云计算实验三报告

作业 1：

```
[root@uhadoop-iyqgxwod-master1 ~]# su hadoop
[hadoop@uhadoop-iyqgxwod-master1 root]$ hdfs dfsadmin -report
Configured Capacity: 23813794778317 (21.66 TB)
Present Capacity: 22613509967053 (20.57 TB)
DFS Remaining: 22613509881037 (20.57 TB)
DFS Used: 86016 (84 KB)
DFS Used%: 0.00%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (3):

Name: 10.23.205.110:50010 (uhadoop-iyqgxwod-core3)
Hostname: uhadoop-iyqgxwod-core3
Decommission Status : Normal
Configured Capacity: 7937949069039 (7.22 TB)
DFS Used: 28672 (28 KB)
Non DFS Used: 0 (0 B)
DFS Remaining: 7537854103279 (6.86 TB)
DFS Used%: 0.00%
DFS Remaining%: 94.96%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
```

```
Name: 10.23.57.249:50010 (uhadoop-iyqgxwod-core1)
Hostname: uhadoop-iyqgxwod-core1
Decommission Status : Normal
Configured Capacity: 7937896640239 (7.22 TB)
DFS Used: 28672 (28 KB)
Non DFS Used: 0 (0 B)
DFS Remaining: 7537801674479 (6.86 TB)
DFS Used%: 0.00%
DFS Remaining%: 94.96%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 2
Last contact: Sat Oct 17 15:24:14 CST 2020


Name: 10.23.58.26:50010 (uhadoop-iyqgxwod-core2)
Hostname: uhadoop-iyqgxwod-core2
Decommission Status : Normal
Configured Capacity: 7937949069039 (7.22 TB)
DFS Used: 28672 (28 KB)
Non DFS Used: 0 (0 B)
DFS Remaining: 7537854103279 (6.86 TB)
DFS Used%: 0.00%
DFS Remaining%: 94.96%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 2
Last contact: Sat Oct 17 15:24:14 CST 2020
```

作业 2：

```
[root@uhadoop-iw0sqedh-master1 ~]# ls
info.txt  install-java.sh
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -test -e info.txt
[root@uhadoop-iw0sqedh-master1 ~]# ls
info.txt  install-java.sh
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -mkdir /test
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -put info.txt /test
```

```
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -ls /test
Found 1 items
-rw-r--r--   3 root supergroup         41 2020-10-19 19:13 /test/info.txt
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -tail /test/info.txt
云计算实验课:10185501402-孙秋实
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -cat /test/info.txt
云计算实验课:10185501402-孙秋实
[root@uhadoop-iw0sqedh-master1 ~]#
```

