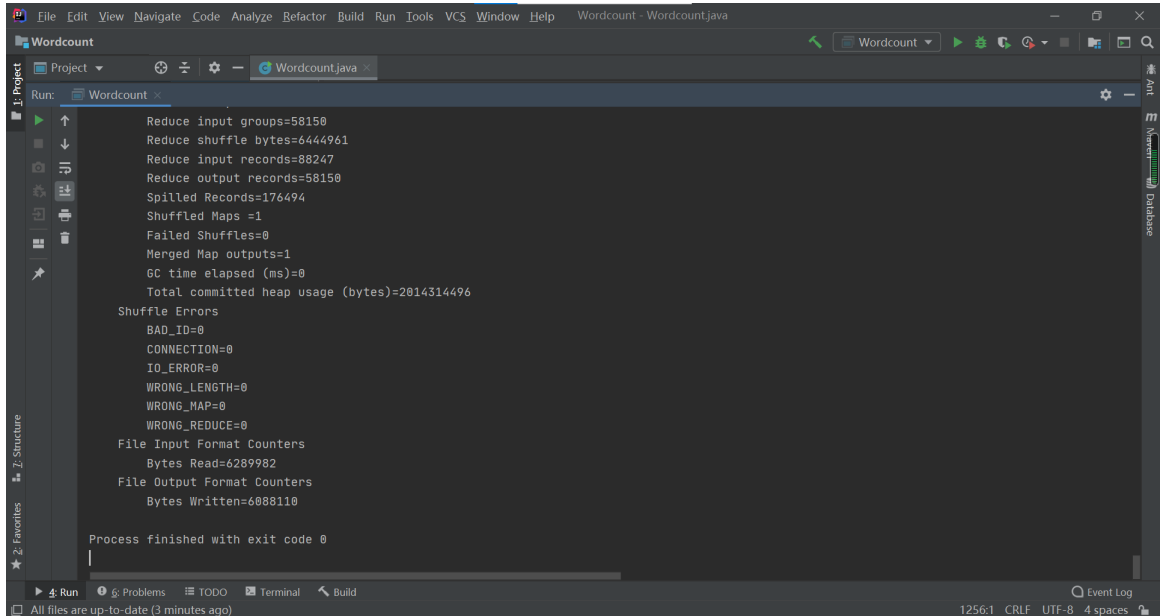


作业6

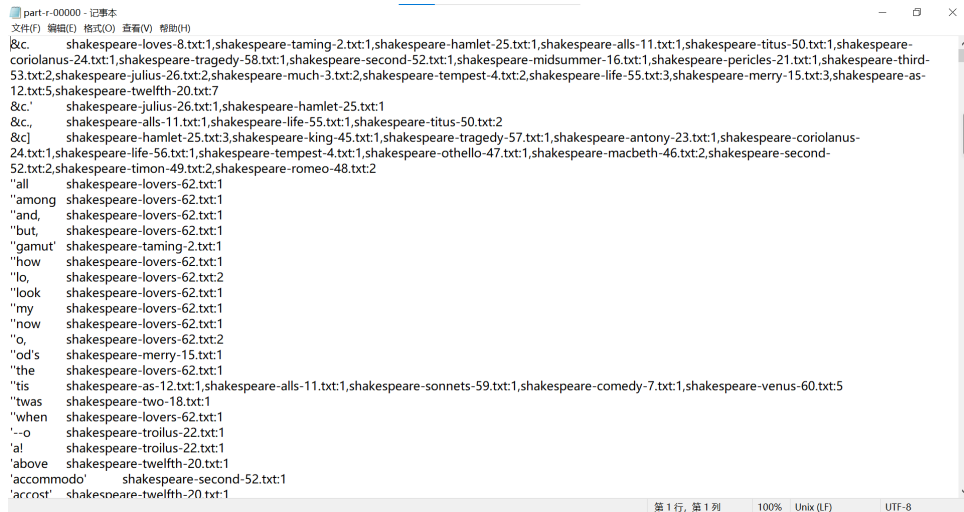
一. IDEA运行

- 运行成功



- 生成fortmp文件夹，输出文件夹为output，打开其中的part-r-00000，为运行结果

.idea	2021/11/4 10:18	文件夹	
fortmp	2021/11/4 10:18	文件夹	
input	2021/11/4 10:18	文件夹	
out	2021/11/4 10:18	文件夹	
output	2021/11/4 10:18	文件夹	
src	2021/11/4 10:18	文件夹	
target	2021/11/4 10:18	文件夹	
pom.xml	2021/10/30 22:18	XML 文档	2 KB
Wordcount.iml	2021/10/30 20:19	IML 文件	1 KB



- 打包成jar包，在文件夹out中，接下来放入虚拟机运行

Homework6 > Wordcount > out > artifacts > Wordcount_jar	▼	↺	🔍 搜索
名称	修改日期	类型	大小
Wordcount.jar	2021/11/4 10:18	Executable Jar File	66,803 KB

