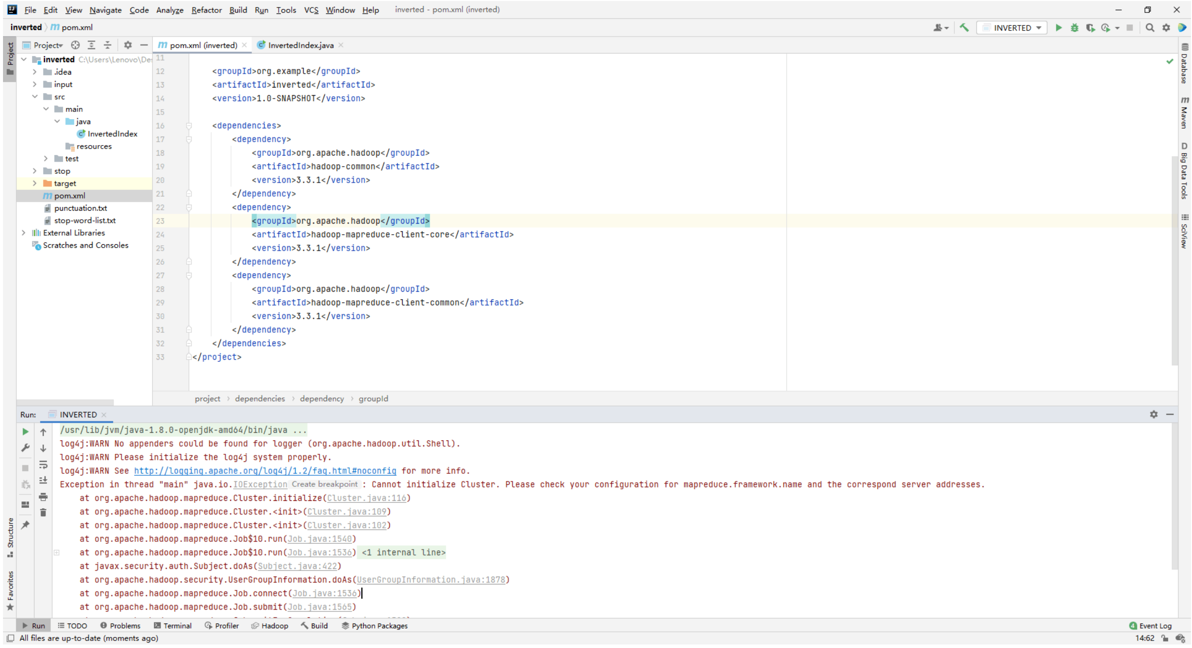
跑样例代码调整好环境



```xml
    <groupId>org.example</groupId>
    <artifactId>inverted</artifactId>
    <version>1.0-SNAPSHOT</version>

    <dependencies>
        <dependency>
            <groupId>org.apache.hadoop</groupId>
            <artifactId>hadoop-common</artifactId>
            <version>3.3.1</version>
        </dependency>
        <dependency>
            <groupId>org.apache.hadoop</groupId>
            <artifactId>hadoop-mapreduce-client-core</artifactId>
            <version>3.3.1</version>
        </dependency>
        <dependency>
            <groupId>org.apache.hadoop</groupId>
            <artifactId>hadoop-mapreduce-client-common</artifactId>
            <version>3.3.1</version>
        </dependency>
    </dependencies>
</project>
```

```
/usr/lib/jvm/java-1.8.0-openjdk-amd64/bin/java ...
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Exception in thread "main" java.io.IOException Create breakpoint : Cannot initialize Cluster. Please check your configuration for mapreduce.framework.name and the correspond server addresses.
	at org.apache.hadoop.mapreduce.Cluster.initialize(Cluster.java:116)
	at org.apache.hadoop.mapreduce.Cluster.<init>(Cluster.java:109)
	at org.apache.hadoop.mapreduce.Cluster.<init>(Cluster.java:102)
	at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1540)
	at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1536) <1 internal line>
	at javax.security.auth.Subject.doAs(Subject.java:422)
	at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
	at org.apache.hadoop.mapreduce.Job.connect(Job.java:1536)
	at org.apache.hadoop.mapreduce.Job.submit(Job.java:1565)
```

