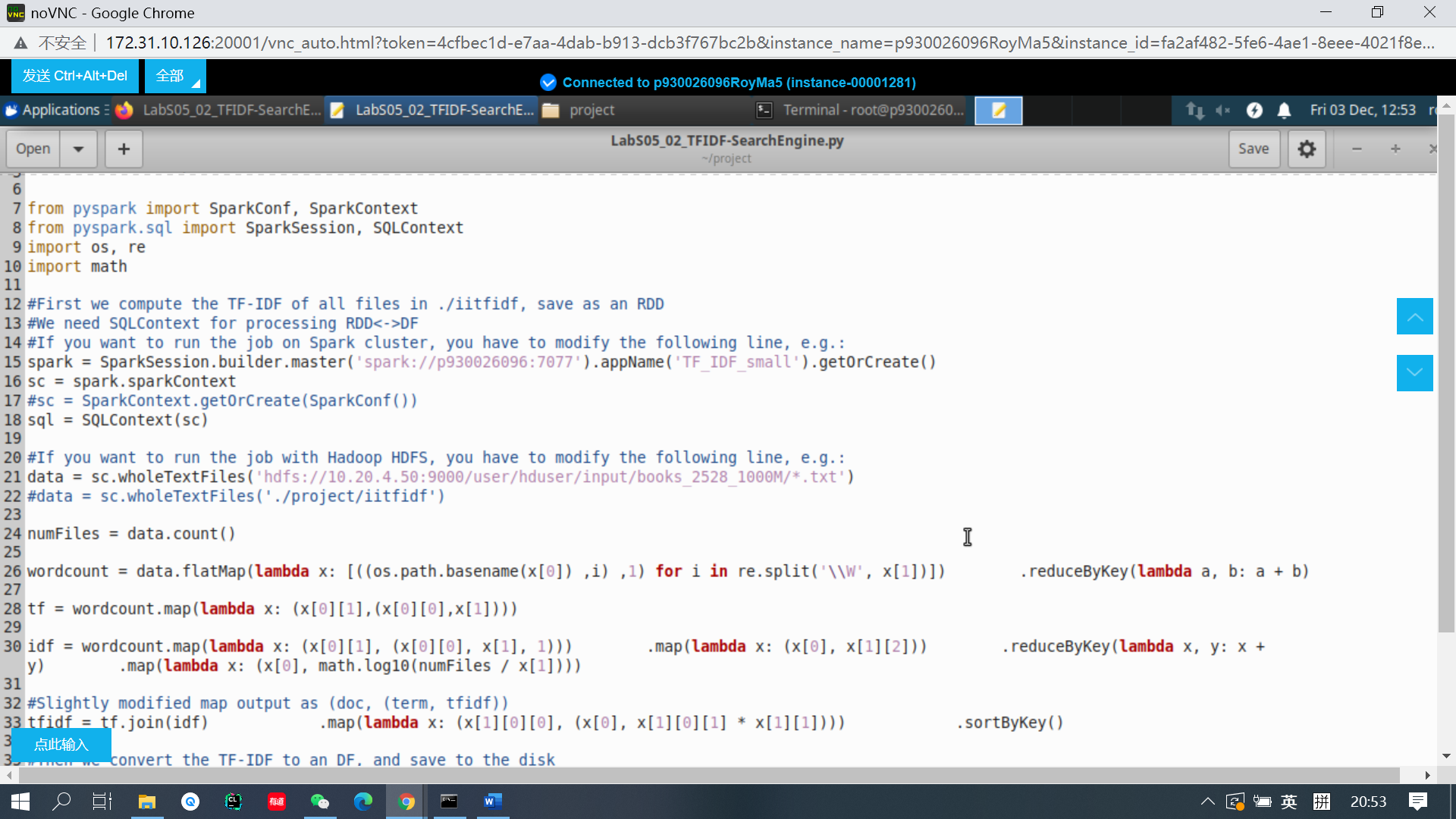
