



哈爾濱工業大學 (深圳)
HARBIN INSTITUTE OF TECHNOLOGY

实验报告

开课学期: 2022 秋季
课程名称: 大数据导论
实验名称: Hadoop 环境配置与基本操作
实验性质: 设计型
实验学时: 2 地点: T2608
学生班级: 保密
学生学号: 保密
学生姓名: wb-yu
评阅教师: _____
报告成绩: _____

实验与创新实践教育中心制

2022 年 9 月

一、实验目的

- 掌握熟悉 Linux 操作系统基础操作，掌握在 Linux 下安装 Java 的技能
- 掌握 Hadoop 分布式环境配置方法，理解 MapReduce 作业的原理和操作方法
- 掌握 Hadoop 集群的启动与关闭操作，并在 Hadoop 上执行词频统计程序

二、实验环境

- CentOS 6.10
- Hadoop 2.6.4
- Java 1.8

三、实验内容

- 在 Linux 下安装 Java，修改 Hadoop 配置文件，独立搭建 Hadoop 分布式环境
- 上传预处理后的实验文本数据至 HDFS
- 启动关闭 Hadoop 集群
- 使用 MapReduce 实现 WordCount 程序任务

四、实验过程

1、配置集群机器，下载安装包。

如图 1 所示，在终端输入 `initnetwork` 命令初始化网络。

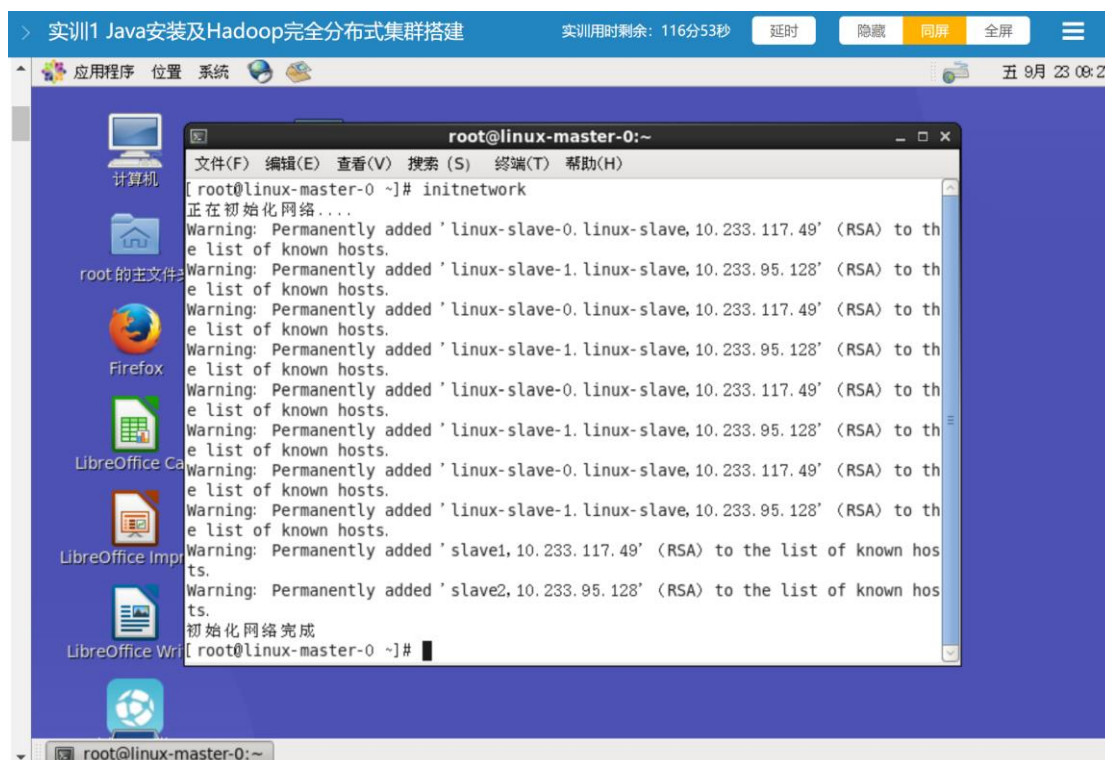


图 1

如图 2 所示，在 Linux 终端执行命令 `wget -P /opt/software http://datasrc.tipdm.net:81/bigdata/hadoop/software/hadoop-2.6.4.tar.gz` 和命令 `wget -P /opt/software http://datasrc.tipdm.net:81/bigdata/hadoop/software/jdk-8u151-linux-x64.rpm`，从而分别将 `jdk-8u151-linux-x64.rpm` 和 `hadoop-2.6.4.tar.gz` 下载到 Linux 本地 `/opt/software` 目录。

```

root@linux-master-0:~
[ root@linux-master-0 ~]# wget -P /opt/software http://datasrc.tipdm.net:81/bigdata/hadoop/software/hadoop-2.6.4.tar.gz
--2022-09-23 09:36:29-- http://datasrc.tipdm.net:81/bigdata/hadoop/software/hadoop-2.6.4.tar.gz
正在解析主机 datasrc.tipdm.net... 192.168.1.158
正在连接 datasrc.tipdm.net|192.168.1.158|:81... 已连接。
已发出 HTTP 请求，正在等待回应... 200 OK
长度：196015975 (187M) [application/x-gzip]
正在保存至：'/opt/software/hadoop-2.6.4.tar.gz'

100%[=====] 196,015,975 1.19G/s in 0.2s

2022-09-23 09:36:29 (1.19 GB/s) - 已保存 '/opt/software/hadoop-2.6.4.tar.gz' [196015975/196015975]

[ root@linux-master-0 ~]# wget -P /opt/software http://datasrc.tipdm.net:81/bigdata/hadoop/software/jdk-8u151-linux-x64.rpm
--2022-09-23 09:38:22-- http://datasrc.tipdm.net:81/bigdata/hadoop/software/jdk-8u151-linux-x64.rpm
正在解析主机 datasrc.tipdm.net... 192.168.1.158
正在连接 datasrc.tipdm.net|192.168.1.158|:81... 已连接。
已发出 HTTP 请求，正在等待回应... 200 OK
长度：174163338 (166M)
正在保存至：'/opt/software/jdk-8u151-linux-x64.rpm'

100%[=====] 174,163,338 --.-K/s in 0.1s

2022-09-23 09:38:22 (1.33 GB/s) - 已保存 '/opt/software/jdk-8u151-linux-x64.rpm' [174163338/174163338]

[ root@linux-master-0 ~]#

```

图 2

如图 3 所示，首先将 Java 安装包发送到子节点上，然后在终端输入命令 `java -version`。由此检测到当前环境并没有安装 Java。因此，进入 `/opt/software` 目录安装 JDK。