二. 虚拟机运行

1. 重新安装配置虚拟机

```
Activities Terminal
severus@ubuntu: ~/hadoop_installs/hadoop-2.9.1
File Edit View Search Terminal Help
Reduce input records=13
Reduce output records=13
Spilled Records=26
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=353
Total committed heap usage (bytes)=307249152
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=488
File Output Format Counters
  Bytes Written=220
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ ls
bin  etc  include  lib  libexec  LICENSE.txt  logs  NOTICE.txt  README.txt  sbin  share
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hdfs dfs -get output output
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ ls
bin  etc  include  lib  libexec  LICENSE.txt  logs  NOTICE.txt  output  README.txt  sbin  share
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ cd output
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1/output$ ls
part-r-000000 SUCCESS
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1/output$ cd ..
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ cat output/*
6 dfs.audit.logger
4 dfs.class
3 dfs.logger
3 dfs.server.namenode,
2 dfs.audit.log.maxbackupindex
2 dfs.period
2 dfs.audit.log.maxfilesize
1 dfs.log
1 dfs.file
1 dfs.servers
1 dfsadmin
1 dfsmetrics.log
1 dfs.replication
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$
```

2. 将jar包和input文件夹放进hadoop，再传至hdfs

```
Activities Terminal
severus@ubuntu: ~/hadoop_installs/hadoop-2.9.1
Firefox Web Browser
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1/homework_6$ cd ..
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hdfs dfs -put input input_home6
21/11/03 23:30:12 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedIOException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
21/11/03 23:30:13 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedIOException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
21/11/03 23:30:13 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedIOException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.closeInternal(DataStreamer.java:844)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:840)
21/11/03 23:30:16 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedIOException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.closeInternal(DataStreamer.java:844)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:840)
21/11/03 23:30:18 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedIOException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$
```



```
Activities Terminal
severus@ubuntu: ~/hadoop_installs/hadoop-2.9.1$ bin/hadoop jar homework_6/Wordcount.jar Wordcount input_home6 output_home6_new
21/11/04 00:14:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://localhost:9000/user/severus/fortmp already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:279)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:145)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1889)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
    at Wordcount.main(Wordcount.java:255)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:193)
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hdfs dfs -rm fortmp
rm: 'fortmp': Is a directory
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hdfs dfs -rm -r fortmp
Deleted fortmp
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hdfs dfs -rm -r output_home6
Deleted output_home6
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ bin/hadoop jar homework_6/Wordcount.jar Wordcount input_home6 output_home6
21/11/04 00:24:54 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/11/04 00:25:03 INFO input.FileInputFormat: Total input files to process : 42
21/11/04 00:25:04 INFO mapreduce.JobSubmitter: number of splits:42
21/11/04 00:25:05 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
21/11/04 00:25:07 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1636009744412_0001
21/11/04 00:25:12 INFO impl.YarnClientImpl: Submitted application application_1636009744412_0001
21/11/04 00:25:13 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1636009744412_0001/
21/11/04 00:25:13 INFO mapreduce.Job: Running job: job_1636009744412_0001
21/11/04 00:26:33 INFO mapreduce.Job: Job Job_1636009744412_0001 running in uber mode : false
21/11/04 00:26:33 INFO mapreduce.Job: map 0% reduce 0%
21/11/04 00:28:27 INFO mapreduce.Job: map 2% reduce 0%
21/11/04 00:28:28 INFO mapreduce.Job: map 3% reduce 0%
21/11/04 00:28:32 INFO mapreduce.Job: map 5% reduce 0%
21/11/04 00:28:33 INFO mapreduce.Job: map 6% reduce 0%
21/11/04 00:28:36 INFO mapreduce.Job: map 7% reduce 0%
```


yarn的情况

All Applications

Namenode information

+

localhost:8088/cluster



All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
2	0	0	2	0	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[MEMORY]	<memory:2048, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Priority	StartTime	FinishTime	State	FinalStatus
application_1636009744412_0002	severus	sort file	MAPREDUCE	default	0	Thu Nov 4 00:40:53 -0700 2021	Thu Nov 4 00:41:36 -0700 2021	FINISHED	SUCCEEDED
application_1636009744412_0001	severus	invertedIndex	MAPREDUCE	default	0	Thu Nov 4 00:25:10 -0700 2021	Thu Nov 4 00:40:49 -0700 2021	FINISHED	SUCCEEDED

Showing 1 to 2 of 2 entries

All Applications

Namenode information

localhost:8088/cluster

Logged in as: dr.who

All Applications

Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0 B	8 GB	0 B	0	8	0

Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
0	0	0	0	0

Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
<memory:2048, vCores:1>	<memory:4096, vCores:4>	0

Search:

Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
	Thu Nov 4 00:40:53 -0700 2021	Thu Nov 4 00:41:36 -0700 2021	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0		History	0
	Thu Nov 4 00:25:10 -0700 2021	Thu Nov 4 00:40:49 -0700 2021	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0		History	0

First Previous 1 Next Last