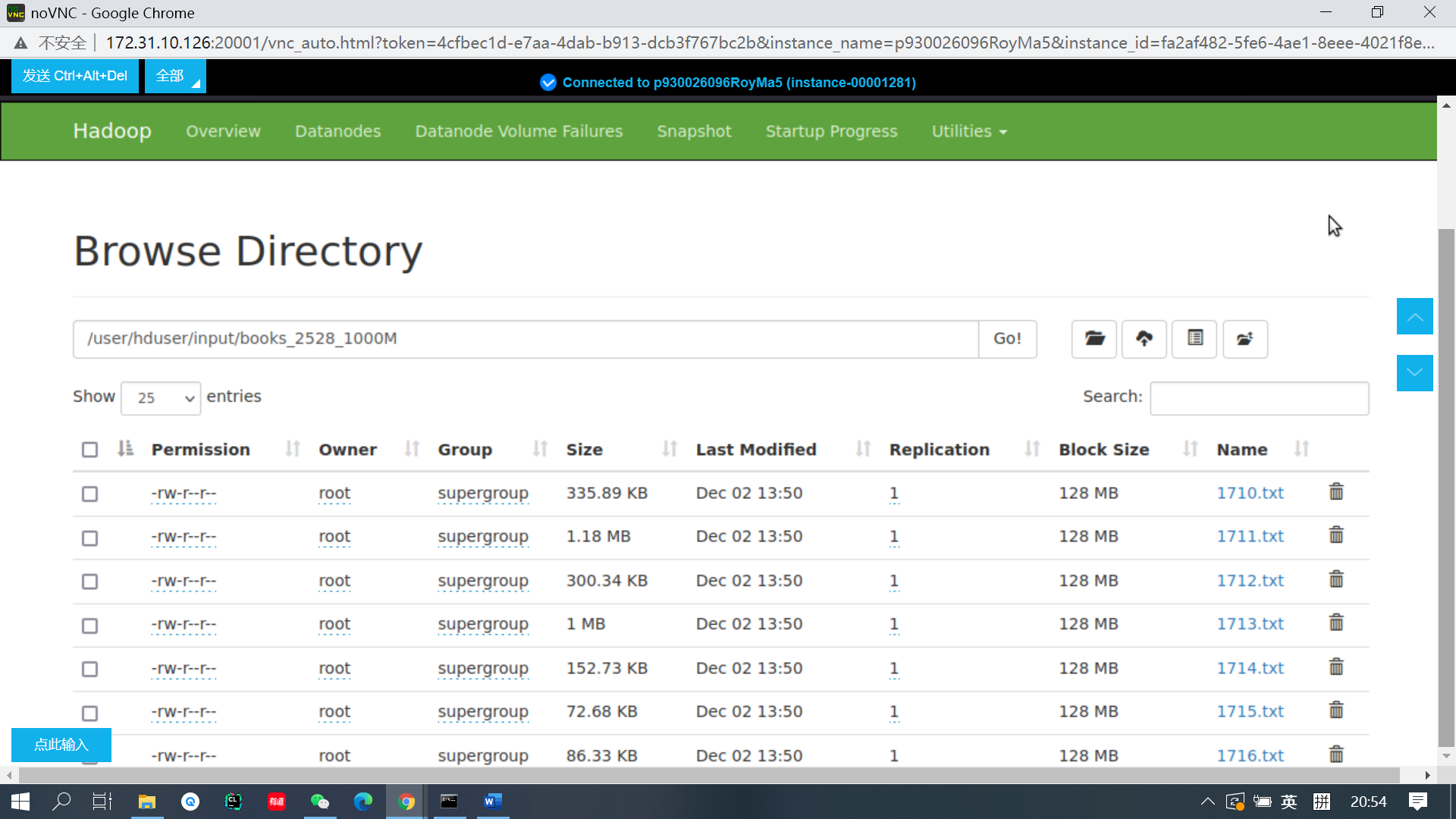
**Bonus**