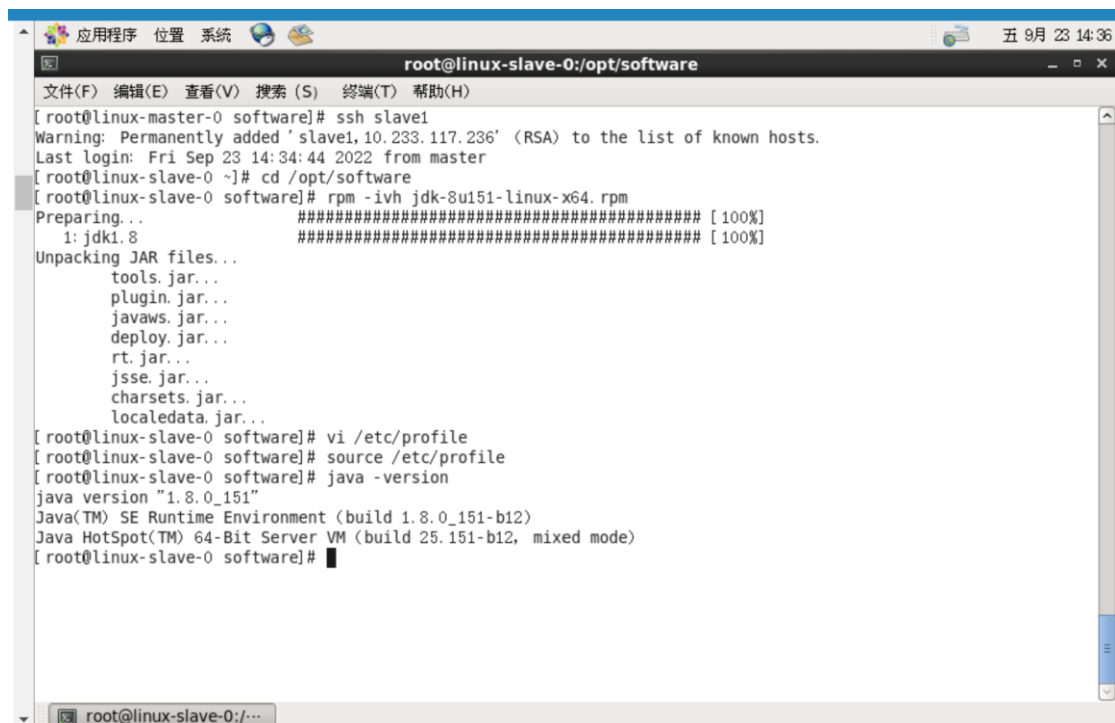
```

> 实训1 Java安装及Hadoop完全分布式集群搭建 实训用时剩余: 84分36秒 延时 隐藏 同屏 全屏
root@linux-master-0:/opt/software
[ root@linux-master-0 ~]# ssh slave1 "mkdir /opt/software"
Warning: Permanently added 'slave1,10.233.117.49' (RSA) to the list of known hosts.
[ root@linux-master-0 ~]# ssh slave2 "mkdir /opt/software"
Warning: Permanently added 'slave2,10.233.95.128' (RSA) to the list of known hosts.
[ root@linux-master-0 ~]# scp /opt/software/jdk-8u151-linux-x64.rpm slave1:/opt/software/
Warning: Permanently added 'slave1,10.233.117.49' (RSA) to the list of known hosts.
jdk-8u151-linux-x64.rpm 100% 166MB 166.1MB/s 00:01
[ root@linux-master-0 ~]# scp /opt/software/jdk-8u151-linux-x64.rpm slave2:/opt/software/
Warning: Permanently added 'slave2,10.233.95.128' (RSA) to the list of known hosts.
jdk-8u151-linux-x64.rpm 100% 166MB 166.1MB/s 00:01
[ root@linux-master-0 ~]# java -version
bash: java: command not found
[ root@linux-master-0 ~]# cd /opt/software/
[ root@linux-master-0 software]# rpm -ivh jdk-8u151-linux-x64.rpm
Preparing... ##### [100%]
1: jdk1.8 ##### [100%]
Unpacking JAR files...
tools.jar...
plugin.jar...
javaws.jar...
deploy.jar...
rt.jar...
jsse.jar...
charsets.jar...
localdata.jar...

```

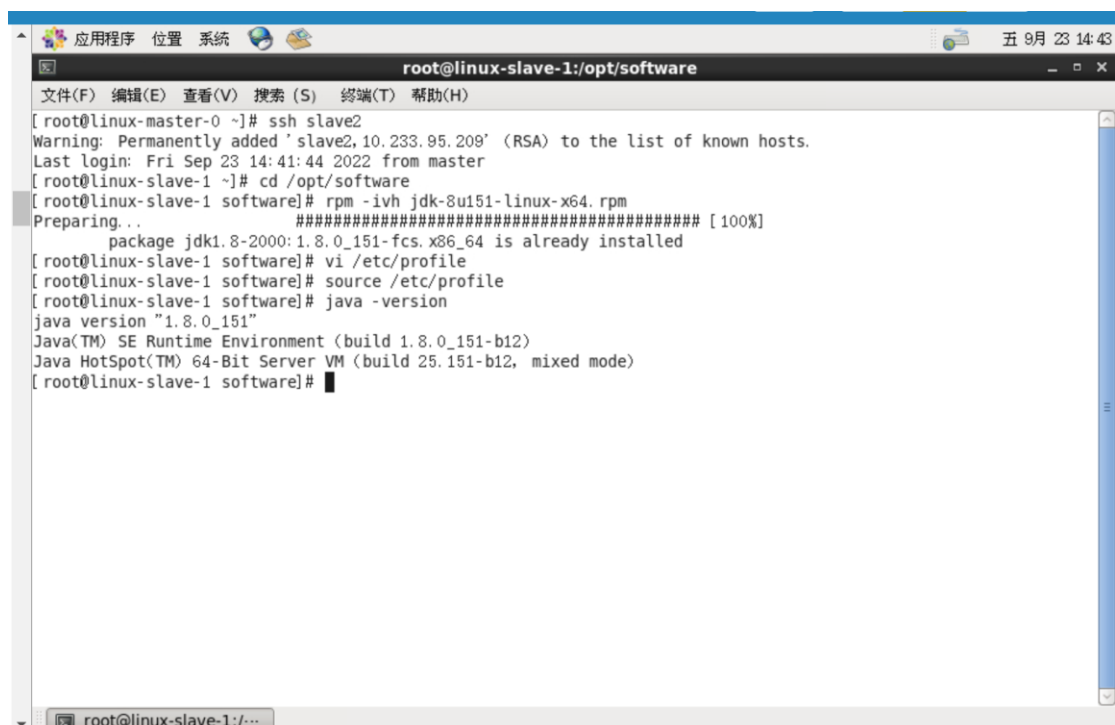
图 3

通过 vi 命令修改/etc/profile 文件，在 master 节点添加 Java 的环境变量。然后，在两个子节点 slave1、slave2 上面执行相同的步骤，安装 Java 并添加环境变量，完成所要求版本 JDK 的安装，如图 4、图 5 所示。



```
root@linux-slave-0:/opt/software
[ root@linux-master-0 software]# ssh slave1
Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
Last login: Fri Sep 23 14:34:44 2022 from master
[ root@linux-slave-0 ~]# cd /opt/software
[ root@linux-slave-0 software]# rpm -ivh jdk-8u151-linux-x64.rpm
Preparing... ##### [ 100%]
1: jdk1.8 ##### [ 100%]
Unpacking JAR files...
  tools.jar...
  plugin.jar...
  javaws.jar...
  deploy.jar...
  rt.jar...
  jsse.jar...
  charsets.jar...
  localedata.jar...
[ root@linux-slave-0 software]# vi /etc/profile
[ root@linux-slave-0 software]# source /etc/profile
[ root@linux-slave-0 software]# java -version
java version "1.8.0_151"
Java(TM) SE Runtime Environment (build 1.8.0_151-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.151-b12, mixed mode)
[ root@linux-slave-0 software]#
```

