

实验一:安装单机 Hadoop 系统与 WordCount 程序实验

171860607

白晋斌

810594956@qq.com

目录

实验内容与要求.....	3
实验过程.....	3
1.单机操作系统安装	3
2.免密码 SSH 访问配置	3
3.安装 Java	3
4.创建用户	4
5.下载 Hadoop	4
6.解压安装 Hadoop	4
7.配置环境变量.....	4
8.修改 Hadoop 配置文件.....	4
8.1 hadoop-env.sh	4
8.2 core-site.xml	4
8.3 hdfs-site.xml	4
8.4 yarn-site.xml	4
8.5 mapred-site.xml.....	5
9.格式化 NameNode.....	5
10.启动 HDFS 和 MapReduce.....	6
11. Hadoop 本地库处理.....	6
12.运行测试	6
自行测试.....	13
1.文本来源	13
2.开始统计	14
3.作业运行状态.....	20
4.统计结果	20
实验体会.....	21
参考资料.....	21

实验内容与要求

1. 每人在自己本地电脑上正确安装和运行伪分布式 Hadoop 系统 安装操作手册和本课程课件请从 MapReduce 课程目录下载。
2. 安装完成后,自己寻找一组英文网页数据,在本机上运行 Hadoop 系统自带的 WordCount 可执行程序文件,并产生输出结果
3. 实验结果提交:要求书写一个实验报告,其中包括:
 - 1.系统安装运行的情况
 - 2.实验数据说明(下载的什么网页数据,多少个 HTML 或 text 文件)
 - 3.程序运行后在 Hadoop Web 作业状态查看界面上的作业运行状态屏幕拷贝
 - 4.实验输出结果开头部分的屏幕拷贝
 - 5.实验体会
 - 6.实验报告文件命名规则:MPLab1-学号-姓名.doc

实验完成时间: 4 月 9 日前完成并提交报告 (有特殊情况的同学联系助教)

实验过程

1.单机操作系统安装

我们直接选用 macOS 操作系统作为实验操作系统.

操作系统版本为 macOS Mojave version 10.14.5.

2.免密码 SSH 访问配置

```
Retina-MacBook-Pro:~ bryan$ ssh-keygen -t rsa -P ''
Generating public/private rsa key pair.
Enter file in which to save the key (/Users/bryan/.ssh/id_rsa):
/Users/bryan/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /Users/bryan/.ssh/id_rsa.
Your public key has been saved in /Users/bryan/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:ssjBk45cH8Vx67TtwXR9QcZjpbU50kZIhGgtdAUad/4 bryan@Retina-MacBook-Pro.local
The key's randomart image is:
+---[RSA 2048]-----+
|      .o++B=. o+ |
|      ==oo ..B. |
|      ..+ o.+ O + |
|      . . . o *.O . |
|      * o S o *E. |
|      . = = + . . |
|      o + o . |
|      |
+----[SHA256]-----+
Retina-MacBook-Pro:~ bryan$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
Retina-MacBook-Pro:~ bryan$ ssh localhost
Last login: Sun Mar 22 17:50:25 2020 from ::1
Retina-MacBook-Pro:~ bryan$
```

3.安装 Java

查看 java 版本 :

java -version

```
[Retina-MacBook-Pro:~ bryan$ java -version
java version "1.8.0_111"
Java(TM) SE Runtime Environment (build 1.8.0_111-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.111-b14, mixed mode)
Retina-MacBook-Pro:~ bryan$
```

4. 创建用户

这里直接使用系统管理员用户 bryan.

5. 下载 Hadoop

地址：<http://hadoop.apache.org/releases.html>

6. 解压安装 Hadoop

解压,放入根目录

7. 配置环境变量

```
export HADOOP_HOME=/Users/bryan/hadoop-2.7.1
```

```
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

关于 Java 的环境变量在更早的时候已完成配置.

8. 修改 Hadoop 配置文件

8.1 hadoop-env.sh

找到 hadoop-env.sh , 打开编辑进行如下设置 :

```
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_111.jdk/Contents/Home
```

8.2 core-site.xml

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/Users/bryan/hadoop-2.7.1/tmp</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
</property>
```

8.3 hdfs-site.xml

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
```

8.4 yarn-site.xml

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>localhost</value>
</property>
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
```

```

    <value>4096</value>
  </property>
  <property>
    <name>yarn.scheduler.minimum-allocation-mb</name>
    <value>2048</value>
  </property>
  <property>
    <name>yarn.nodemanager.vmem-pmem-ratio</name>
    <value>2.1</value>
  </property>
  <property>
    <name>yarn.log-aggregation-enable</name>
    <value>true</value>
  </property>

```

8.5 mapred-site.xml

```

  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>