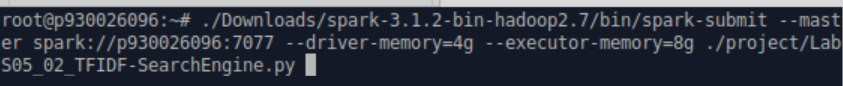
**Bonus(1)**



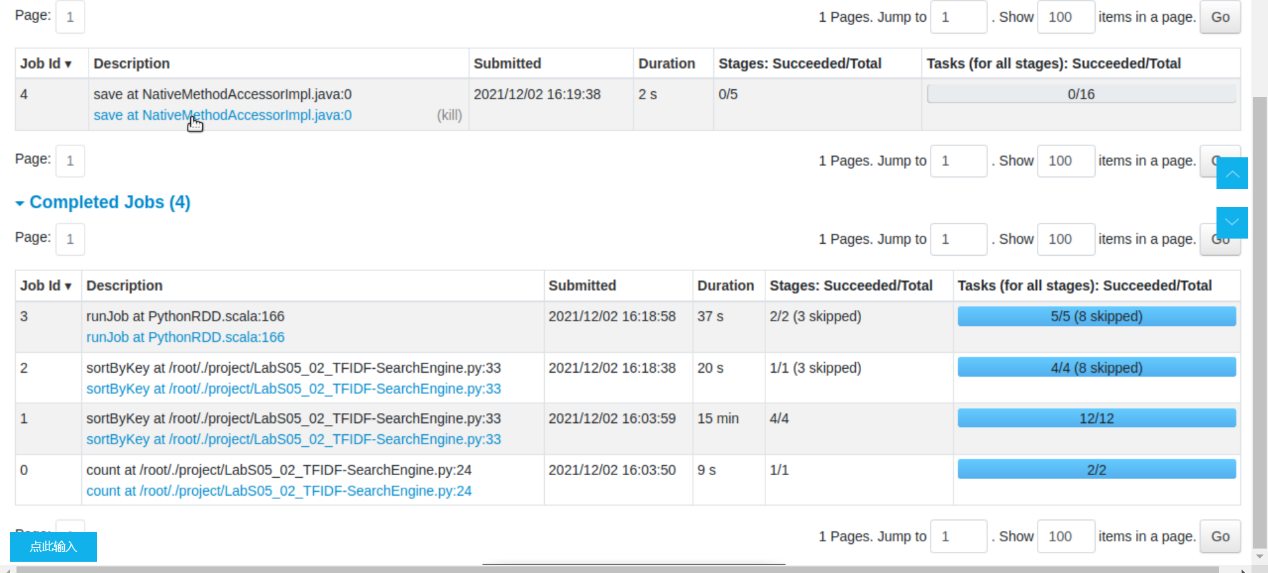
**Run a large file (1G) from our HDFS, since there was an error when I ran the code in