图 4

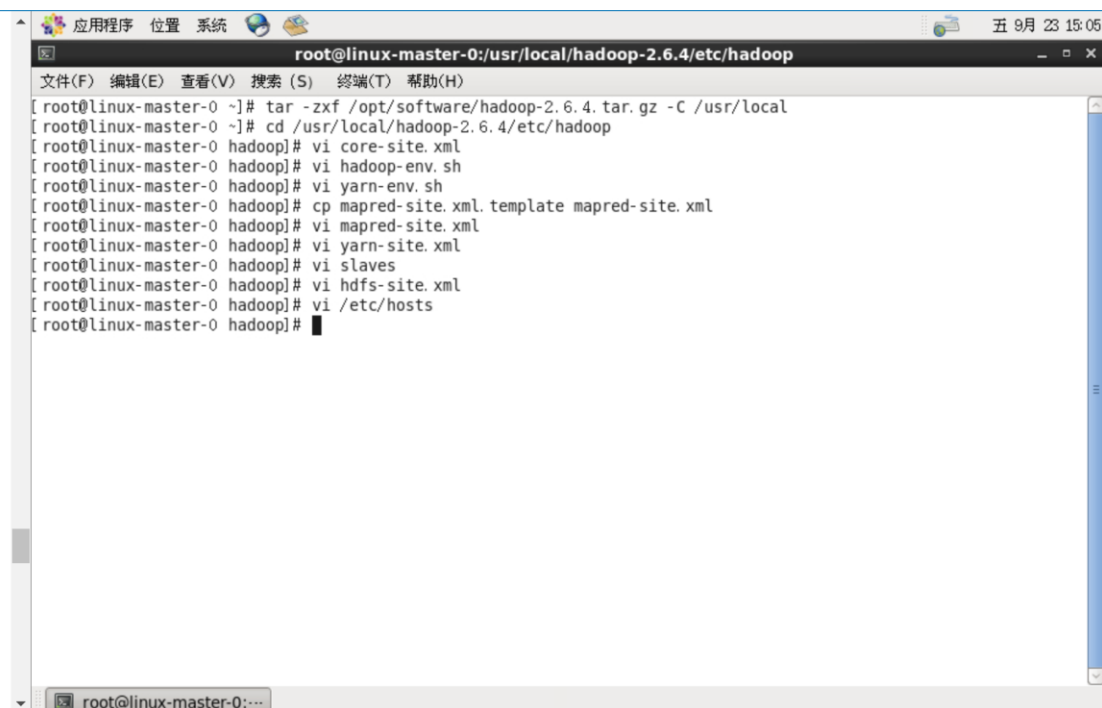


```
root@linux-slave-1:/opt/software
[ root@linux-master-0 ~]# ssh slave2
Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
Last login: Fri Sep 23 14:41:44 2022 from master
[ root@linux-slave-1 ~]# cd /opt/software
[ root@linux-slave-1 software]# rpm -ivh jdk-8u151-linux-x64.rpm
Preparing... ##### [ 100%]
package jdk1.8-2000:1.8.0_151-fcs.x86_64 is already installed
[ root@linux-slave-1 software]# vi /etc/profile
[ root@linux-slave-1 software]# source /etc/profile
[ root@linux-slave-1 software]# java -version
java version "1.8.0_151"
Java(TM) SE Runtime Environment (build 1.8.0_151-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.151-b12, mixed mode)
[ root@linux-slave-1 software]#
```

图 5

2、修改 Hadoop 配置文件。

首先执行 `tar -zxf /opt/software/hadoop-2.6.4.tar.gz -C /usr/local`，将 Hadoop 安装包解压至 master 的 `/usr/local` 目录下，随后进入 `/usr/local/hadoop-2.6.4/etc/hadoop` 目录，用 `vi` 命令修改相关配置文件，如图 6 所示。



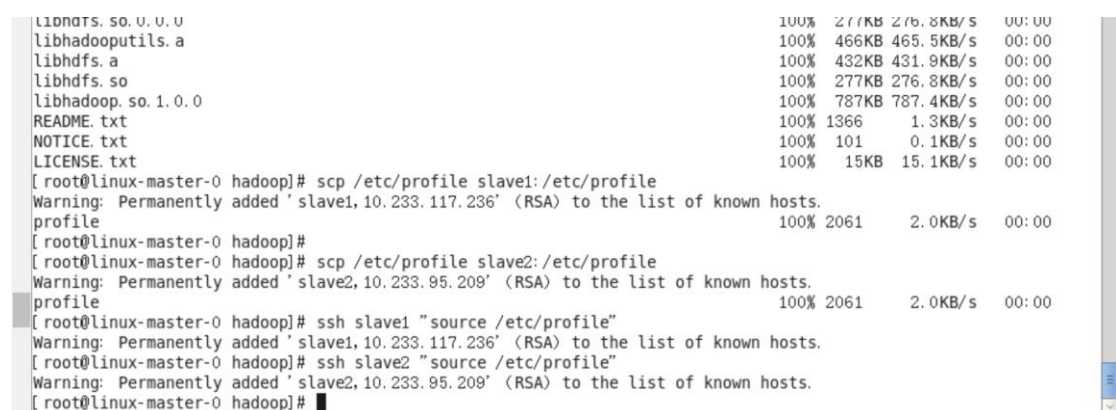
```

root@linux-master-0: /usr/local/hadoop-2.6.4/etc/hadoop
[ root@linux-master-0 ~]# tar -zxf /opt/software/hadoop-2.6.4.tar.gz -C /usr/local
[ root@linux-master-0 ~]# cd /usr/local/hadoop-2.6.4/etc/hadoop
[ root@linux-master-0 hadoop]# vi core-site.xml
[ root@linux-master-0 hadoop]# vi hadoop-env.sh
[ root@linux-master-0 hadoop]# vi yarn-env.sh
[ root@linux-master-0 hadoop]# cp mapred-site.xml.template mapred-site.xml
[ root@linux-master-0 hadoop]# vi mapred-site.xml
[ root@linux-master-0 hadoop]# vi yarn-site.xml
[ root@linux-master-0 hadoop]# vi slaves
[ root@linux-master-0 hadoop]# vi hdfs-site.xml
[ root@linux-master-0 hadoop]# vi /etc/hosts
[ root@linux-master-0 hadoop]#

```

图 6

随后，在 master 上修改 `/etc/profile` 文件，以配置 Hadoop 的环境变量，使用命令 `source /etc/profile` 使配置生效。将 Hadoop 安装包、`/etc/profile` 文件复制传输到 `slave1`、`slave2` 节点，如图 7 所示。



```

libnfs.so.0.0.0          100% 277KB 276.8KB/s  00:00
libhadooputils.a         100% 466KB 465.5KB/s  00:00
libhdfs.a                100% 432KB 431.9KB/s  00:00
libhdfs.so               100% 277KB 276.8KB/s  00:00
libhadoop.so.1.0.0       100% 787KB 787.4KB/s  00:00
README.txt               100% 1366   1.3KB/s  00:00
NOTICE.txt               100% 101    0.1KB/s  00:00
LICENSE.txt              100% 15KB   15.1KB/s  00:00
[ root@linux-master-0 hadoop]# scp /etc/profile slave1:/etc/profile
Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
profile                  100% 2061   2.0KB/s  00:00
[ root@linux-master-0 hadoop]# scp /etc/profile slave2:/etc/profile
Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
profile                  100% 2061   2.0KB/s  00:00
[ root@linux-master-0 hadoop]# ssh slave1 "source /etc/profile"
Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
[ root@linux-master-0 hadoop]# ssh slave2 "source /etc/profile"
Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
[ root@linux-master-0 hadoop]#