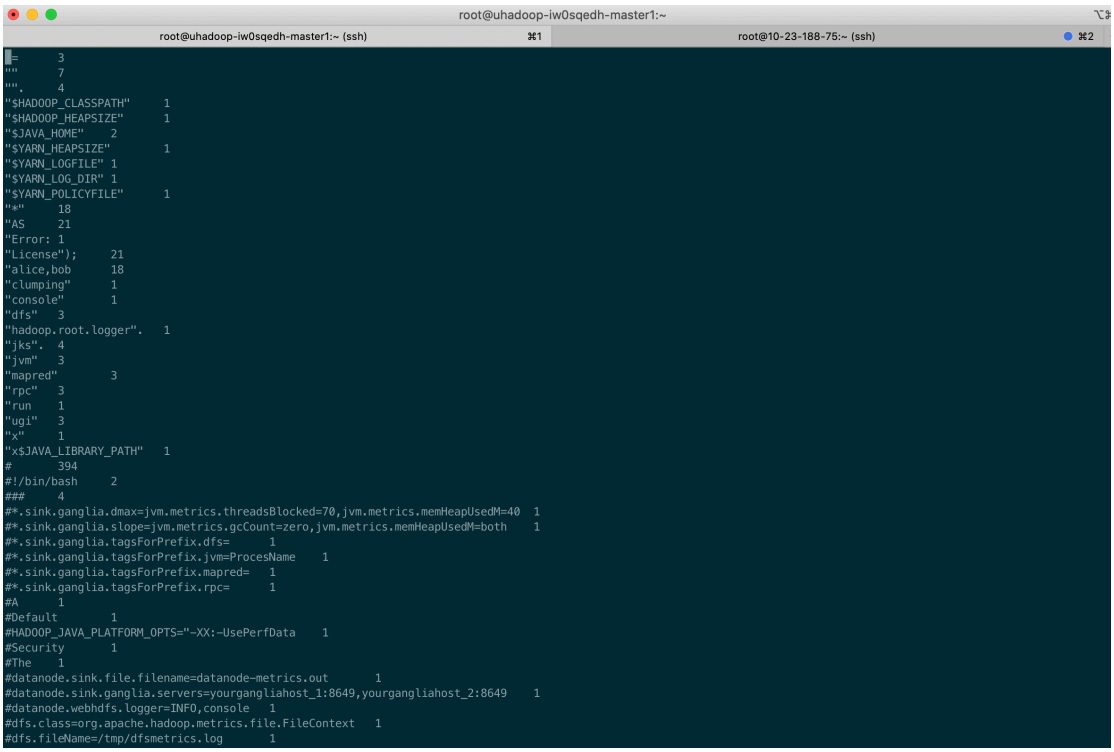
作业 3：

```
 ✘ sunqiushi@sunqiushideMacBook-Pro  ~  ssh root@106.75.229.238
The authenticity of host '106.75.229.238 (106.75.229.238)' can't be established.
ECDSA key fingerprint is SHA256:7bKxlPi4gy1b7lfjii4bZKd7VSn3iNKVxelj5tATW9g.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '106.75.229.238' (ECDSA) to the list of known hosts.
root@106.75.229.238's password:
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[root@10-23-188-75 ~]# source ~/.bashrc
[root@10-23-188-75 ~]# hadoop fs -ls /test
Found 1 items
-rw-r--r--   3 root supergroup         41 2020-10-19 19:13 /test/info.txt
[root@10-23-188-75 ~]# hadoop fs -rm -r -skipTrash /test
Deleted /test
[root@10-23-188-75 ~]# hadoop fs -ls /test
ls: `/test': No such file or directory
[root@10-23-188-75 ~]#
```

作业 4：

```
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -mkdir /myinput
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -put /home/hadoop/etc/hadoop/* /myinput
[root@uhadoop-iw0sqedh-master1 ~]# hadoop jar /home/hadoop/hadoop-examples.jar wordcount /myinput /myoutput
```

```
[root@uhadoop-iw0sqedh-master1 ~]# hadoop fs -cat /myoutput/part-r-00000 > mapreduce_result.txt
[root@uhadoop-iw0sqedh-master1 ~]#
```