首先调试环境，发现idea无法初始化集群，查阅资料(21条消息) Hadoop 解决本地运行出错Cannot initialize Cluster. Please check your configuration for mapreduce.framework... Pineapple的博客-CSDN博客发现，hadoop-mapreduce-client-core.jar是支持放在集群上运行的，hadoop-mapreduce-client-common.jar是支持在本地运行的，应该加上需要的依赖，补充maven配置文件。配置好后正常运行。参考代码：Hadoop MapReduce: 带词频属性的文档倒排索引 - Penguin (polarxiong.com)

代码正文介绍

运行命令

`bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.output.textoutputformat.separator=: input/ output/ -skip punctuation.txt -skip stop-word-list.txt`

这次有经验，对代码构架更熟悉，写的代码更加简洁干净，文件参数均为实际路径。

具体实现：

mapper：对输入的Text切分为多个word,每个word作为一个key输出

- 输入：key:当前行偏移位置, value:当前行内容；
- 输出：key:word#filename, value:1

combiner：将Mapper输出的中间结果相同key部分的value累加，减少向Reduce节点传输的数据量

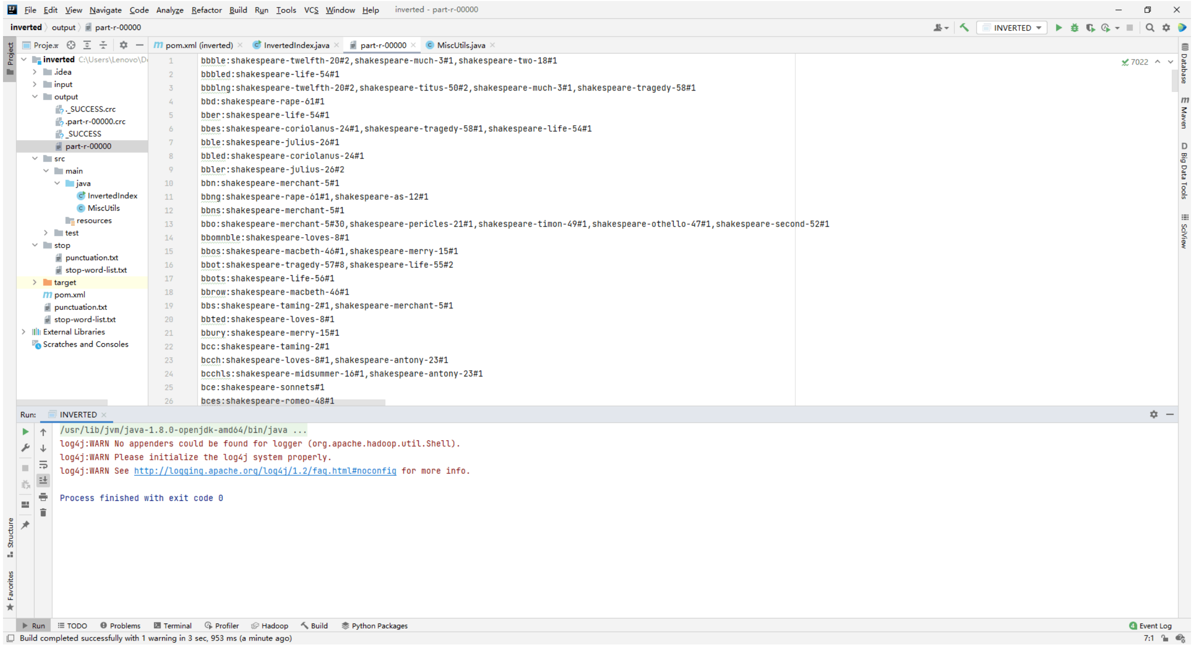
- 输出：key:word#filename, value:累加和

partitioner：为了将同一个word的键值对发送到同一个Reduce节点，对key进行临时处理，将原key的(word, filename)临时拆开，使Partitioner只按照word值进行选择Reduce节点
reduce：利用每个Reducer接收到的键值对中，是按照同一个word排好序一起出现的，将相同的word对应的(word,filename)拆分开，将filename与累加和构成键值对存储到临时Hashmap中。当累加出现到word不同时，将前面的结果利用MiscUtils排好序再写入临时StringBuilder中来写入文件即完成了旧word的统计，则将新word作为key再次开始。

- 输入：key:word#filename, value:[NUM,NUM,...]