```

图 7

在 master 上执行命令 `hdfs namenode -format` 以进行格式化，观察到终端输出提示 Storage directory `hadoop/hdfs/name` has been successfully formatted，说明已经格式化成功。如图 8 所示。

```
22/09/23 15:18:09 INFO namenode.FSImage: Allocated new BlockPoolId: BP-2078097544-10.233.116.187-1663917489818
22/09/23 15:18:09 INFO common.Storage: Storage directory /hadoop/hdfs/name has been successfully formatted.
22/09/23 15:18:10 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
22/09/23 15:18:10 INFO util.ExitUtil: Exiting with status 0
22/09/23 15:18:10 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at master/10.233.116.187
*****/
[root@linux-master-0 hadoop]#
```

图 8

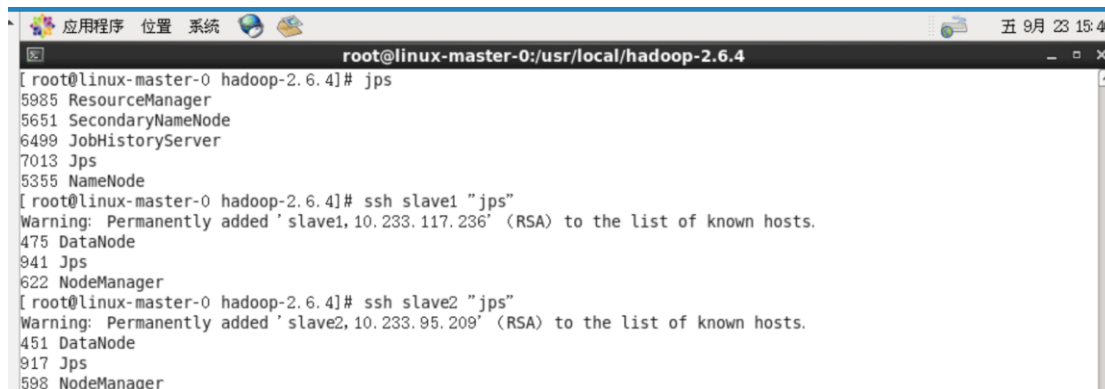
3.启动、关闭 Hadoop 集群

在 master 节点，进入 Hadoop 安装目录，启动 HDFS 相关服务、YARN 相关服务和日志相关服务，如图 9 所示。

```
root@linux-master-0:/usr/local/hadoop-2.6.4
[root@linux-master-0 hadoop]# cd $HADOOP_HOME
[root@linux-master-0 hadoop-2.6.4]# sbin/start-dfs.sh
22/09/23 15:31:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.. using bui
ltin-java classes where applicable
Starting namenodes on [master]
master: Warning: Permanently added 'master,10.233.116.187' (RSA) to the list of known hosts.
master: starting namenode, logging to /usr/local/hadoop-2.6.4/logs/hadoop-root-namenode-linux-master-0.out
slave2: Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
slave1: Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
slave2: starting datanode, logging to /usr/local/hadoop-2.6.4/logs/hadoop-root-datanode-linux-slave-1.out
slave1: starting datanode, logging to /usr/local/hadoop-2.6.4/logs/hadoop-root-datanode-linux-slave-0.out
Starting secondary namenodes [master]
master: Warning: Permanently added 'master,10.233.116.187' (RSA) to the list of known hosts.
master: starting secondarynamenode, logging to /usr/local/hadoop-2.6.4/logs/hadoop-root-secondarynamenode-linux-m
aster-0.out
22/09/23 15:31:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.. using bui
ltin-java classes where applicable
[root@linux-master-0 hadoop-2.6.4]# sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.6.4/logs/yarn-root-resourcemanager-linux-master-0.out
slave1: Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
slave2: Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
slave1: starting nodemanager, logging to /usr/local/hadoop-2.6.4/logs/yarn-root-nodemanager-linux-slave-0.out
slave2: starting nodemanager, logging to /usr/local/hadoop-2.6.4/logs/yarn-root-nodemanager-linux-slave-1.out
[root@linux-master-0 hadoop-2.6.4]# sbin/mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /usr/local/hadoop-2.6.4/logs/mapred-root-historyserver-linux-master-0.out
```

图 9

集群启动之后，在主节点 master，子节点 slave1、slave2 分别执行 `jps`，结果如图 10 所示。这说明集群的启动没有问题。



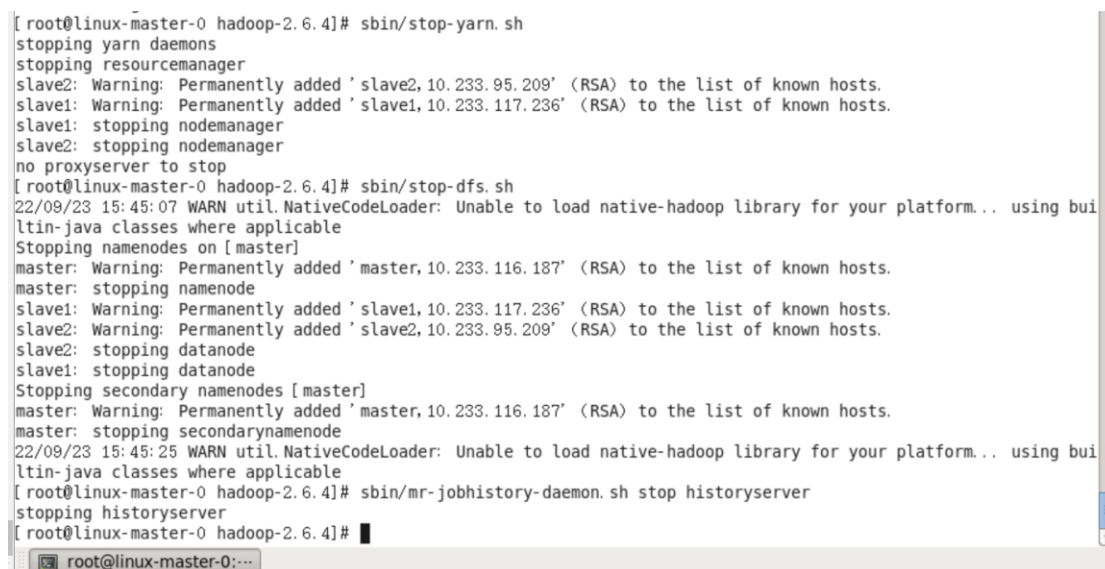
```

root@linux-master-0:/usr/local/hadoop-2.6.4
[root@linux-master-0 hadoop-2.6.4]# jps
5985 ResourceManager
5651 SecondaryNameNode
6499 JobHistoryServer
7013 Jps
5355 NameNode
[root@linux-master-0 hadoop-2.6.4]# ssh slave1 "jps"
Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
475 DataNode
941 Jps
622 NodeManager
[root@linux-master-0 hadoop-2.6.4]# ssh slave2 "jps"
Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
451 DataNode
917 Jps
598 NodeManager

```

图 10

对 Hadoop 集群进行关闭，如图 11 所示。



```

[root@linux-master-0 hadoop-2.6.4]# sbin/stop-yarn.sh
stopping yarn daemons
stopping resourcemanager
slave2: Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
slave1: Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
slave1: stopping nodemanager
slave2: stopping nodemanager
no proxyserver to stop
[root@linux-master-0 hadoop-2.6.4]# sbin/stop-dfs.sh
22/09/23 15:45:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bui
ltin-java classes where applicable
Stopping namenodes on [master]
master: Warning: Permanently added 'master,10.233.116.187' (RSA) to the list of known hosts.
master: stopping namenode
slave1: Warning: Permanently added 'slave1,10.233.117.236' (RSA) to the list of known hosts.
slave2: Warning: Permanently added 'slave2,10.233.95.209' (RSA) to the list of known hosts.
slave2: stopping datanode
slave1: stopping datanode
Stopping secondary namenodes [master]
master: Warning: Permanently added 'master,10.233.116.187' (RSA) to the list of known hosts.
master: stopping secondarynamenode
22/09/23 15:45:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bui
ltin-java classes where applicable
[root@linux-master-0 hadoop-2.6.4]# sbin/mr-jobhistory-daemon.sh stop historyserver
stopping historyserver
[root@linux-master-0 hadoop-2.6.4]#

