

## 云计算实验三报告

作业 1:

```
[root@uhadoop-3k3ptjdu-master1 ~]# su hadoop
[hadoop@uhadoop-3k3ptjdu-master1 root]$ hdfs dfsadmin -report
Configured Capacity: 23813794778317 (21.66 TB)
Present Capacity: 22613509967053 (20.57 TB)
DFS Remaining: 22613509893325 (20.57 TB)
DFS Used: 73728 (72 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: 10.23.116.205:50010 (uhadoop-3k3ptjdu-core2)
Hostname: uhadoop-3k3ptjdu-core2
Decommission Status : Normal
Configured Capacity: 7937949069039 (7.22 TB)
DFS Used: 24576 (24 KB)
Non DFS Used: 0 (0 B)
DFS Remaining: 7537854107375 (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: Tue Oct 20 13:23:42 CST 2020

Name: 10.23.180.73:50010 (uhadoop-3k3ptjdu-core3)
Hostname: uhadoop-3k3ptjdu-core3
Decommission Status : Normal
Configured Capacity: 7937949069039 (7.22 TB)
DFS Used: 24576 (24 KB)
Non DFS Used: 0 (0 B)
DFS Remaining: 7537854107375 (6.86 TB)
DFS Used%: 0.00%
```

作业 2:

```
[root@uhadoop-3k3ptjdu-master1 ~]# hadoop fs -cat /test/info.txt
云计算实验课:10195501425-张雯怡
```

### 作业 3:

```
[root@10-23-65-212 ~]# hadoop fs -rm /test/info.txt
20/10/20 14:04:25 INFO fs.TrashPolicyDefault: Moved: 'hdfs://Ucluster/test/info.txt'
to trash at: hdfs://Ucluster/user/root/.Trash/Current/test/info.txt
[root@10-23-65-212 ~]# hadoop fs -ls /test

[root@10-23-65-212 ~]# hadoop fs -rmdir /test
[root@10-23-65-212 ~]# hadoop fs -ls /
Found 4 items
drwxrwxrwt - hadoop supergroup          0 2020-10-20 13:15 /hbase
drwxrwxrwt - hadoop supergroup          0 2020-10-20 13:17 /tmp
drwxrwxrwt - hadoop supergroup          0 2020-10-20 13:27 /user
drwxrwxrwt - hadoop supergroup          0 2020-10-20 13:15 /var
```

### 作业 4:

```
values 6
variable 4
variables 4
version 1
version="1.0" 7
version="1.0"> 1
version="1.0"?> 6
via 3
view, 1
viewing 2
w/ 1
want 2
warnings 1
warnings. 1
weak 1
when 10
where 4
which 7
while 1
who 6
will 23
window 1
window, 1
with 57
within 4
without 1
work 11
writing, 21
written 2
xmlns:xsl="http://www.w3.org/1999/XSL/Transform" 1
yarn.ewma.cleanupInterval=300 1
yarn.ewma.maxUniqueMessages=250 1
yarn.ewma.messageAgeLimitSeconds=86400 1
yarn.nodemanager.linux-container-executor.group 1
yarn.nodemanager.linux-container-executor.group=#configured 1
yarn.server.resourcemanager.appsummary.log.file 1
yarn.server.resourcemanager.appsummary.log.file=rm-appsummary.log 1
yarn.server.resourcemanager.appsummary.logger 2
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.logger} 1
you 26
zookeeper 1
[root@uhadoop-bgeulxao-master1 ~]#
```

## 作业 5:

```
#print(words)
ft.close()
fo.close()

fn = open('C:\\Users\\86137\\file_list.txt','rt') # 打开文件
words = fn.read() # 读出整个文件
fn.close() # 关闭文件

wordlist = []

words = re.sub('[\s+|!/,./%&*@+\\"{}]|(?![\.\-\_\~\#\&\$])+=','', words) #去除标点

if re.match(r'[a-zA-Z0-9]*', words): #避免中文影响

    wordlist.extend(words.strip().split())

Count = collections.Counter(wordlist) #统计词频

end_time=time()
run_time=end_time-begin_time

print('运行时间:', run_time)
print(collections.Counter(wordlist))

运行时间: 0.05873270988464355
Counter({'the': 373, 'log': 205, 'name': 204, '<property>': 201, 'value': 201, 'property': 201, 'to': 170, 'License': 147, 'for': 146, 'of': 142, 'append': 128, 'log': 126, 'apache': 124, 'is': 123, 'YARN': 122, 'HADOOP': 119, 'hadoop': 113, 'The': 102, 'and': 95, 'file': 92, 'a': 89, 'in': 87, 'by': 87, 'description': 85, 'export': 78, 'cnameyar': 77, 'under': 74, 'this': 73, 'rem': 73, 'o': 71, 'or': 67, '<value>': 67, 'set': 66, 'with': 57, 'distributed': 55, 'user': 53, 'not': 50, 'be': 50, 'org': 49, 'OPTS': 48, 'default': 47, 'have': 47, 'may': 46, 'users': 45, 'are': 45, 'list': 44, 'description': 42, 'OPTS': 41, 'See': 40, 'used': 40, 'resourceManager': 39, 'ACL': 38, 'logger': 38, 'group': 3, '7, 'http': 37, 'svr': 36, 'hadoop': 36, 'sink': 35, 'use': 34, 'if': 34, 'metrics': 34, 'ganglia': 34, 'Apache': 33, 'licenses': 32, 'JAVA': 32, 'HOME': 32, '-': 31, '--': 3, 'protocol': 31, 'server': 31, 'on': 30, 'all': 30, 'mapred': 30, 'you': 29, 'an': 29, 'at': 28, 'escher': 28, 'specific': 28, '<namedfs>': 28, 'yarn': 27, 'KMS': 27, 'you': 26, 'req': 26, 'For': 26, 'scheduler': 25, 'nodeManager': 25, 'defined': 25, 'classorg': 25, 'queue': 24, 'audit': 24, 'm': 24, 'A': 23, 'will': 23, 'http': 22, 'www': 22, 'that': 22, 'e': 22, 'specified': 22, 'allowed': 22, 'eckl': 22, 'licensed': 21, 'Version': 21, '2': 21, 'j': 21, 'except': 21, 'compliance': 21, 'obtain': 21, 'copy': 21, 'LICENSE -2': 21, 'Unless': 21, 'applicable': 21, 'law': 21, 'agreed': 21, 'writing': 21, 'software': 21, 'AS': 21, 'IS': 21, 'BASIS': 21, 'WITHOUT': 21, 'WARRANTIES': 21, 'OR': 21, 'CONDITION
```