- 输出：key:word, value:filename#NUM,filename#NUM,...

cleanup：　上述reduce()只会在遇到新word时，处理并输出前一个word，故最后一个word还需要额外的处理。重载cleanup()，处理最后一个word并输出。



```
bbble:shakespeare-twelfth-20#2,shakespeare-much-3#1,shakespeare-two-18#1
bbbled:shakespeare-life-54#1
bbblng:shakespeare-twelfth-20#2,shakespeare-titus-50#2,shakespeare-much-3#1,shakespeare-tragedy-58#1
bbd:shakespeare-rape-61#1
bber:shakespeare-life-54#1
bbes:shakespeare-coriolanus-24#1,shakespeare-tragedy-58#1,shakespeare-life-54#1
bble:shakespeare-julius-26#1
bbled:shakespeare-coriolanus-24#1
bbler:shakespeare-julius-26#2
bbn:shakespeare-merchant-5#1
bbng:shakespeare-rape-61#1,shakespeare-as-12#1
bbns:shakespeare-merchant-5#1
bbo:shakespeare-merchant-5#30,shakespeare-pericles-21#1,shakespeare-timon-49#1,shakespeare-othello-47#1,shakespeare-second-52#1
bbomnble:shakespeare-loves-8#1
bbos:shakespeare-macbeth-46#1,shakespeare-merry-15#1
bbot:shakespeare-tragedy-57#8,shakespeare-life-55#2
bbots:shakespeare-life-56#1
bbrow:shakespeare-macbeth-46#1
bbs:shakespeare-taming-2#1,shakespeare-merchant-5#1
bbted:shakespeare-loves-8#1
bbury:shakespeare-merry-15#1
bcc:shakespeare-taming-2#1
bcch:shakespeare-loves-8#1,shakespeare-antony-23#1
bcchls:shakespeare-midsummer-16#1,shakespeare-antony-23#1
bce:shakespeare-sonnets#1
bces:shakespeare-romeo-48#1
```

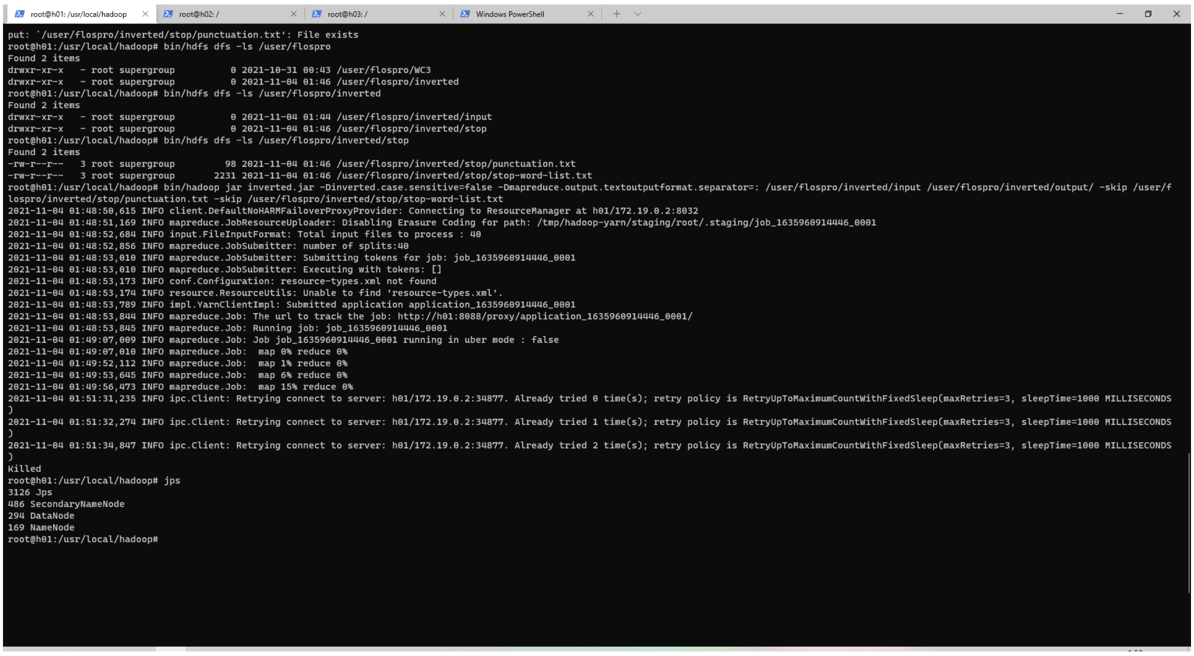
```
/usr/lib/jvm/java-1.8.0-openjdk-amd64/bin/java ...
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

Process finished with exit code 0
```