```

图 11

3、使用 MapReduce 完成 WordCount 程序任务。

如图 12 所示，进入“实训 3 运行首个 MapReduce 任务”实训环境，首先将预处理后得到的文本上传至实训环境的/data 目录下。



图 12

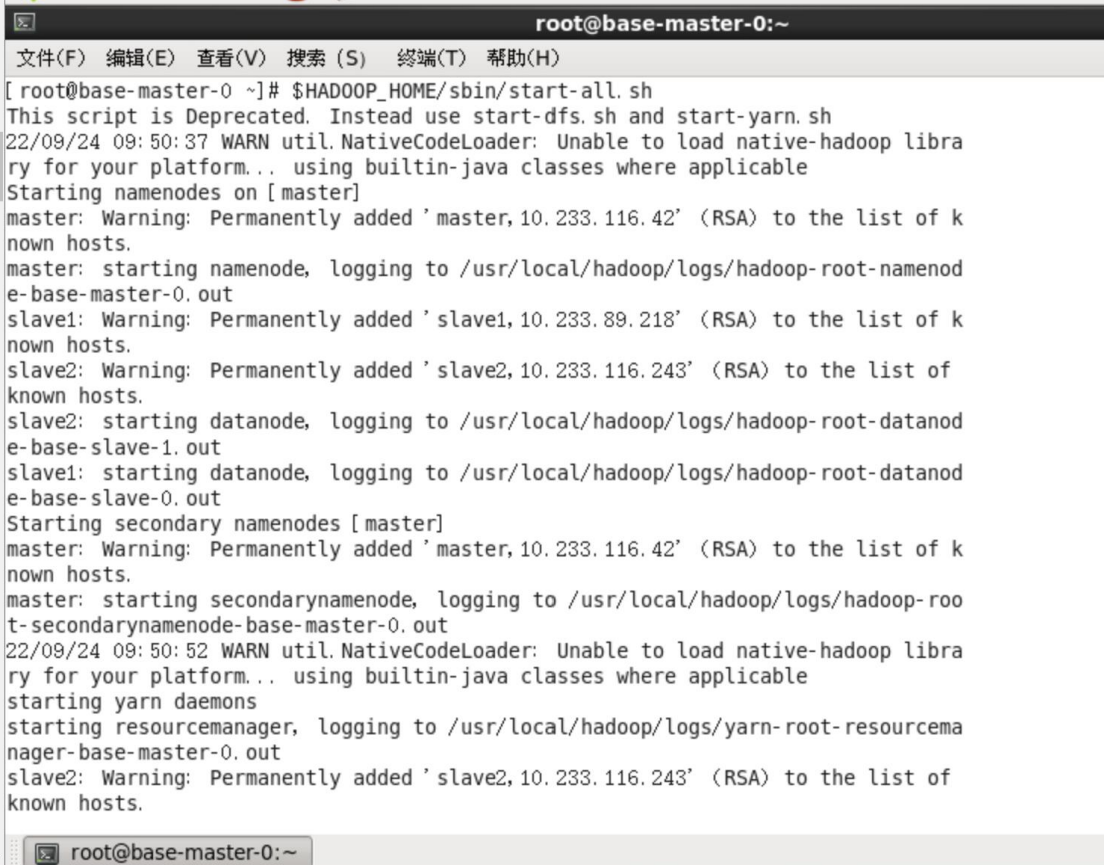
如图 13、图 14 所示，打开终端，使用命令 `initnetwork` 初始化网络，并启动 Hadoop 集群。

```

[root@base-master-0 ~]# initnetwork
正在初始化网络....
Warning: Permanently added 'base-slave-0, base-slave, 10.233.116.216' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-1, base-slave, 10.233.99.177' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-0, base-slave, 10.233.116.216' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-1, base-slave, 10.233.99.177' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-0, base-slave, 10.233.116.216' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-1, base-slave, 10.233.99.177' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-0, base-slave, 10.233.116.216' (RSA) to the
list of known hosts.
Warning: Permanently added 'base-slave-1, base-slave, 10.233.99.177' (RSA) to the
list of known hosts.
Warning: Permanently added 'slave1, 10.233.116.216' (RSA) to the list of known ho
sts.
Warning: Permanently added 'slave2, 10.233.99.177' (RSA) to the list of known hos
ts.
初始化网络完成

```

图 13



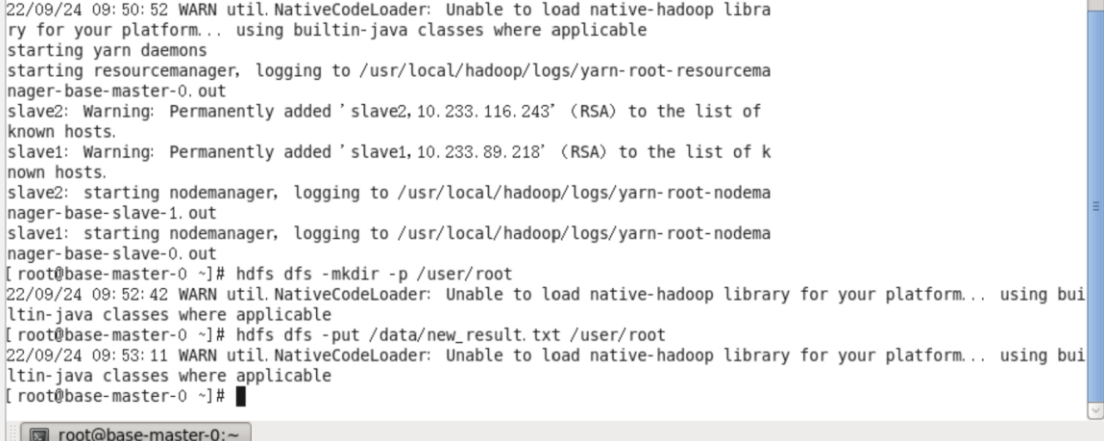
```

root@base-master-0:~
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
[root@base-master-0 ~]# $HADOOP_HOME/sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
22/09/24 09:50:37 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [master]
master: Warning: Permanently added 'master, 10.233.116.42' (RSA) to the list of k
nown hosts.
master: starting namenode, logging to /usr/local/hadoop/logs/hadoop-root-namenod
e-base-master-0.out
slave1: Warning: Permanently added 'slave1, 10.233.89.218' (RSA) to the list of k
nown hosts.
slave2: Warning: Permanently added 'slave2, 10.233.116.243' (RSA) to the list of
known hosts.
slave2: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanod
e-base-slave-1.out
slave1: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanod
e-base-slave-0.out
Starting secondary namenodes [master]
master: Warning: Permanently added 'master, 10.233.116.42' (RSA) to the list of k
nown hosts.
master: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-roo
t-secondarynamenode-base-master-0.out
22/09/24 09:50:52 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-root-resourcem
anager-base-master-0.out
slave2: Warning: Permanently added 'slave2, 10.233.116.243' (RSA) to the list of
known hosts.
root@base-master-0:~

