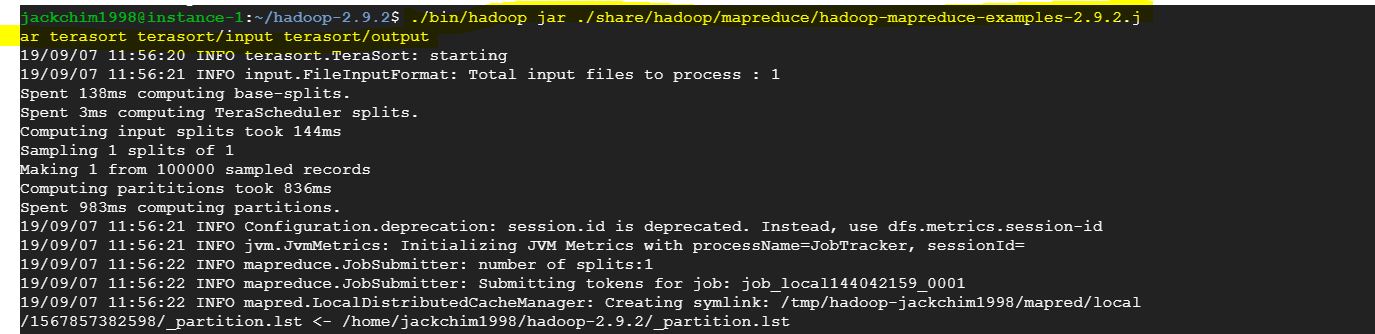


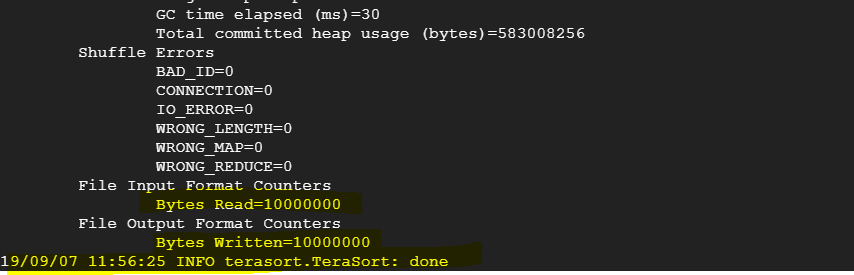
We try to run terasort example. Firstly, generate input file of random number.



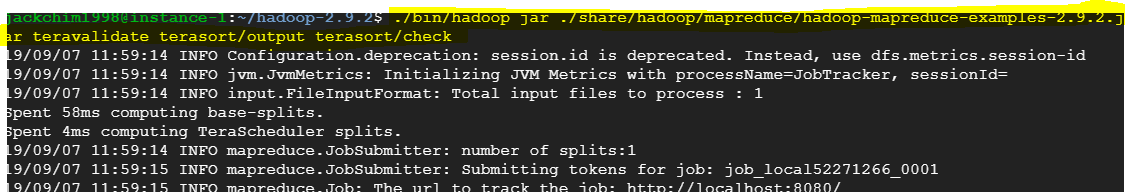


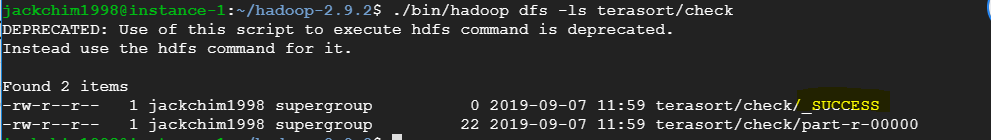
Then sorting it to output file.