　　正确运行。

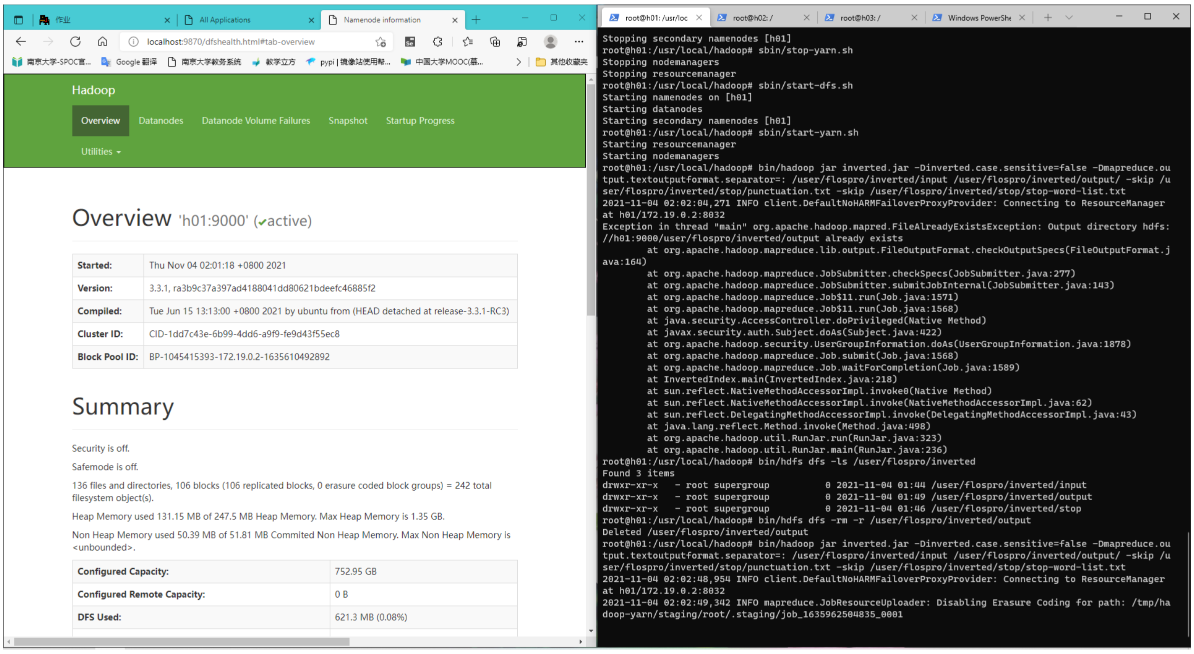
在docker集群中运行

　　迁移好文件后，执行命令：



```
put: `/user/flospro/inverted/stop/punctuation.txt': File exists
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro
Found 2 items
drwxr-xr-x   - root supergroup          0 2021-10-31 00:43 /user/flospro/WC3
drwxr-xr-x   - root supergroup          0 2021-11-04 01:46 /user/flospro/inverted
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro/inverted
Found 2 items
drwxr-xr-x   - root supergroup          0 2021-11-04 01:44 /user/flospro/inverted/input
drwxr-xr-x   - root supergroup          0 2021-11-04 01:46 /user/flospro/inverted/stop
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro/inverted/stop
Found 2 items
-rw-r--r--   3 root supergroup         98 2021-11-04 01:46 /user/flospro/inverted/stop/punctuation.txt
-rw-r--r--   3 root supergroup       2231 2021-11-04 01:46 /user/flospro/inverted/stop/stop-word-list.txt
root@h01:/usr/local/hadoop# bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.output.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /user/f
lospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 01:48:50,615 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at h01/172.19.0.2:8032
2021-11-04 01:48:51,169 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1635960914446_0001
2021-11-04 01:48:52,684 INFO input.FileInputFormat: Total input files to process : 40
2021-11-04 01:48:52,856 INFO mapreduce.JobSubmitter: number of splits:40
2021-11-04 01:48:53,010 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1635960914446_0001
2021-11-04 01:48:53,010 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-11-04 01:48:53,173 INFO conf.Configuration: resource-types.xml not found
