

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 ~]#
```

```
root@uhadoop-1w0sqedh-master1:~ (ssh)
root@uhadoop-1w0sqedh-master1:~
root@10-23-188-75:~ (ssh)

# 3
### 7
### 4
"SHADOOP_CLASSPATH" 1
"SHADOOP_HEAPSIZE" 1
"$JAVA_HOME" 2
"SYARN_HEAPSIZE" 1
"SYARN_LOGFILE" 1
"SYARN_LOG_DIR" 1
"SYARN_POLICYFILE" 1
"AS" 18
"AS" 21
"Error: 1
"License"); 21
"alice,bob" 18
"clumping" 1
"console" 1
"dfs" 3
"hadoop.root.logger", 1
"jks", 4
"jvm" 3
"mapred" 3
"rpc" 3
"run" 1
"ugi" 3
"u" 1
"$JAVA_LIBRARY_PATH" 1
# 394
#!/bin/bash 2
### 4
##.sink.ganglia.dmax=jvm.metrics.threadsBlocked=70,jvm.metrics.memHeapUsedM=40 1
##.sink.ganglia.slope=jvm.metrics.gcCount=zero,jvm.metrics.memHeapUsedM=both 1
##.sink.ganglia.tagsForPrefix.dfs= 1
##.sink.ganglia.tagsForPrefix.jvm=ProcesName 1
##.sink.ganglia.tagsForPrefix.mapred= 1
##.sink.ganglia.tagsForPrefix.rpc= 1
#A 1
#Default 1
#HADOOP_JAVA_PLATFORM_OPTS="-XX:-UsePerfData 1
#Security 1
#The 1
#datanode.sink.file.filename=datanode-metrics.out 1
#datanode.sink.ganglia.server=yourgangliahost_1:8649,yourgangliahost_2:8649 1
#datanode.webdfs.logger=INFO,console 1
#dfs.class=org.apache.hadoop.metrics.file.FileContext 1
#dfs.fileName=/tmp/dfsmetrics.log 1
```

作业 5：

Python 实现单线程 Map-Reduce 的代码如下

```
1. # Cloud-Computing 2020
2. # Author:QiushiSun
3. # Lab 3 Mapreduce:Mapper.py
4. import sys
5. import time
6. time_start = time.time() # start
7.
8. for each_line in sys.stdin: #stream data
9.     each_line = each_line.strip() #stripping
10.    words = each_line.split() #stripping
11.    for word in words:
12.        print("%s\t%s" % (word, 1)) # mapper
13.
14. time_end = time.time() #end
15. time_c= time_end - time_start # get time
16. print('mapper time cost', time_c, 's') # delete it when reducing
17.
18. # Cloud-Computing 2020
19. # Author:QiushiSun
20. # Lab 3 Mapreduce:Mapper.py
21. import sys
22. 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

```

sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > cat hadoop/* > raw.txt
sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > cat raw.txt | python mapper.py | python reducer.py > mapreduce_outcome.txt
sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce >

```

查看 MapReduce 的结果：

```

Licensed      21
under       74
the        351
Apache      33
License,     21
Version 2.1
2.0         21
(the        21
"License");  21
you         26
may         46
not         50
use         28
this        73
file        58
except      21
in          86
compliance  21
with        57
License.    42
You         29
obtain      21
a           90
copy        21
of          142
License     63
NORMAL  mapreduce_outcome.txt
unix < latin1 1:1

```

所耗时间：

```

sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > cat raw.txt | python mapper.py > maptime.txt
sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > tail -1 maptime.txt
mapper time cost 0.1677800750732421875 s

sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > cat raw.txt | python mapper.py | python reducer.py > count.txt
sunqiushi@sunqiushideMacBook-Pro ~/Desktop/云计算应用与开发/lab3/python mapreduce > tail -1 count.txt
Reduce time cost 0.020264894485473633 s

```

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:

```
sunqiushi@sunqiushideMacBook-Pro ~/desktop/云计算应用与开发/lab3/python mapreduce > grep -a "jvm" mapreduce_outcome.txt
"jvm" 3
```

Test2:

```
sunqiushi@sunqiushideMacBook-Pro ~/desktop/云计算应用与开发/lab3/python mapreduce > grep -a "#Security" mapreduce_outcome.txt
#Security 1
```

Test3:

```
sunqiushi@sunqiushideMacBook-Pro ~/desktop/云计算应用与开发/lab3/python mapreduce > grep -a "rpc" mapreduce_outcome.txt
"rpc" 3
```

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。