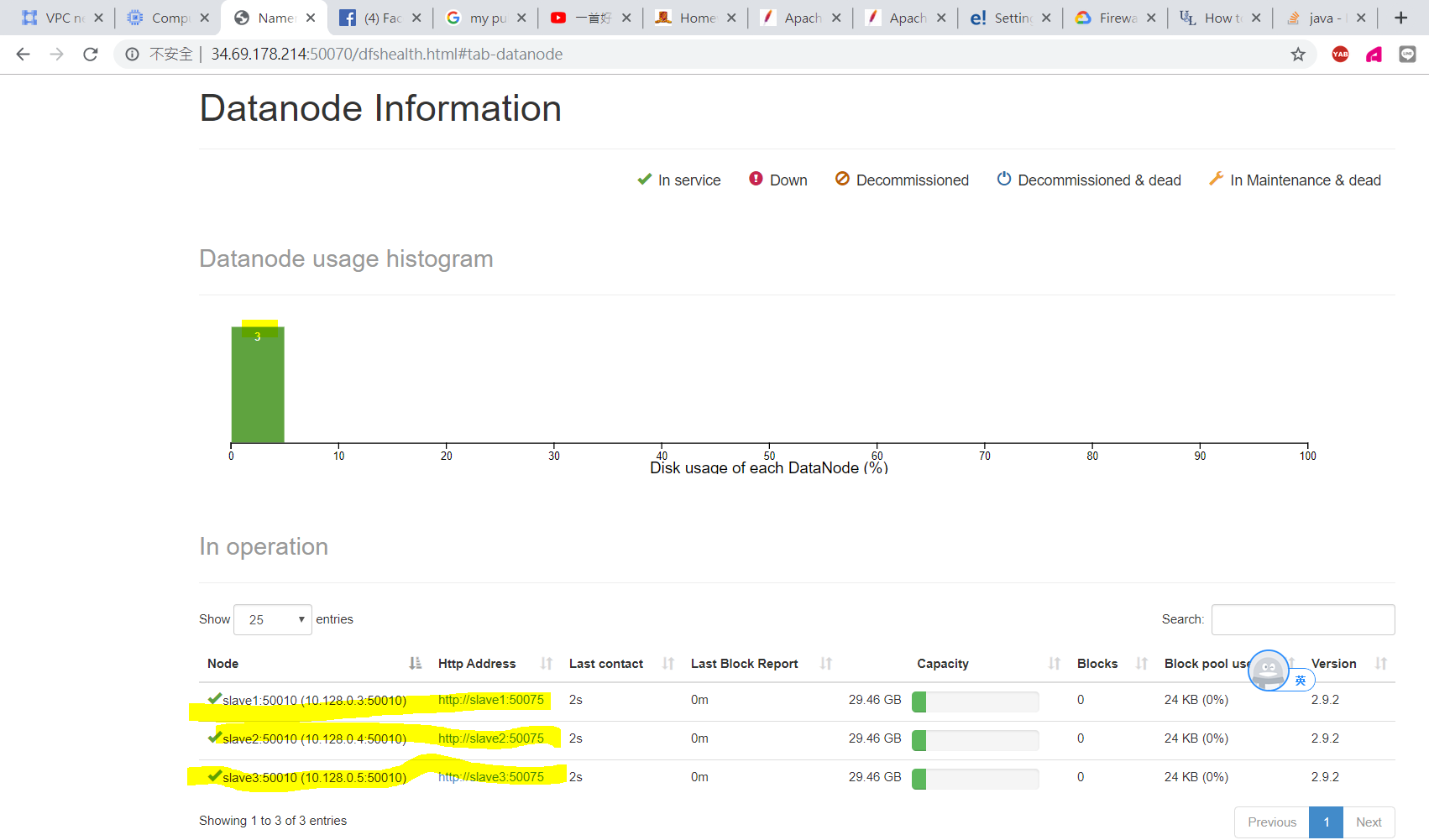
Finally, run teravalidate to validate the sorting result.