2021-11-04 01:48:53,174 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-11-04 01:48:53,789 INFO impl.YarnClientImpl: Submitted application application_1635960914446_0001
2021-11-04 01:48:53,844 INFO mapreduce.Job: The url to track the job: http://h01:8088/proxy/application_1635960914446_0001/
2021-11-04 01:48:53,845 INFO mapreduce.Job: Running job: job_1635960914446_0001
2021-11-04 01:49:07,009 INFO mapreduce.Job: Job job_1635960914446_0001 running in uber mode : false
2021-11-04 01:49:07,010 INFO mapreduce.Job:  map 0% reduce 0%
2021-11-04 01:49:52,112 INFO mapreduce.Job:  map 1% reduce 0%
2021-11-04 01:49:53,645 INFO mapreduce.Job:  map 6% reduce 0%
2021-11-04 01:49:56,473 INFO mapreduce.Job:  map 15% reduce 0%
2021-11-04 01:51:31,235 INFO ipc.Client: Retrying connect to server: h01/172.19.0.2:34877. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS
)
2021-11-04 01:51:32,274 INFO ipc.Client: Retrying connect to server: h01/172.19.0.2:34877. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS
)
2021-11-04 01:51:34,847 INFO ipc.Client: Retrying connect to server: h01/172.19.0.2:34877. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS
)
Killed
root@h01:/usr/local/hadoop# jps
3126 Jps
486 SecondaryNameNode
294 DataNode
169 NameNode
root@h01:/usr/local/hadoop#
```

　　运行到15%之前一切正常，但是由于电脑承受不住运载，服务宕机了，任务被kill。从报错：连不上服务器和jps可以看出，确实是性能问题。



Overview 'h01:9000' (✓active)

Started:	Thu Nov 04 02:01:18 +0800 2021
Version:	3.3.1, ra3b9c37a397ad4188041dd80621bdeefc46885f2
Compiled:	Tue Jun 15 13:13:00 +0800 2021 by ubuntu from (HEAD detached at release-3.3.1-RC3)
Cluster ID:	CID-1dd7c43e-6b99-4dd6-a9f9-fe9d43f55ec8
Block Pool ID:	BP-1045415393-172.19.0.2-1635610492892

Summary

Security is off.

Safemode is off.

136 files and directories, 106 blocks (106 replicated blocks, 0 erasure coded block groups) = 242 total filesystem object(s).