```

9.格式化 NameNode

格式化成功，返回有关 NameNode 的启动信息，其中会有一句“.... has been successfully formatted.”

```

Retina-MacBook-Pro:bin bryan$ ./hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

20/03/22 18:50:47 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = Retina-MacBook-Pro.local/192.168.1.109
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.1
STARTUP_MSG: classpath = /Users/bryan/hadoop-2.7.1/etc/hadoop:/Users/bryan/hadoop-2.7.1/share/hadoop/common/lib/hadoop-auth-2.7.1.jar:/Users/bryan/hadoop-2.7.1/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/Users/bryan/hadoop-2.7.1/share/hadoop/common/lib/commons-junit4-4.11.jar:/Users/bryan/hadoop-2.7.1/share/hadoop/common/lib/commons
20/03/22 23:56:58 INFO namenode.FSDirectory: ACLs enabled? false
20/03/22 23:56:58 INFO namenode.FSDirectory: XAttrs enabled? true
20/03/22 23:56:58 INFO namenode.FSDirectory: Maximum size of an xattr: 16384
20/03/22 23:56:58 INFO namenode.NameNode: Caching file names occurring more than 10 times
20/03/22 23:56:58 INFO util.GSet: Computing capacity for map cachedBlocks
20/03/22 23:56:58 INFO util.GSet: VM type = 64-bit
20/03/22 23:56:58 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
20/03/22 23:56:58 INFO util.GSet: capacity = 2^18 = 262144 entries
20/03/22 23:56:58 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
20/03/22 23:56:58 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
20/03/22 23:56:58 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
20/03/22 23:56:58 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
20/03/22 23:56:58 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/03/22 23:56:58 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/03/22 23:56:58 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/03/22 23:56:58 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/03/22 23:56:58 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/03/22 23:56:58 INFO util.GSet: VM type = 64-bit
20/03/22 23:56:58 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/03/22 23:56:58 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory /Users/bryan/hadoop-2.7.1/tmp/dfs/name ? (Y or N) y
20/03/22 23:57:08 INFO namenode.FSImage: Allocated new BlockPoolId: BP-598796757-192.168.1.109-1584892628782
20/03/22 23:57:08 INFO common.Storage: Storage directory /Users/bryan/hadoop-2.7.1/tmp/dfs/name has been successfully formatted.
20/03/22 23:57:08 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/03/22 23:57:08 INFO util.ExitUtil: Exiting with status 0
20/03/22 23:57:08 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at Retina-MacBook-Pro.local/192.168.1.109
*****/
Retina-MacBook-Pro:bin bryan$

```

10. 启动 HDFS 和 MapReduce

```
Retina-MacBook-Pro:~ bryan$ ./start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
20/03/22 23:58:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /Users/bryan/hadoop-2.7.1/logs/hadoop-bryan-namenode-Retina-MacBook-Pro.local.out
localhost: starting datanode, logging to /Users/bryan/hadoop-2.7.1/logs/hadoop-bryan-datanode-Retina-MacBook-Pro.local.out
Starting secondary namenodes [account.jetbrains.com]
account.jetbrains.com: starting secondarynamenode, logging to /Users/bryan/hadoop-2.7.1/logs/hadoop-bryan-secondarynamenode-Retina-MacBook-Pro.local.out
20/03/22 23:58:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting yarn daemons
resourcemanager running as process 80788. Stop it first.
localhost: nodemanager running as process 80870. Stop it first.

[Retina-MacBook-Pro:~ bryan$ jps
86899 SecondaryNameNode
80788 ResourceManager
80870 NodeManager
86795 DataNode
86714 NameNode
87069 Jps
58797
```

11. Hadoop 本地库处理

这里我们会发现,每句指令后面都有一句 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable Found 2 items

这是因为：Hadoop 本地库是为了提高效率或者某些不能用 Java 实现的功能组件库。目前只支持 *unix, 在 Mac OS X 和 Cygwin 上不提供直接支持。因此需要自己重新编译, 这里有篇文章可供参考：Mac OSX 下 Hadoop 使用本地库提高效率, 详见参考资料[12][13][14].从而我们可以手动加入本地库。

12. 运行测试

文件内容：

file1.txt:hello hadoop hello world

file2.txt: goodbye hadoop

输入指令：

hadoop fs -put ~/Desktop/test/file1.txt /input

hadoop fs -put ~/Desktop/test/file2.txt /input

hadoop fs -ls /input

终端显示：

```
Retina-MacBook-Pro:~ bryan$ hadoop fs -ls /input
Found 2 items
-rw-r--r--  1 bryan supergroup    24 2020-03-24 17:18 /input/file1.txt
-rw-r--r--  1 bryan supergroup    14 2020-03-24 17:18 /input/file2.txt
```

输入指令：

cd ~/hadoop-2.7.1/share/hadoop/mapreduce/

hadoop jar hadoop-mapreduce-examples-2.7.1.jar wordcount /input /output

终端显示：

```

20/03/24 17:26:14 INFO mapreduce.Job: Job job_local1787702454_0001 completed successfully
20/03/24 17:26:14 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=821804
    FILE: Number of bytes written=1682633
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=100
    HDFS: Number of bytes written=35
    HDFS: Number of read operations=22
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=5
  Map-Reduce Framework
    Map input records=2
    Map output records=6
    Map output bytes=64
    Map output materialized bytes=76
    Input split bytes=204
    Combine input records=6
    Combine output records=5
    Reduce input groups=4
    Reduce shuffle bytes=76
    Reduce input records=5
    Reduce output records=4
    Spilled Records=10
    Shuffled Maps =2
    Failed Shuffles=0
    Merged Map outputs=2
    GC time elapsed (ms)=5
    Total committed heap usage (bytes)=1135083520
  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=38
  File Output Format Counters
    Bytes Written=35
Retina-MacBook-Pro:mapreduce bryan$

```

详细文本如下:

```
Retina-MacBook-Pro:mapreduce bryan$ hadoop jar hadoop-mapreduce-examples-2.7.1.jar
wordcount /input /output
```

```
20/03/24 17:26:11 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
```

```
20/03/24 17:26:11 INFO jvm.JvmMetrics: Initializing JVM Metrics with
processName=JobTracker, sessionId=
```

```
20/03/24 17:26:12 INFO input.FileInputFormat: Total input paths to process : 2
```

```
20/03/24 17:26:12 INFO mapreduce.JobSubmitter: number of splits:2
```

```
20/03/24 17:26:12 INFO mapreduce.JobSubmitter: Submitting tokens for job:
```

```
job_local1787702454_0001
```

```
20/03/24 17:26:12 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
```