```

图 14

如图 15 所示，首先在 HDFS 新建目录/user/root，然后将/data 目录下的待统计文本文件 new_result.txt 上传至 HDFS。



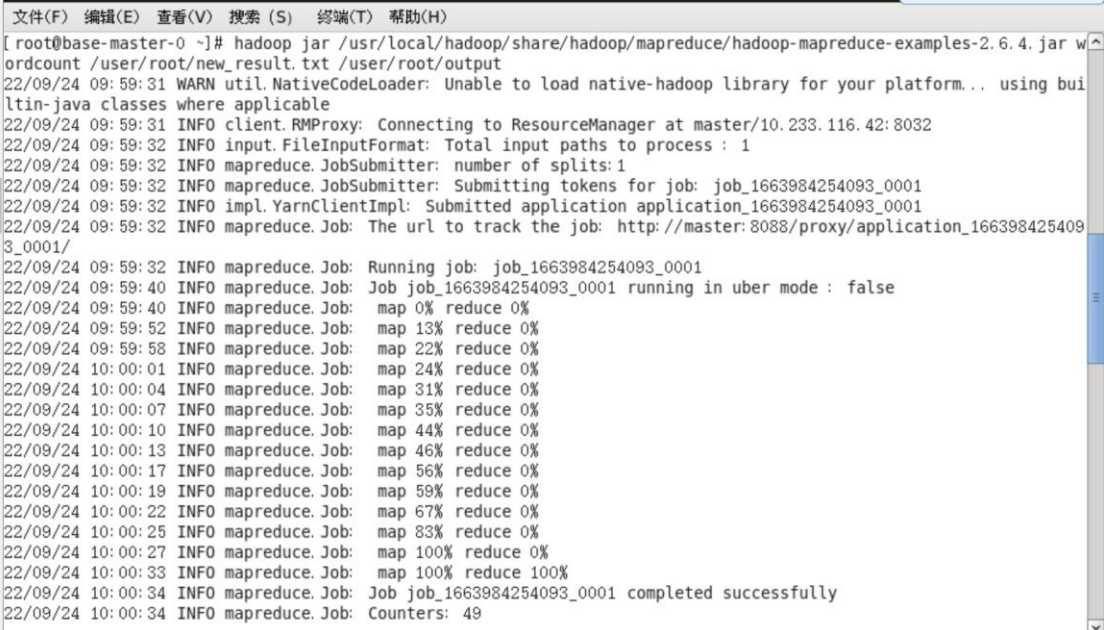
```

22/09/24 09:50:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-root-resourcemanager-base-master-0.out
slave2: Warning: Permanently added 'slave2,10.233.116.243' (RSA) to the list of known hosts.
slave1: Warning: Permanently added 'slave1,10.233.89.218' (RSA) to the list of known hosts.
slave2: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-base-slave-1.out
slave1: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-root-nodemanager-base-slave-0.out
[root@base-master-0 ~]# hdfs dfs -mkdir -p /user/root
22/09/24 09:52:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@base-master-0 ~]# hdfs dfs -put /data/new_result.txt /user/root
22/09/24 09:53:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@base-master-0 ~]#

```

图 15

如图 16 所示，提交 MapReduce 任务，对所上传的 new_result.txt 文件执行词频统计程序。



```

文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
[root@base-master-0 ~]# hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.4.jar wordcount /user/root/new_result.txt /user/root/output
22/09/24 09:59:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
22/09/24 09:59:31 INFO client.RMProxy: Connecting to ResourceManager at master/10.233.116.42:8032
22/09/24 09:59:32 INFO input.FileInputFormat: Total input paths to process : 1
22/09/24 09:59:32 INFO mapreduce.JobSubmitter: number of splits:1
22/09/24 09:59:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1663984254093_0001
22/09/24 09:59:32 INFO impl.YarnClientImpl: Submitted application application_1663984254093_0001
22/09/24 09:59:32 INFO mapreduce.Job: The url to track the job: http://master:8088/proxy/application_1663984254093_0001/
22/09/24 09:59:32 INFO mapreduce.Job: Running job: job_1663984254093_0001
22/09/24 09:59:40 INFO mapreduce.Job: Job job_1663984254093_0001 running in uber mode : false
22/09/24 09:59:40 INFO mapreduce.Job: map 0% reduce 0%
22/09/24 09:59:52 INFO mapreduce.Job: map 13% reduce 0%
22/09/24 09:59:58 INFO mapreduce.Job: map 22% reduce 0%
22/09/24 10:00:01 INFO mapreduce.Job: map 24% reduce 0%
22/09/24 10:00:04 INFO mapreduce.Job: map 31% reduce 0%
22/09/24 10:00:07 INFO mapreduce.Job: map 35% reduce 0%
22/09/24 10:00:10 INFO mapreduce.Job: map 44% reduce 0%
22/09/24 10:00:13 INFO mapreduce.Job: map 46% reduce 0%
22/09/24 10:00:17 INFO mapreduce.Job: map 56% reduce 0%
22/09/24 10:00:19 INFO mapreduce.Job: map 59% reduce 0%
22/09/24 10:00:22 INFO mapreduce.Job: map 67% reduce 0%
22/09/24 10:00:25 INFO mapreduce.Job: map 83% reduce 0%
22/09/24 10:00:27 INFO mapreduce.Job: map 100% reduce 0%
22/09/24 10:00:33 INFO mapreduce.Job: map 100% reduce 100%
22/09/24 10:00:34 INFO mapreduce.Job: Job job_1663984254093_0001 completed successfully
22/09/24 10:00:34 INFO mapreduce.Job: Counters: 49

```

图 16

如图 17 所示，可以在/user/root/output 目录下观察到新生成的文件。其中文件_SUCCESS 表示任务执行完成，而 part-r-00000 则为任务执行完成后产生的结果文件。

Browse Directory

<input type="text" value="/user/root/output"/>						<input data-bbox="1257 342 1289 365" type="button" value="Go!"/>
Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	root	supergroup	0 B	3	128 MB	_SUCCESS
-rw-r--r--	root	supergroup	2.89 MB	3	128 MB	part-r-00000

Hadoop, 2014.

图 17

下载 `part-r-00000` 文件（如图 18），并更名为 `statistical_result.txt`，即可得到分词统计结果。将该文件从实训环境中导出，并在 Visual Studio Code 中查看，结果如图 19 所示。