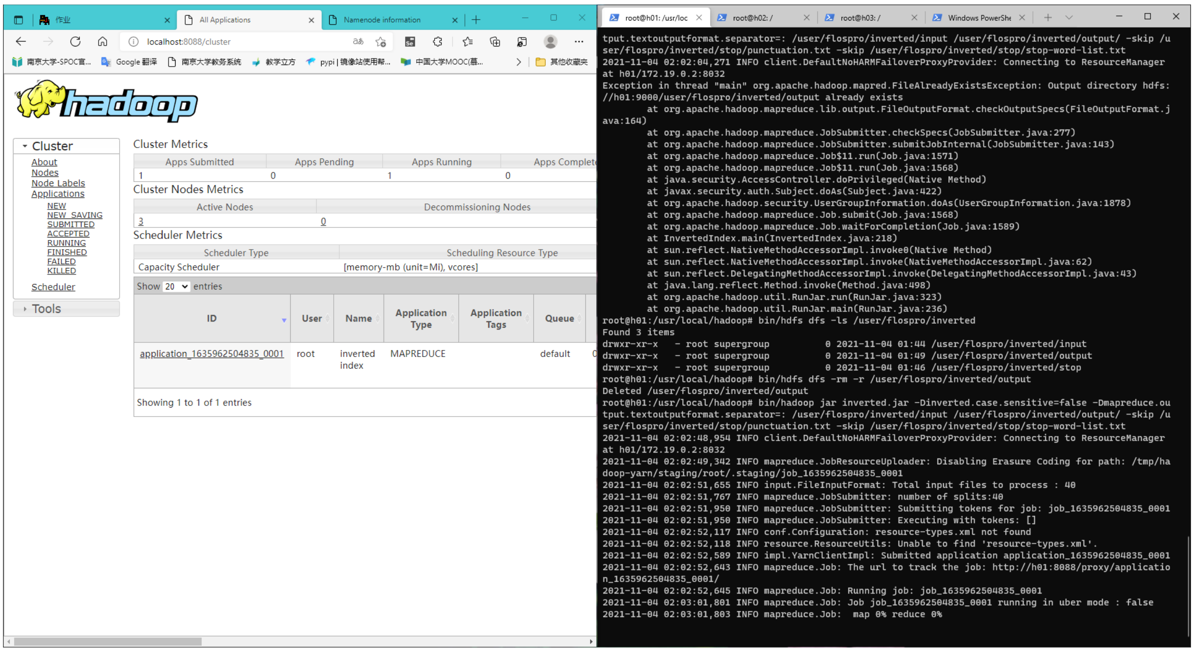
Heap Memory used 131.15 MB of 247.5 MB Heap Memory. Max Heap Memory is 1.35 GB.

Non Heap Memory used 50.39 MB of 51.81 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	752.95 GB
Configured Remote Capacity:	0 B
DFS Used:	621.3 MB (0.08%)

```
Stopping secondary namenodes [h01]
root@h01:/usr/local/hadoop# sbin/stop-yarn.sh
Stopping nodemanagers
Stopping resourcemanager
root@h01:/usr/local/hadoop# sbin/start-dfs.sh
Starting namenodes on [h01]
Starting datanodes
Starting secondary namenodes [h01]
root@h01:/usr/local/hadoop# sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
root@h01:/usr/local/hadoop# bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.ou
tput.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /u
ser/flospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 02:02:04,271 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager
at h01/172.19.0.2:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs:
//h01:9000/user/flospro/inverted/output already exists
        at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.j
ava:164)
        at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
        at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1571)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1568)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1568)
        at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1589)
        at InvertedIndex.main(InvertedIndex.java:218)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro/inverted
Found 3 items
drwxr-xr-x   - root supergroup          0 2021-11-04 01:44 /user/flospro/inverted/input
drwxr-xr-x   - root supergroup          0 2021-11-04 01:49 /user/flospro/inverted/output
drwxr-xr-x   - root supergroup          0 2021-11-04 01:46 /user/flospro/inverted/stop
root@h01:/usr/local/hadoop# bin/hdfs dfs -rm -r /user/flospro/inverted/output
Deleted /user/flospro/inverted/output
root@h01:/usr/local/hadoop# bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.ou
tput.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /u
ser/flospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 02:02:48,954 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager
at h01/172.19.0.2:8032
2021-11-04 02:02:49,342 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/ha
doop-yarn/staging/root/.staging/job_1635962504835_0001
```

再次尝试：



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Complete
1	0	1	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes
3	0