20/03/24 17:26:12 INFO mapreduce.Job: Running job: job_local1787702454_0001
20/03/24 17:26:12 INFO mapred.LocalJobRunner: OutputCommitter set in config null
20/03/24 17:26:12 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
20/03/24 17:26:12 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
20/03/24 17:26:12 INFO mapred.LocalJobRunner: Waiting for map tasks
20/03/24 17:26:12 INFO mapred.LocalJobRunner: Starting task: attempt_local1787702454_0001_m_000000_0
20/03/24 17:26:12 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
20/03/24 17:26:12 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
20/03/24 17:26:12 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
20/03/24 17:26:12 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/input/file1.txt:0+24
20/03/24 17:26:13 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
20/03/24 17:26:13 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
20/03/24 17:26:13 INFO mapred.MapTask: soft limit at 83886080
20/03/24 17:26:13 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
20/03/24 17:26:13 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
20/03/24 17:26:13 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask\$MapOutputBuffer
20/03/24 17:26:13 INFO mapred.LocalJobRunner:
20/03/24 17:26:13 INFO mapred.MapTask: Starting flush of map output
20/03/24 17:26:13 INFO mapred.MapTask: Spilling map output
20/03/24 17:26:13 INFO mapred.MapTask: bufstart = 0; bufend = 41; bufvoid = 104857600
20/03/24 17:26:13 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214384(104857536); length = 13/6553600
20/03/24 17:26:13 INFO mapred.MapTask: Finished spill 0
20/03/24 17:26:13 INFO mapred.Task: Task:attempt_local1787702454_0001_m_000000_0 is done. And is in the process of committing

20/03/24 17:26:13 INFO mapred.LocalJobRunner: map

20/03/24 17:26:13 INFO mapred.Task: Task 'attempt_local1787702454_0001_m_000000_0'
done.

20/03/24 17:26:13 INFO mapred.LocalJobRunner: Finishing task:
attempt_local1787702454_0001_m_000000_0

20/03/24 17:26:13 INFO mapred.LocalJobRunner: Starting task:
attempt_local1787702454_0001_m_000001_0

20/03/24 17:26:13 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 1

20/03/24 17:26:13 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is
supported only on Linux.

20/03/24 17:26:13 INFO mapred.Task: Using ResourceCalculatorProcessTree : null

20/03/24 17:26:13 INFO mapred.MapTask: Processing split:
hdfs://localhost:9000/input/file2.txt:0+14

20/03/24 17:26:13 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

20/03/24 17:26:13 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

20/03/24 17:26:13 INFO mapred.MapTask: soft limit at 83886080

20/03/24 17:26:13 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

20/03/24 17:26:13 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

20/03/24 17:26:13 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer

20/03/24 17:26:13 INFO mapred.LocalJobRunner:

20/03/24 17:26:13 INFO mapred.MapTask: Starting flush of map output

20/03/24 17:26:13 INFO mapred.MapTask: Spilling map output

20/03/24 17:26:13 INFO mapred.MapTask: bufstart = 0; bufend = 23; bufvoid = 104857600

20/03/24 17:26:13 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend =
26214392(104857568); length = 5/6553600

20/03/24 17:26:13 INFO mapred.MapTask: Finished spill 0

20/03/24 17:26:13 INFO mapred.Task: Task:attempt_local1787702454_0001_m_000001_0
is done. And is in the process of committing

20/03/24 17:26:13 INFO mapred.LocalJobRunner: map

20/03/24 17:26:13 INFO mapred.Task: Task 'attempt_local1787702454_0001_m_000001_0'
done.

20/03/24 17:26:13 INFO mapred.LocalJobRunner: Finishing task:
attempt_local1787702454_0001_m_000001_0

20/03/24 17:26:13 INFO mapred.LocalJobRunner: map task executor complete.

20/03/24 17:26:13 INFO mapred.LocalJobRunner: Waiting for reduce tasks

20/03/24 17:26:13 INFO mapred.LocalJobRunner: Starting task:
attempt_local1787702454_0001_r_000000_0

20/03/24 17:26:13 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 1

20/03/24 17:26:13 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is
supported only on Linux.

20/03/24 17:26:13 INFO mapred.Task: Using ResourceCalculatorProcessTree : null

20/03/24 17:26:13 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@659865c8

20/03/24 17:26:13 INFO reduce.MergeManagerImpl: MergerManager:
memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=220663392,
ioSortFactor=10, memToMemMergeOutputsThreshold=10

20/03/24 17:26:13 INFO reduce.EventFetcher: attempt_local1787702454_0001_r_000000_0
Thread started: EventFetcher for fetching Map Completion Events

20/03/24 17:26:13 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
attempt_local1787702454_0001_m_000001_0 decomp: 29 len: 33 to MEMORY

20/03/24 17:26:13 INFO reduce.InMemoryMapOutput: Read 29 bytes from map-output for
attempt_local1787702454_0001_m_000001_0

20/03/24 17:26:13 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of
size: 29, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->29

20/03/24 17:26:13 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
attempt_local1787702454_0001_m_000000_0 decomp: 39 len: 43 to MEMORY

20/03/24 17:26:13 INFO reduce.InMemoryMapOutput: Read 39 bytes from map-output for
attempt_local1787702454_0001_m_000000_0