Jupyter notebook, so, I submit the task from the console.**



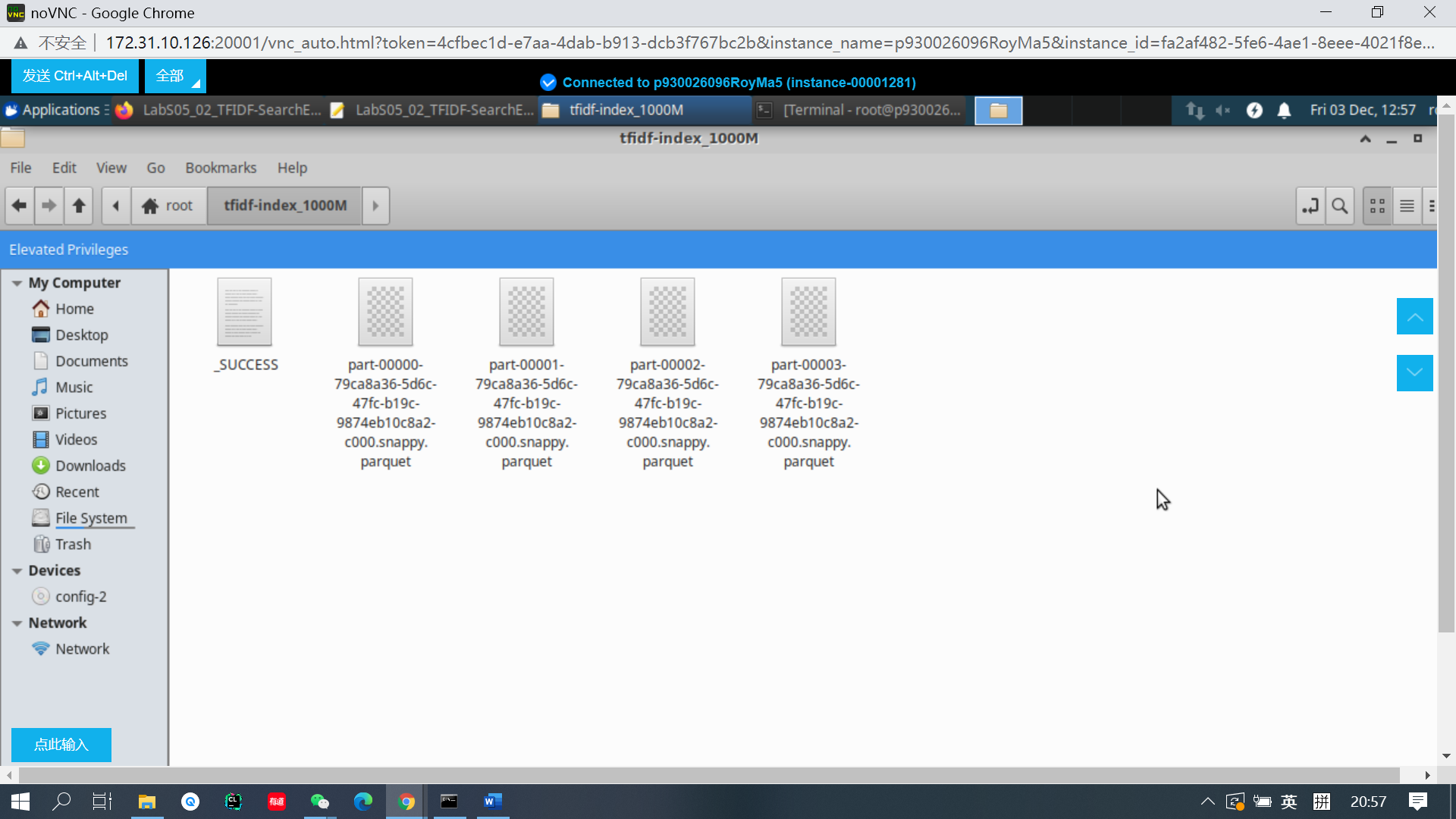
**Here is a screen shout of my file directory in HDFS which contains more than 1G.**