Scheduler Metrics

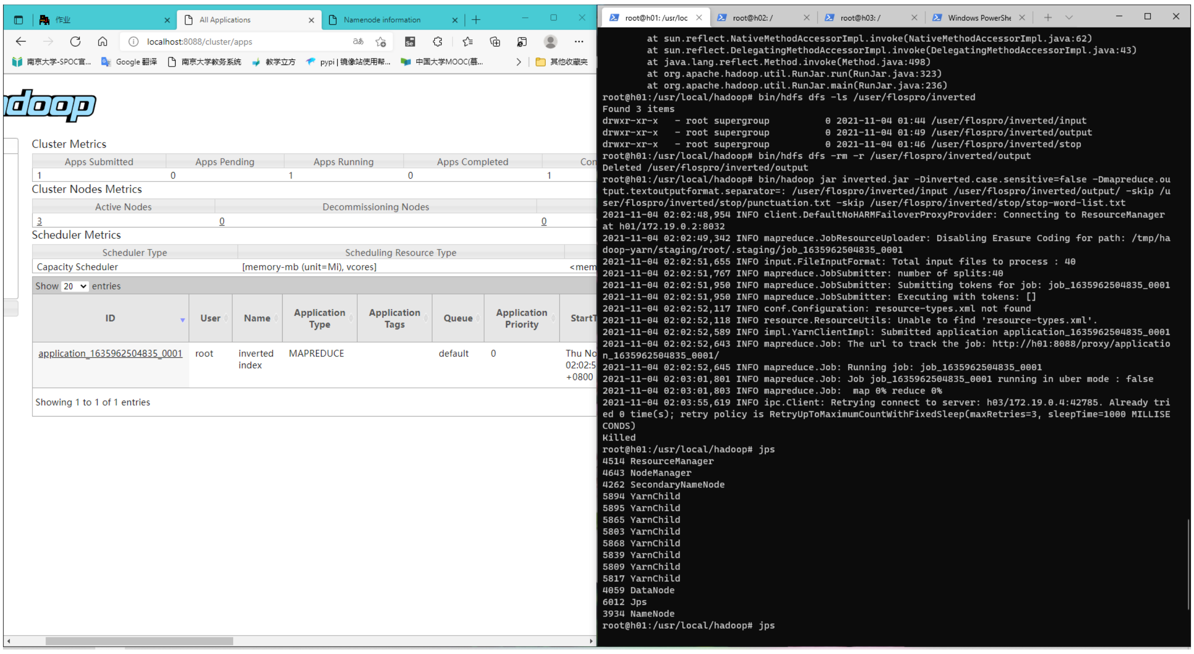
Scheduler Type	Scheduling Resource Type
Capacity Scheduler	[memory-mb (unit=Mi), vcores]

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue
application_1635962504835_0001	root	inverted index	MAPREDUCE		default

Showing 1 to 1 of 1 entries

```
tput.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /u
ser/flospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 02:02:04,271 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager
at h01/172.19.0.2:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs:
//h01:9000/user/flospro/inverted/output already exists
        at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.j
ava:164)
        at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:277)
        at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1571)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1568)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1878)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1568)
        at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1589)
        at InvertedIndex.main(InvertedIndex.java:218)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro/inverted
Found 3 items
drwxr-xr-x   - root supergroup          0 2021-11-04 01:44 /user/flospro/inverted/input
drwxr-xr-x   - root supergroup          0 2021-11-04 01:49 /user/flospro/inverted/output
drwxr-xr-x   - root supergroup          0 2021-11-04 01:46 /user/flospro/inverted/stop
root@h01:/usr/local/hadoop# bin/hdfs dfs -rm -r /user/flospro/inverted/output
Deleted /user/flospro/inverted/output
root@h01:/usr/local/hadoop# bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.ou
tput.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /u
ser/flospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 02:02:48,954 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager
at h01/172.19.0.2:8032
2021-11-04 02:02:49,342 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/ha
doop-yarn/staging/root/.staging/job_1635962504835_0001
2021-11-04 02:02:51,655 INFO input.FileInputFormat: Total input files to process : 40
2021-11-04 02:02:51,767 INFO mapreduce.JobSubmitter: number of splits:40
2021-11-04 02:02:51,950 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1635962504835_0001
2021-11-04 02:02:51,950 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-11-04 02:02:52,117 INFO conf.Configuration: resource-types.xml not found
2021-11-04 02:02:52,118 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-11-04 02:02:52,589 INFO impl.YarnClientImpl: Submitted application application_1635962504835_0001
2021-11-04 02:02:52,643 INFO mapreduce.Job: The url to track the job: http://h01:8088/proxy/applicatio
n_1635962504835_0001/
2021-11-04 02:02:52,645 INFO mapreduce.Job: Running job: job_1635962504835_0001
2021-11-04 02:03:01,801 INFO mapreduce.Job: Job job_1635962504835_0001 running in uber mode : false
2021-11-04 02:03:01,803 INFO mapreduce.Job:  map 0% reduce 0%
```