20/03/24 17:26:13 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of
size: 39, inMemoryMapOutputs.size() -> 2, commitMemory -> 29, usedMemory ->68

20/03/24 17:26:13 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
20/03/24 17:26:13 INFO mapred.LocalJobRunner: 2 / 2 copied.
20/03/24 17:26:13 INFO reduce.MergeManagerImpl: finalMerge called with 2 in-memory
map-outputs and 0 on-disk map-outputs
20/03/24 17:26:13 INFO mapred.Merger: Merging 2 sorted segments
20/03/24 17:26:13 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left
of total size: 49 bytes
20/03/24 17:26:13 INFO reduce.MergeManagerImpl: Merged 2 segments, 68 bytes to disk
to satisfy reduce memory limit
20/03/24 17:26:13 INFO reduce.MergeManagerImpl: Merging 1 files, 70 bytes from disk
20/03/24 17:26:13 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from
memory into reduce
20/03/24 17:26:13 INFO mapred.Merger: Merging 1 sorted segments
20/03/24 17:26:13 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left
of total size: 56 bytes
20/03/24 17:26:13 INFO mapred.LocalJobRunner: 2 / 2 copied.
20/03/24 17:26:13 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead,
use mapreduce.job.skiprecords
20/03/24 17:26:13 INFO mapreduce.Job: Job job_local1787702454_0001 running in uber
mode : false
20/03/24 17:26:13 INFO mapreduce.Job: map 100% reduce 0%
20/03/24 17:26:13 INFO mapred.Task: Task:attempt_local1787702454_0001_r_000000_0 is
done. And is in the process of committing
20/03/24 17:26:13 INFO mapred.LocalJobRunner: 2 / 2 copied.
20/03/24 17:26:13 INFO mapred.Task: Task attempt_local1787702454_0001_r_000000_0 is
allowed to commit now
20/03/24 17:26:14 INFO output.FileOutputCommitter: Saved output of task
'attempt_local1787702454_0001_r_000000_0' to
hdfs://localhost:9000/output/_temporary/0/task_local1787702454_0001_r_000000
20/03/24 17:26:14 INFO mapred.LocalJobRunner: reduce > reduce
20/03/24 17:26:14 INFO mapred.Task: Task 'attempt_local1787702454_0001_r_000000_0'
done.

20/03/24 17:26:14 INFO mapred.LocalJobRunner: Finishing task:
attempt_local1787702454_0001_r_000000_0
20/03/24 17:26:14 INFO mapred.LocalJobRunner: reduce task executor complete.
20/03/24 17:26:14 INFO mapreduce.Job: map 100% reduce 100%
20/03/24 17:26:14 INFO mapreduce.Job: Job job_local1787702454_0001 completed
successfully
20/03/24 17:26:14 INFO mapreduce.Job: Counters: 35

File System Counters

FILE: Number of bytes read=821804
FILE: Number of bytes written=1682633
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=100
HDFS: Number of bytes written=35
HDFS: Number of read operations=22
HDFS: Number of large read operations=0
HDFS: Number of write operations=5

Map-Reduce Framework

Map input records=2
Map output records=6
Map output bytes=64
Map output materialized bytes=76
Input split bytes=204
Combine input records=6
Combine output records=5
Reduce input groups=4
Reduce shuffle bytes=76
Reduce input records=5
Reduce output records=4
Spilled Records=10
Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=5

Total committed heap usage (bytes)=1135083520

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=38

File Output Format Counters

Bytes Written=35

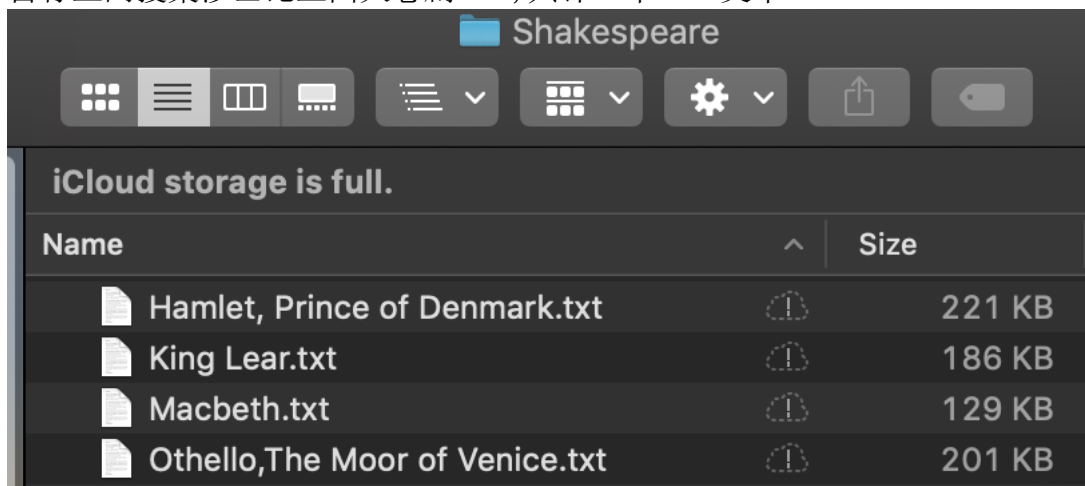
查看统计结果:

```
Retina-MacBook-Pro:mapreduce bryan$ hadoop fs -ls /output
Found 2 items
-rw-r--r--  1 bryan supergroup          0 2020-03-24 17:26 /output/_SUCCESS
-rw-r--r--  1 bryan supergroup        35 2020-03-24 17:26 /output/part-r-00000
Retina-MacBook-Pro:mapreduce bryan$ hadoop fs -cat /output/part-r-00000
goodbye 1
hadoop  2
hello   2
world   1
```

自行测试

1.文本来源

自行上网搜集莎士比亚四大悲剧 txt, 共计 4 个 txt 文本.