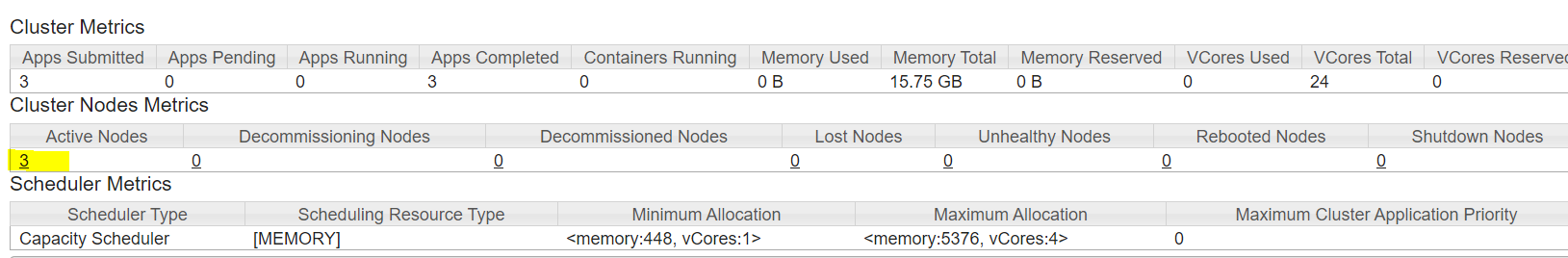


[40 marks] Multi-node Hadoop Cluster Setup

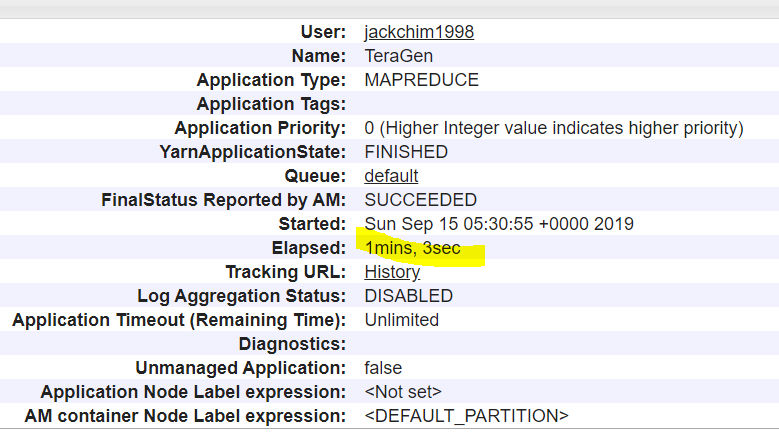
Set up 3 slaves successfully



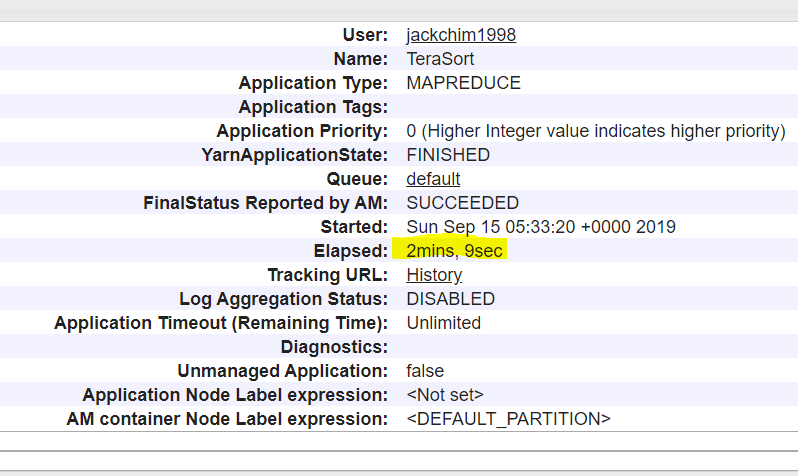
They are all active nodes



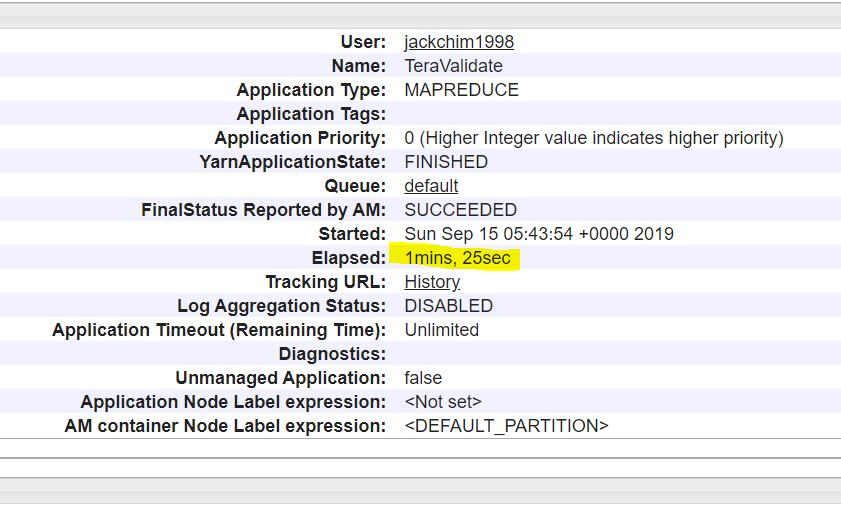
We teragen 2GB data with 1min 3 sec



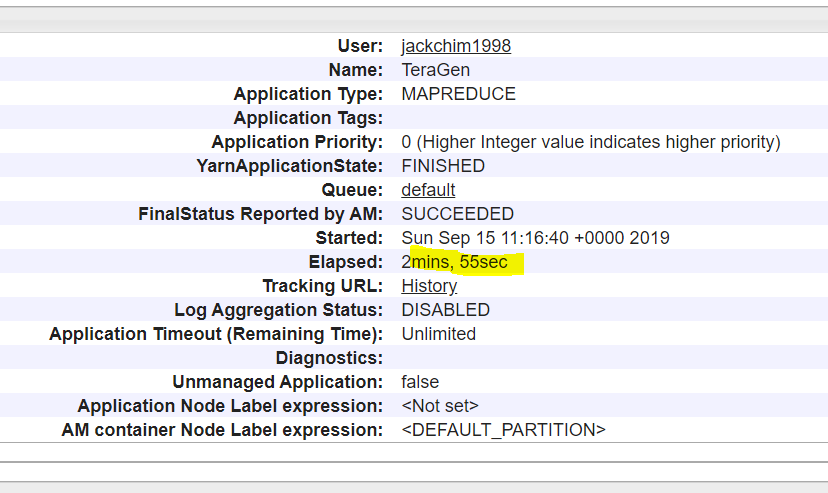
Terasort 2GB data with 2 mins and 9 sec



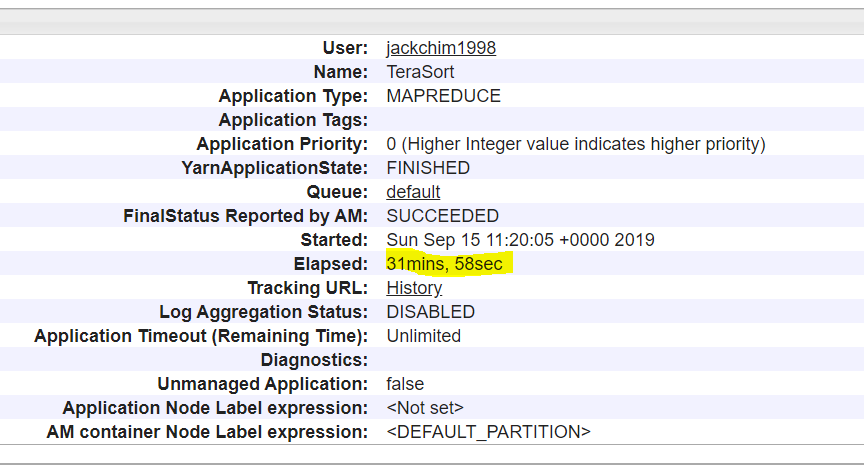