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Con
1	0	1	0	1

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	
3	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<mem

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartT
application_1635962504835_0001	root	inverted index	MAPREDUCE		default	0	Thu No 02:02:5 +0800

Showing 1 to 1 of 1 entries

```
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
root@h01:/usr/local/hadoop# bin/hdfs dfs -ls /user/flospro/inverted
Found 3 items
drwxr-xr-x   - root supergroup          0 2021-11-04 01:44 /user/flospro/inverted/input
drwxr-xr-x   - root supergroup          0 2021-11-04 01:49 /user/flospro/inverted/output
drwxr-xr-x   - root supergroup          0 2021-11-04 01:46 /user/flospro/inverted/stop
root@h01:/usr/local/hadoop# bin/hdfs dfs -rm -r /user/flospro/inverted/output
Deleted /user/flospro/inverted/output
root@h01:/usr/local/hadoop# bin/hadoop jar inverted.jar -Dinverted.case.sensitive=false -Dmapreduce.ou
tput.textoutputformat.separator=: /user/flospro/inverted/input /user/flospro/inverted/output/ -skip /u
ser/flospro/inverted/stop/punctuation.txt -skip /user/flospro/inverted/stop/stop-word-list.txt
2021-11-04 02:02:48,954 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager
at h01/172.19.0.2:8032
2021-11-04 02:02:49,342 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/ha
doop-yarn/staging/root/.staging/job_1635962504835_0001
2021-11-04 02:02:51,655 INFO input.FileInputFormat: Total input files to process : 40
2021-11-04 02:02:51,767 INFO mapreduce.JobSubmitter: number of splits:40
2021-11-04 02:02:51,950 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1635962504835_0001
2021-11-04 02:02:51,950 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-11-04 02:02:52,117 INFO conf.Configuration: resource-types.xml not found
2021-11-04 02:02:52,118 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-11-04 02:02:52,589 INFO impl.YarnClientImpl: Submitted application application_1635962504835_0001
2021-11-04 02:02:52,643 INFO mapreduce.Job: The url to track the job: http://h01:8088/proxy/applicatio
n_1635962504835_0001/
2021-11-04 02:02:52,645 INFO mapreduce.Job: Running job: job_1635962504835_0001
2021-11-04 02:03:01,801 INFO mapreduce.Job: Job job_1635962504835_0001 running in uber mode : false
2021-11-04 02:03:01,803 INFO mapreduce.Job:  map 0% reduce 0%
2021-11-04 02:03:55,619 INFO ipc.Client: Retrying connect to server: h03/172.19.0.4:42785. Already tri
ed 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISE
CONDS)
Killed
root@h01:/usr/local/hadoop# jps
4514 ResourceManager
4643 NodeManager
4262 SecondaryNameNode
5894 YarnChild
5895 YarnChild
5865 YarnChild
5803 YarnChild
5868 YarnChild
5839 YarnChild
5809 YarnChild
5817 YarnChild
4059 DataNode
6012 Jps
3934 NameNode
root@h01:/usr/local/hadoop# jps
```

虽然在单机上表现优秀，集群还是宕机，个人使用的电脑是小新air14 2019。本来是最近的作业跑不动时就有考虑申请bdkit，但一方面考虑到最近的wordcount3.0中我不恰当地同时创建了多个MR任务，是没有认真把控开销导致的宕机，正常操作可能能运行有侥幸心理，一方面老师说到最近南大的bdkit需要维护只能后续申请阿里云的需要注意作业完成时间就没有去申请，这次一定得找助教申请bdkit了。