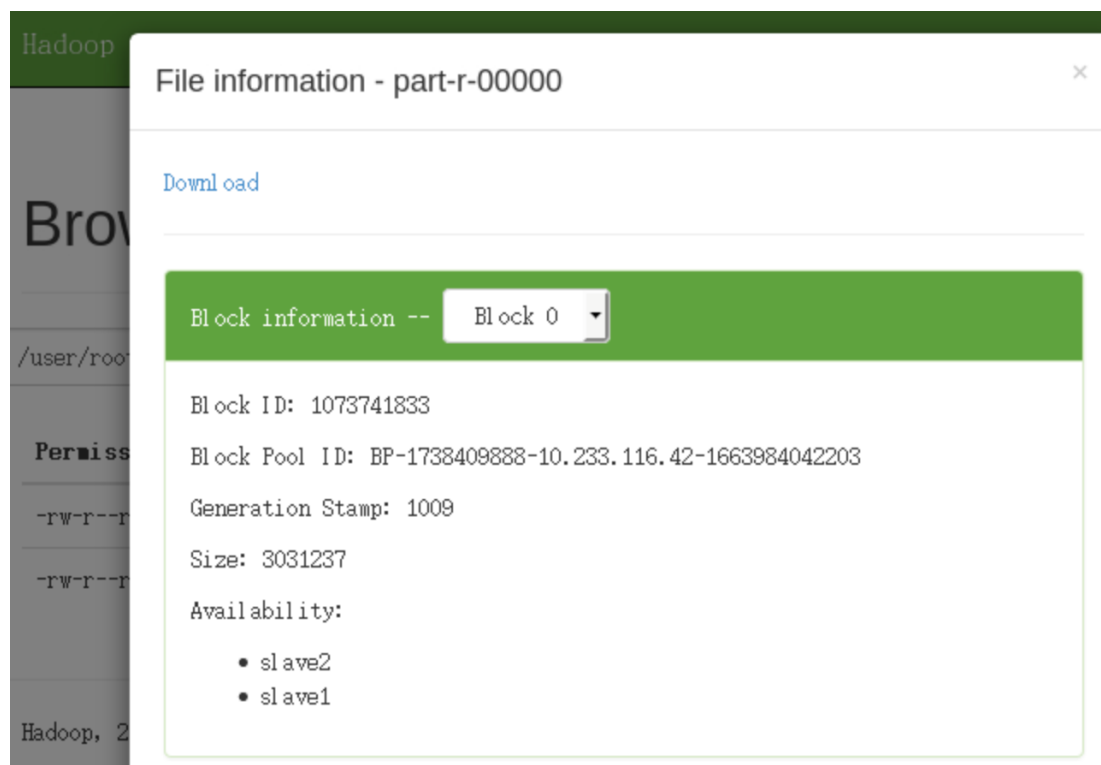


图 18

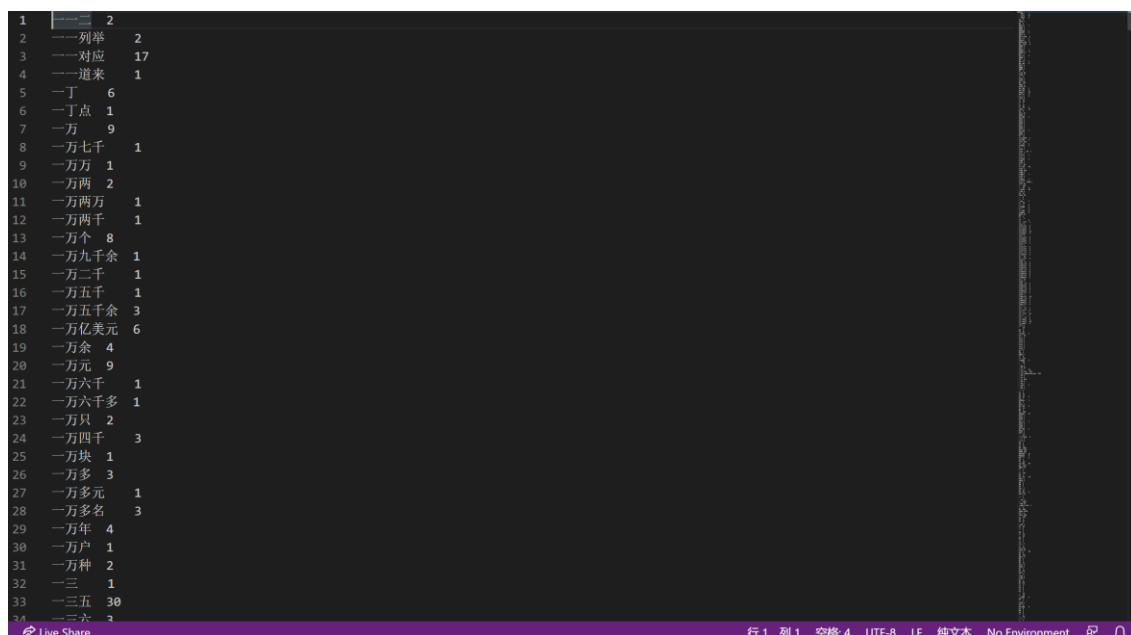


图 19

五、实验结果

WordCount 程序任务的完整运行命令、执行结果如下图 20、21、22 所示。

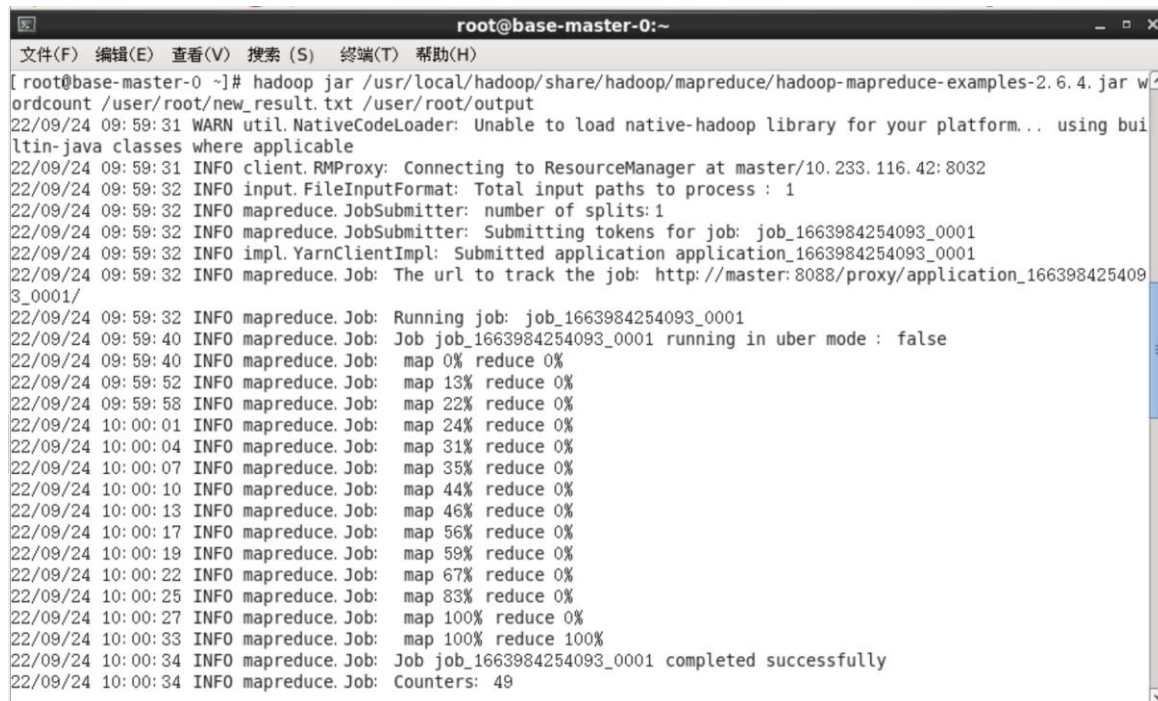
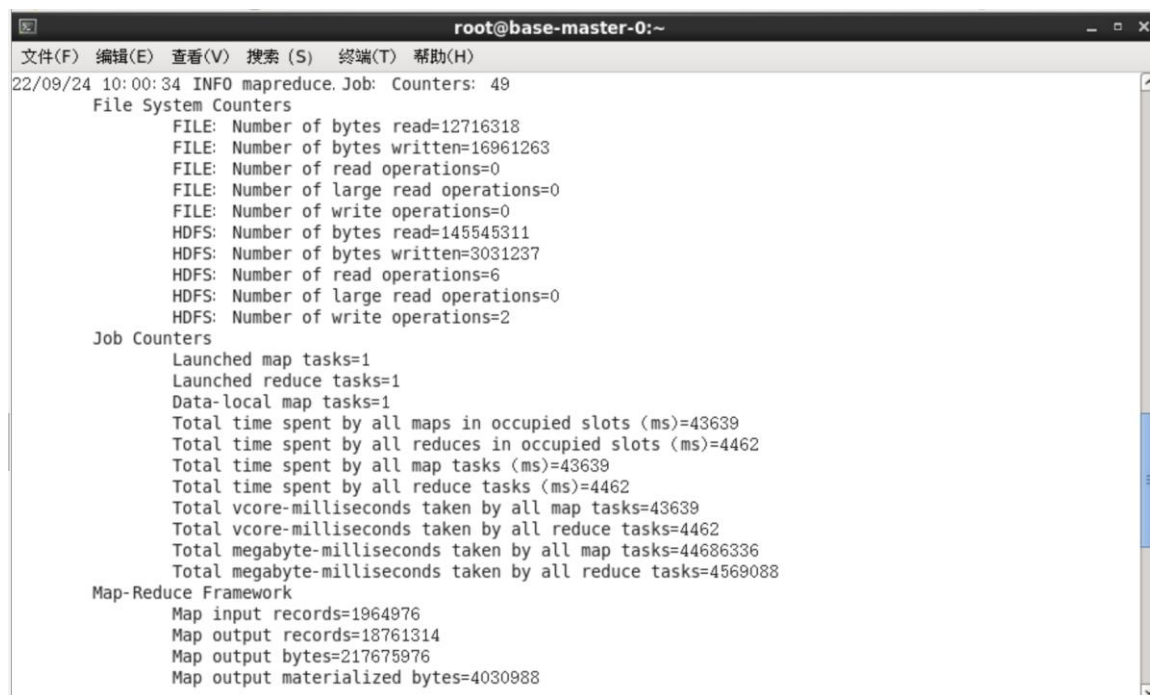