We teravalidate 2 GB data with 1 mins 25 sec



We teragen 20GB data with 2min 55 sec



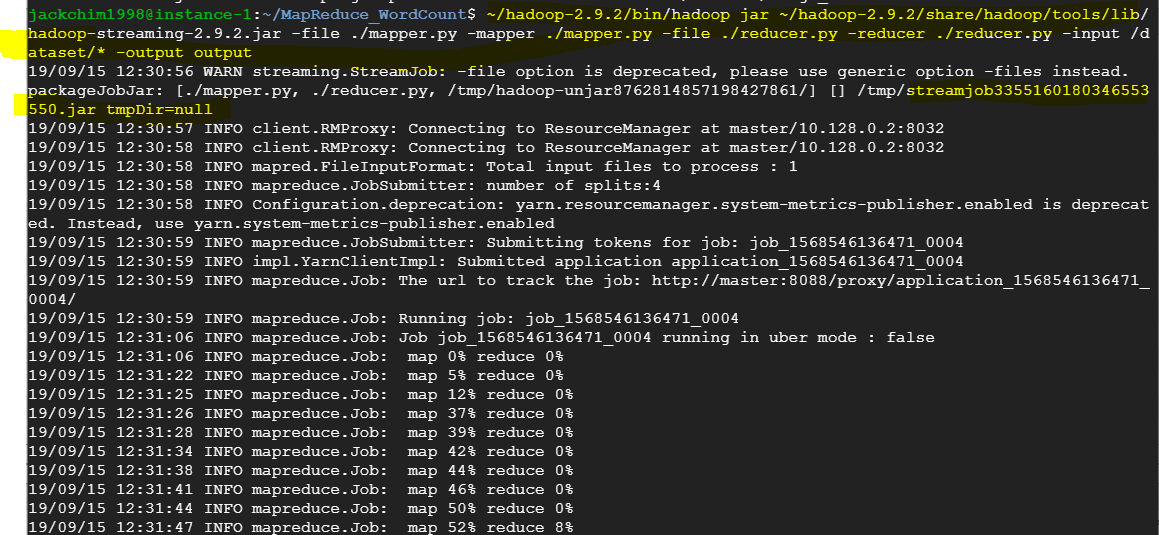
Terasort 20GB data with 31 mins and 58 sec



It is time-consuming, so that I would not do TearValidate.

[40 marks] Running the Python Code on Hadoop

The below screen is the command, we can see the job name is streamjob3355160……….



The below screen shows that it elapsed 3mins 46sec.