作业 5：

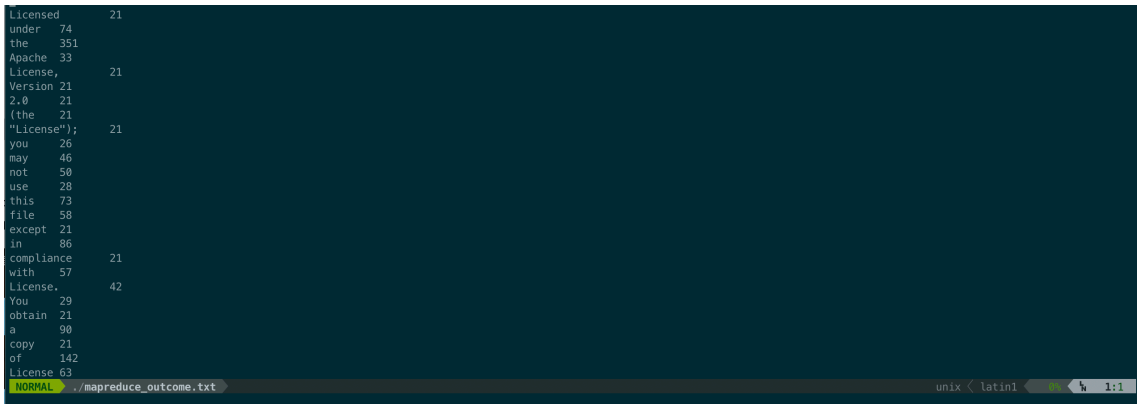
Python 实现单线程 Map-Reduce 的代码如下

```
# Cloud-Computing 2020
# Author:QiushiSun
# Lab 3 Mapreduce:Mapper.py
import sys
import time
time_start = time.time() # start

for each_line in sys.stdin: #stream data
    each_line = each_line.strip() #stripping
    words = each_line.split()  #stripping
    for word in words:
        print("%s\t%s" % (word, 1)) # mapper

time_end = time.time()    #end
time_c= time_end - time_start   # get time
print('mapper time cost', time_c, 's')  # delete it when reducing

# Cloud-Computing 2020
# Author:QiushiSun
# Lab 3 Mapreduce:Mapper.py
import sys
import time
```

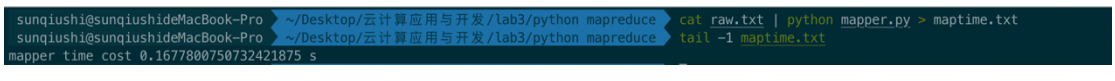
```
23. time_start = time.time() # start
24. mr_dic = {} #init
25. for each_line in sys.stdin:# stream data
26.     each_line = each_line.strip()
27.     word,counting = each_line.split('\t') #seperate word and count
28.
29.     try: # string -> integer
30.         counting = int(counting)
31.     except ValueError:
32.         continue
33.
34.     if word in mr_dic:
35.         mr_dic[word] += 1 #already in dic->++
36.     else:
37.         mr_dic.setdefault(word, 1) #new dict
38.
39. for key, value in mr_dic.items():# print outcome
40.     print('%s\t%s' % (key, value))
41.
42. time_end = time.time()    #end
43. time_c= time_end - time_start   #cost
44. print('reduce time cost', time_c, 's')
```

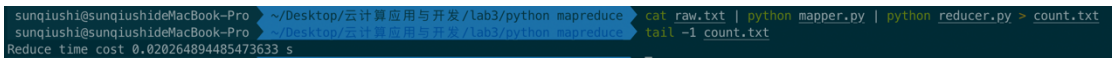
在命令行中先对 hadoop 文件夹中的文件进行预处理，然后用于这个手写的 MapReduce 进行测试

查看 MapReduce 的结果：



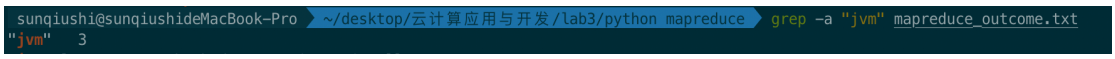
所耗时间：



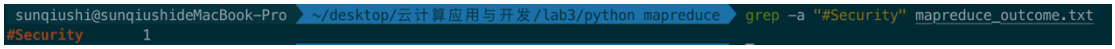


这个手写 MapReduce 执行的总时间约为 0.188 秒

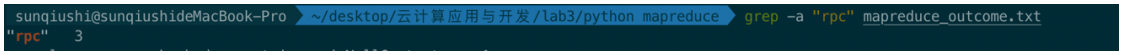
最后对比一下 Q4 中 MapReduce 结果与这个单线程 MapReduce 的结果是否相同：
Test1:



Test2:



Test3:



如果您愿意把这三个结果和问题 4 中的结果对比的话，会发现它们是一样的。