**Submit the file in the console by setting the driver memory and executor memory.**

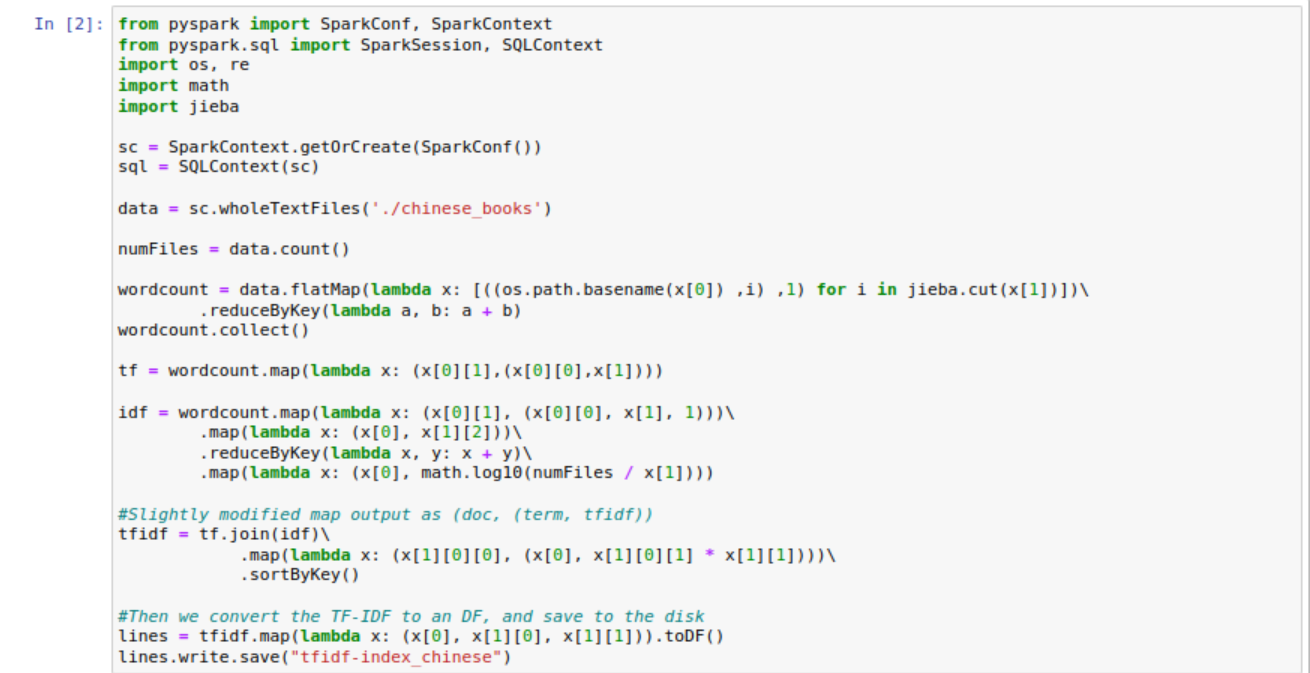


**The result of task. It takes approximately 16 mins.**

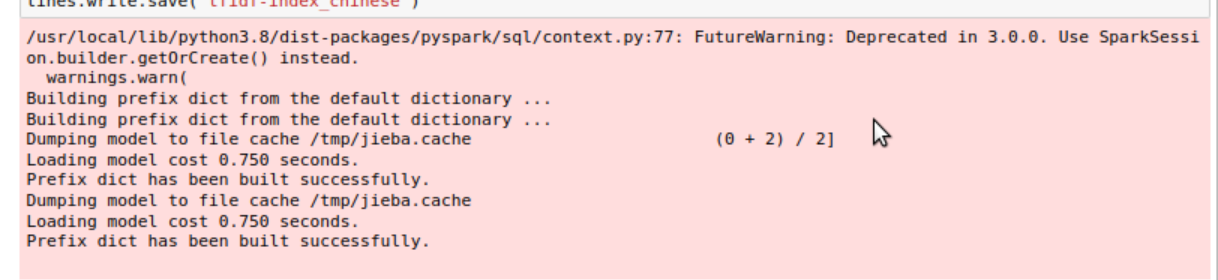


**Here is the output files of the codes.**

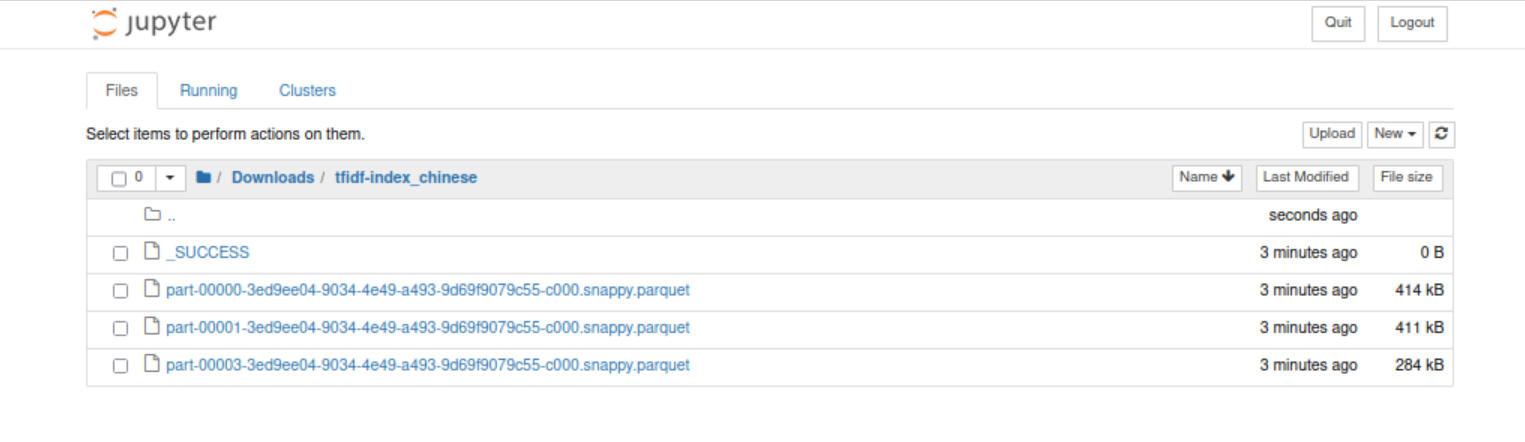
**Bonus（2）**



**Run the code based on some Chinese books by using the package jieba.**



**The program is running.**



**The files that generate by the codes.**