图 20

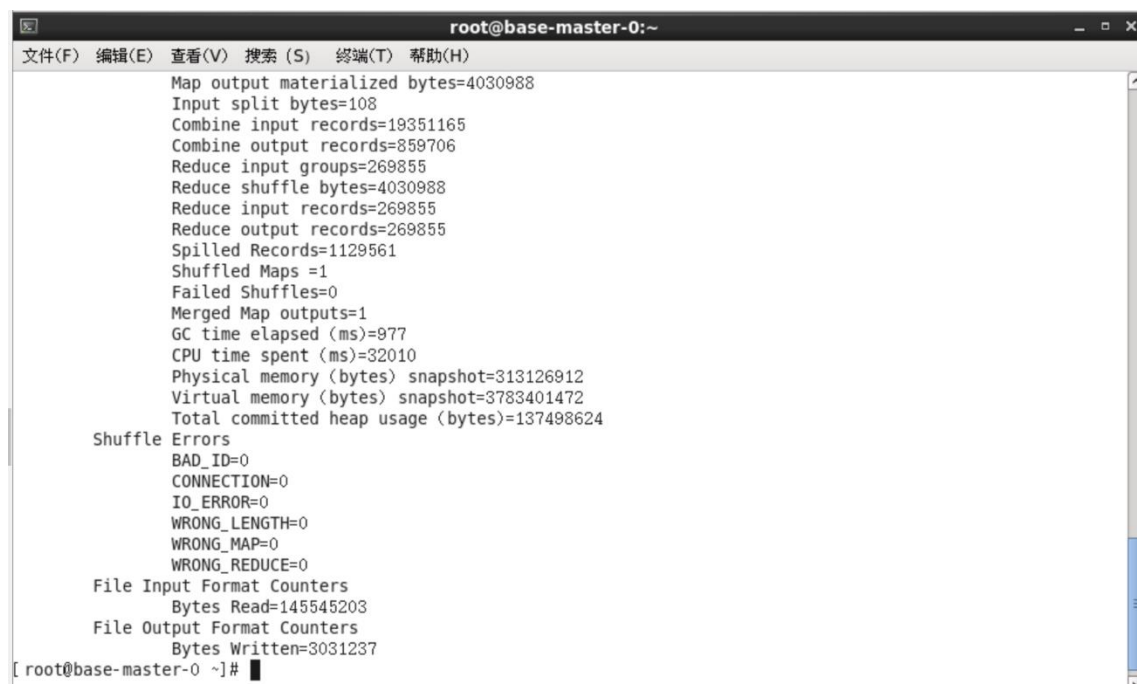


```

root@base-master-0:~
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
22/09/24 10:00:34 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=12716318
  FILE: Number of bytes written=16961263
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=145545311
  HDFS: Number of bytes written=3031237
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=43639
  Total time spent by all reduces in occupied slots (ms)=4462
  Total time spent by all map tasks (ms)=43639
  Total time spent by all reduce tasks (ms)=4462
  Total vcore-milliseconds taken by all map tasks=43639
  Total vcore-milliseconds taken by all reduce tasks=4462
  Total megabyte-milliseconds taken by all map tasks=44686336
  Total megabyte-milliseconds taken by all reduce tasks=4569088
Map-Reduce Framework
  Map input records=1964976
  Map output records=18761314
  Map output bytes=217675976
  Map output materialized bytes=4030988

```

图 21



```

root@base-master-0:~
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
  Map output materialized bytes=4030988
  Input split bytes=108
  Combine input records=19351165
  Combine output records=859706
  Reduce input groups=269855
  Reduce shuffle bytes=4030988
  Reduce input records=269855
  Reduce output records=269855
  Spilled Records=1129561
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=977
  CPU time spent (ms)=32010
  Physical memory (bytes) snapshot=313126912
  Virtual memory (bytes) snapshot=3783401472
  Total committed heap usage (bytes)=137498624
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=145545203
File Output Format Counters
  Bytes Written=3031237
[ root@base-master-0 ~]#

```

图 22

文件 `statistical_result.txt` 存储了 WordCount 程序的运行输出（使用 UTF-8 编码）。部分词频统计结果如图 23、24 所示。

9251	中共中央政治局	2165
9252	中共中央政治局常委会	23
9253	中共中央政法委员会	15
9254	中共中央文献研究室	5
9255	中共中央纪委	12
9256	中共中央纪律检查委员会	66
9257	中共中央组织部	162
9258	中共中央统战部	47
9259	中共二大	24
9260	中共云南省委	28
9261	中共党史	134
9262	中共党员	231
9263	中共八大	10
9264	中共六届六中全会	2
9265	中共军队	1
9266	中共北京市委	41
9267	中共十一届三中全会	24
9268	中共十五大	1
9269	中共南京市委	11
9270	中共四川省委	64
9271	中共天津市委	13
9272	中共安徽省委	95
9273	中共山东省委	44
9274	中共广东	11
9275	中共广东省委	75
9276	中共江苏省委	20
9277	中共江西省委	49
9278	中共河南省委	69
9279	中共浙江省委	42
9280	中共湖北省委	40
9281	中共湖南省委	68
9282	中共湘区委员会	1
9283	中共甘肃省委	16
9284	中共福建省委	274

图 23

70416	国防委员会	8
70417	国防工业	34
70418	国防建设	113
70419	国防报	36
70420	国防报印海军	1
70421	国防报图片	1
70422	国防报美海军	1
70423	国防教育	1102
70424	国防法	11
70425	国防生	29
70426	国防科学技术大学	3
70427	国防科学技术工业委员会	2
70428	国防科工办	17
70429	国防科工委	10
70430	国防科技	1256
70431	国防科技大学	63
70432	国防绿	4
70433	国防部	1608
70434	国防部长	76
70435	国际	30524
70436	国际主义	16
70437	国际主义者	3
70438	国际乒联	3
70439	国际争端	7
70440	国际事务	168
70441	国际交往	74
70442	国际交流	1729
70443	国际会计	2
70444	国际会议中心	25
70445	国际公法	9
70446	国际公约	43
70447	国际共产主义运动	12
70448	国际共运	5
70449	国际关系学院	38

图 24