2.开始统计

输入命令:

```
hadoop fs -copyFromLocal ~/Desktop/Shakespeare/ /
mapreduce bryan$ hadoop fs -ls /
cd ~/hadoop-2.7.1/share/hadoop/mapreduce/
hadoop jar hadoop-mapreduce-examples-2.7.1.jar wordcount /Shakespeare
/ShakespeareStatistics
终端显示:
```

```
Map-Reduce Framework
  Map input records=20271
  Map output records=111064
  Map output bytes=1062229
  Map output materialized bytes=377260
  Input split bytes=479
  Combine input records=111064
  Combine output records=27768
  Reduce input groups=18461
  Reduce shuffle bytes=377260
  Reduce input records=27768
  Reduce output records=18461
  Spilled Records=55536
  Shuffled Maps =4
  Failed Shuffles=0
  Merged Map outputs=4
  GC time elapsed (ms)=13
  Total committed heap usage (bytes)=2364014592
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=736752
File Output Format Counters
  Bytes Written=184888
Retina-MacBook-Pro:mapreduce bryan$
```

详细输出内容如下:

```
Retina-MacBook-Pro:mapreduce bryan$ hadoop jar hadoop-mapreduce-examples-2.7.1.jar
wordcount /Shakespeare /ShakespeareStatistics
20/03/25 11:41:49 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
20/03/25 11:41:49 INFO jvm.JvmMetrics: Initializing JVM Metrics with
processName=JobTracker, sessionId=
20/03/25 11:41:50 INFO input.FileInputFormat: Total input paths to process : 4
20/03/25 11:41:50 INFO mapreduce.JobSubmitter: number of splits:4
20/03/25 11:41:50 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_local741435052_0001
20/03/25 11:41:50 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
20/03/25 11:41:50 INFO mapreduce.Job: Running job: job_local741435052_0001
20/03/25 11:41:50 INFO mapred.LocalJobRunner: OutputCommitter set in config null
20/03/25 11:41:50 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 1
20/03/25 11:41:50 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
```

20/03/25 11:41:50 INFO mapred.LocalJobRunner: Waiting for map tasks

20/03/25 11:41:50 INFO mapred.LocalJobRunner: Starting task:
attempt_local741435052_0001_m_000000_0

20/03/25 11:41:50 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 1

20/03/25 11:41:50 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is
supported only on Linux.

20/03/25 11:41:50 INFO mapred.Task: Using ResourceCalculatorProcessTree : null

20/03/25 11:41:50 INFO mapred.MapTask: Processing split:
hdfs://localhost:9000/Shakespeare/Hamlet, Prince of Denmark.TXT:0+221121

20/03/25 11:41:50 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

20/03/25 11:41:50 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

20/03/25 11:41:50 INFO mapred.MapTask: soft limit at 83886080

20/03/25 11:41:50 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

20/03/25 11:41:50 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

20/03/25 11:41:50 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer

20/03/25 11:41:50 INFO mapred.LocalJobRunner:

20/03/25 11:41:50 INFO mapred.MapTask: Starting flush of map output

20/03/25 11:41:50 INFO mapred.MapTask: Spilling map output

20/03/25 11:41:50 INFO mapred.MapTask: bufstart = 0; bufend = 317851; bufvoid =
104857600

20/03/25 11:41:50 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend =
26080708(104322832); length = 133689/6553600

20/03/25 11:41:50 INFO mapred.MapTask: Finished spill 0

20/03/25 11:41:50 INFO mapred.Task: Task:attempt_local741435052_0001_m_000000_0 is
done. And is in the process of committing

20/03/25 11:41:50 INFO mapred.LocalJobRunner: map

20/03/25 11:41:50 INFO mapred.Task: Task 'attempt_local741435052_0001_m_000000_0'
done.

20/03/25 11:41:50 INFO mapred.LocalJobRunner: Finishing task:
attempt_local741435052_0001_m_000000_0

20/03/25 11:41:50 INFO mapred.LocalJobRunner: Starting task:
attempt_local741435052_0001_m_000001_0

20/03/25 11:41:50 INFO output.FileOutputCommitter: File Output Committer Algorithm
version is 1

20/03/25 11:41:50 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is
supported only on Linux.

20/03/25 11:41:50 INFO mapred.Task: Using ResourceCalculatorProcessTree : null

20/03/25 11:41:50 INFO mapred.MapTask: Processing split:
hdfs://localhost:9000/Shakespeare/Othello,The Moor of Venice.TXT:0+200717

20/03/25 11:41:50 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

20/03/25 11:41:50 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

20/03/25 11:41:50 INFO mapred.MapTask: soft limit at 83886080

20/03/25 11:41:50 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

20/03/25 11:41:50 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

20/03/25 11:41:50 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer

20/03/25 11:41:51 INFO mapred.LocalJobRunner:

20/03/25 11:41:51 INFO mapred.MapTask: Starting flush of map output

20/03/25 11:41:51 INFO mapred.MapTask: Spilling map output
 20/03/25 11:41:51 INFO mapred.MapTask: bufstart = 0; bufend = 280514; bufvoid = 104857600
 20/03/25 11:41:51 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26097444(104389776); length = 116953/6553600
 20/03/25 11:41:51 INFO mapred.MapTask: Finished spill 0
 20/03/25 11:41:51 INFO mapred.Task: Task:attempt_local741435052_0001_m_000001_0 is done. And is in the process of committing
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: map
 20/03/25 11:41:51 INFO mapred.Task: Task 'attempt_local741435052_0001_m_000001_0' done.
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Finishing task: attempt_local741435052_0001_m_000001_0
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Starting task: attempt_local741435052_0001_m_000002_0
 20/03/25 11:41:51 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
 20/03/25 11:41:51 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
 20/03/25 11:41:51 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
 20/03/25 11:41:51 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/Shakespeare/King Lear.TXT:0+185797
 20/03/25 11:41:51 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
 20/03/25 11:41:51 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
 20/03/25 11:41:51 INFO mapred.MapTask: soft limit at 83886080
 20/03/25 11:41:51 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
 20/03/25 11:41:51 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
 20/03/25 11:41:51 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask\$MapOutputBuffer
 20/03/25 11:41:51 INFO mapred.LocalJobRunner:
 20/03/25 11:41:51 INFO mapred.MapTask: Starting flush of map output
 20/03/25 11:41:51 INFO mapred.MapTask: Spilling map output
 20/03/25 11:41:51 INFO mapred.MapTask: bufstart = 0; bufend = 269651; bufvoid = 104857600
 20/03/25 11:41:51 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26101400(104405600); length = 112997/6553600
 20/03/25 11:41:51 INFO mapred.MapTask: Finished spill 0
 20/03/25 11:41:51 INFO mapred.Task: Task:attempt_local741435052_0001_m_000002_0 is done. And is in the process of committing
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: map
 20/03/25 11:41:51 INFO mapred.Task: Task 'attempt_local741435052_0001_m_000002_0' done.
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Finishing task: attempt_local741435052_0001_m_000002_0
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Starting task: attempt_local741435052_0001_m_000003_0
 20/03/25 11:41:51 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
 20/03/25 11:41:51 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.

20/03/25 11:41:51 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
 20/03/25 11:41:51 INFO mapred.MapTask: Processing split:
 hdfs://localhost:9000/Shakespeare/Macbeth.txt:0+129117
 20/03/25 11:41:51 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
 20/03/25 11:41:51 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
 20/03/25 11:41:51 INFO mapred.MapTask: soft limit at 83886080
 20/03/25 11:41:51 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
 20/03/25 11:41:51 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
 20/03/25 11:41:51 INFO mapred.MapTask: Map output collector class =
 org.apache.hadoop.mapred.MapTask\$MapOutputBuffer
 20/03/25 11:41:51 INFO mapred.LocalJobRunner:
 20/03/25 11:41:51 INFO mapred.MapTask: Starting flush of map output
 20/03/25 11:41:51 INFO mapred.MapTask: Spilling map output
 20/03/25 11:41:51 INFO mapred.MapTask: bufstart = 0; bufend = 194213; bufvoid =
 104857600
 20/03/25 11:41:51 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend =
 26133792(104535168); length = 80605/6553600
 20/03/25 11:41:51 INFO mapred.MapTask: Finished spill 0
 20/03/25 11:41:51 INFO mapred.Task: Task:attempt_local741435052_0001_m_000003_0 is
 done. And is in the process of committing
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: map
 20/03/25 11:41:51 INFO mapred.Task: Task 'attempt_local741435052_0001_m_000003_0'
 done.
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Finishing task:
 attempt_local741435052_0001_m_000003_0
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: map task executor complete.
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Waiting for reduce tasks
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Starting task:
 attempt_local741435052_0001_r_000000_0
 20/03/25 11:41:51 INFO output.FileOutputCommitter: File Output Committer Algorithm
 version is 1
 20/03/25 11:41:51 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is
 supported only on Linux.
 20/03/25 11:41:51 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
 20/03/25 11:41:51 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
 org.apache.hadoop.mapreduce.task.reduce.Shuffle@4bd35b4a
 20/03/25 11:41:51 INFO reduce.MergeManagerImpl: MergerManager:
 memoryLimit=368102592, maxSingleShuffleLimit=92025648, mergeThreshold=242947728,
 ioSortFactor=10, memToMemMergeOutputsThreshold=10
 20/03/25 11:41:51 INFO reduce.EventFetcher: attempt_local741435052_0001_r_000000_0
 Thread started: EventFetcher for fetching Map Completion Events
 20/03/25 11:41:51 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
 attempt_local741435052_0001_m_000003_0 decomp: 77762 len: 77766 to MEMORY
 20/03/25 11:41:51 INFO reduce.InMemoryMapOutput: Read 77762 bytes from map-output
 for attempt_local741435052_0001_m_000003_0
 20/03/25 11:41:51 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of
 size: 77762, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory -> 77762
 20/03/25 11:41:51 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
 attempt_local741435052_0001_m_000000_0 decomp: 110588 len: 110592 to MEMORY

20/03/25 11:41:51 INFO reduce.InMemoryMapOutput: Read 110588 bytes from map-output for attempt_local741435052_0001_m_000000_0

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 110588, inMemoryMapOutputs.size() -> 2, commitMemory -> 77762, usedMemory ->188350

20/03/25 11:41:51 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local741435052_0001_m_000002_0 decomp: 98186 len: 98190 to MEMORY

20/03/25 11:41:51 INFO reduce.InMemoryMapOutput: Read 98186 bytes from map-output for attempt_local741435052_0001_m_000002_0

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 98186, inMemoryMapOutputs.size() -> 3, commitMemory -> 188350, usedMemory ->286536

20/03/25 11:41:51 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local741435052_0001_m_000001_0 decomp: 90708 len: 90712 to MEMORY

20/03/25 11:41:51 INFO reduce.InMemoryMapOutput: Read 90708 bytes from map-output for attempt_local741435052_0001_m_000001_0

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 90708, inMemoryMapOutputs.size() -> 4, commitMemory -> 286536, usedMemory ->377244

20/03/25 11:41:51 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning

20/03/25 11:41:51 INFO mapred.LocalJobRunner: 4 / 4 copied.

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: finalMerge called with 4 in-memory map-outputs and 0 on-disk map-outputs

20/03/25 11:41:51 INFO mapred.Merger: Merging 4 sorted segments

20/03/25 11:41:51 INFO mapred.Merger: Down to the last merge-pass, with 4 segments left of total size: 377214 bytes

20/03/25 11:41:51 INFO mapreduce.Job: Job job_local741435052_0001 running in uber mode : false

20/03/25 11:41:51 INFO mapreduce.Job: map 100% reduce 0%

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: Merged 4 segments, 377244 bytes to disk to satisfy reduce memory limit

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: Merging 1 files, 377242 bytes from disk

20/03/25 11:41:51 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce

20/03/25 11:41:51 INFO mapred.Merger: Merging 1 sorted segments

20/03/25 11:41:51 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 377230 bytes

20/03/25 11:41:51 INFO mapred.LocalJobRunner: 4 / 4 copied.

20/03/25 11:41:51 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords

20/03/25 11:41:51 INFO mapred.Task: Task:attempt_local741435052_0001_r_000000_0 is done. And is in the process of committing

20/03/25 11:41:51 INFO mapred.LocalJobRunner: 4 / 4 copied.

20/03/25 11:41:51 INFO mapred.Task: Task attempt_local741435052_0001_r_000000_0 is allowed to commit now

20/03/25 11:41:51 INFO output.FileOutputCommitter: Saved output of task 'attempt_local741435052_0001_r_000000_0' to

hdfs://localhost:9000/ShakespeareStatistics/_temporary/0/task_local741435052_0001_r_000000

20/03/25 11:41:51 INFO mapred.LocalJobRunner: reduce > reduce
 20/03/25 11:41:51 INFO mapred.Task: Task 'attempt_local741435052_0001_r_000000_0'
 done.
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: Finishing task:
 attempt_local741435052_0001_r_000000_0
 20/03/25 11:41:51 INFO mapred.LocalJobRunner: reduce task executor complete.
 20/03/25 11:41:52 INFO mapreduce.Job: map 100% reduce 100%
 20/03/25 11:41:52 INFO mapreduce.Job: Job job_local741435052_0001 completed
 successfully
 20/03/25 11:41:52 INFO mapreduce.Job: Counters: 35

File System Counters

- FILE: Number of bytes read=2129322
- FILE: Number of bytes written=4541690
- FILE: Number of read operations=0
- FILE: Number of large read operations=0
- FILE: Number of write operations=0
- HDFS: Number of bytes read=2724098
- HDFS: Number of bytes written=184888
- HDFS: Number of read operations=46
- HDFS: Number of large read operations=0
- HDFS: Number of write operations=7

Map-Reduce Framework

- Map input records=20271
- Map output records=111064
- Map output bytes=1062229
- Map output materialized bytes=377260
- Input split bytes=479
- Combine input records=111064
- Combine output records=27768
- Reduce input groups=18461
- Reduce shuffle bytes=377260
- Reduce input records=27768
- Reduce output records=18461
- Spilled Records=55536
- Shuffled Maps =4
- Failed Shuffles=0
- Merged Map outputs=4
- GC time elapsed (ms)=13
- Total committed heap usage (bytes)=2364014592

Shuffle Errors

- BAD_ID=0
- CONNECTION=0
- IO_ERROR=0
- WRONG_LENGTH=0
- WRONG_MAP=0
- WRONG_REDUCE=0

File Input Format Counters

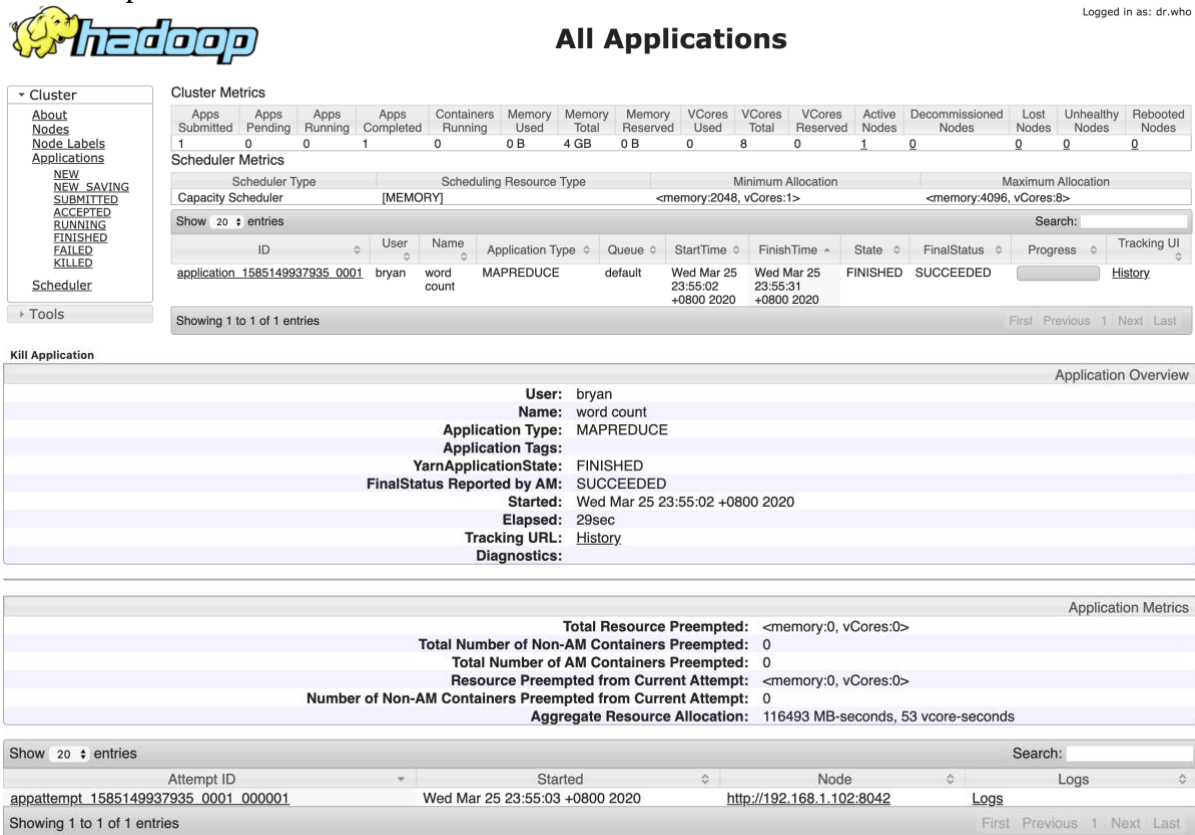
- Bytes Read=736752

File Output Format Counters

- Bytes Written=184888

3.作业运行状态

在 Hadoop Web 作业状态查看界面上的作业运行状态屏幕拷贝：
<http://localhost:8088/>



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Vcores Used	Vcores Total	Vcores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	4 GB	0 B	0	8	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:2048, vCores:1>	<memory:4096, vCores:8>

Application Overview

User: bryan
Name: word count
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: FINISHED
FinalStatus Reported by AM: SUCCEEDED
Started: Wed Mar 25 23:55:02 +0800 2020
Elapsed: 29sec
Tracking URL: History
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 116493 MB-seconds, 53 vcore-seconds

Attempt ID | **Started** | **Node** | **Logs**

attempt_1585149937935_0001_000001	Wed Mar 25 23:55:03 +0800 2020	http://192.168.1.102:8042	Logs
-----------------------------------	--------------------------------	---------------------------	------

4.统计结果

统计结果如图所示, 已将结果文件拷贝至本地, 更详细的统计结果见附件.

```
Retina-MacBook-Pro:mapreduce bryan$ hadoop fs -ls /ShakespeareStatistics
Found 2 items
-rw-r--r-- 1 bryan supergroup 0 2020-03-25 11:41 /ShakespeareStatistics/_SUCCESS
-rw-r--r-- 1 bryan supergroup 184888 2020-03-25 11:41 /ShakespeareStatistics/part-r-00000
Retina-MacBook-Pro:mapreduce bryan$ hadoop fs -cat /ShakespeareStatistics/part-r-00000
20/03/25 11:43:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
'Tis 1
'A 3
"AS-IS". 1
"Amen" 2
"Amen"? 1
"Amen," 1
"And 1
"Aroint 1
"Black 1
"Certes," 1
"Come 1
"Cursed 1
"Defect" 1
"Fear 2
"Give 1
"Glamis 1
"God 2
"Good 1
"He 1
"Help, 1
"Here 1
"Hold, 2
"I 4
```

`hadoop fs -get /ShakespeareStatistics ~/Desktop/Shakespeare/`

实验体会

Hadoop 的核心是 MapReduce，而 MapReduce 的核心又在于 map 和 reduce 函数。它们是交给用户实现的，这两个函数定义了任务本身。通过对 Hadoop 的应用，可以大大提升我们处理大数据的效率。

参考资料

- [1]<https://www.jianshu.com/p/6d32b166f47d>
- [2]https://blog.csdn.net/huxingguang_ios/article/details/78709428
- [3]<http://www.hihubs.com/article/341>
- [4]<https://blog.csdn.net/huanghanqian/article/details/78185285>
- [5]https://blog.csdn.net/yuxin_csdn/article/details/39703683
- [6]<https://blog.csdn.net/xbwer/article/details/35614679>
- [7]<https://blog.csdn.net/edogawachia/article/details/96492689>
- [8]<https://blog.csdn.net/xiaohang20103/article/details/9624771>
- [9]<https://blog.csdn.net/dai451954706/article/details/50464036>
- [10]<https://blog.csdn.net/ViMan1204/article/details/89048469>
- [11]<https://www.jianshu.com/p/d19ce17234b7>
- [12]<https://www.jianshu.com/p/dea5a5ec3f2c>
- [13]http://rockyfeng.me/hadoop_native_library_mac.html
- [14]<https://blog.csdn.net/pengzonglu7292/article/details/89055414>
- [15]https://blog.csdn.net/tswc_byy/article/details/90577551
- [16]https://blog.csdn.net/th_num/article/details/79916430
- [17]<https://blog.csdn.net/xiakexiaohu/article/details/54580971>
- [18]<https://segmentfault.com/a/1190000002672666>