

Hadoop2.X 64 位环境搭建

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，博主为石山园，博客地址为 <http://www.cnblogs.com/shishanyuan> 。该系列课程是应邀实验楼整理编写的，这里需要赞一下实验楼提供了学习的新方式，可以边看博客边上机实验，课程地址为 <https://www.shiyanlou.com/courses/237>

【注】该系列所使用到安装包、测试数据和代码均可在百度网盘下载，具体地址为 <http://pan.baidu.com/s/10PnDs>，下载该 PDF 文件

1 搭建环境

部署节点操作系统为 CentOS，防火墙和 SELinux 禁用，创建了一个 shiyanlou 用户并在系统根目录下创建/app 目录，用于存放 Hadoop 等组件运行包。因为该目录用于安装 hadoop 等组件程序，用户对 shiyanlou 必须赋予 rwx 权限（一般做法是 root 用户在根目录下创建/app 目录，并修改该目录拥有者为 shiyanlou(chown -R shiyanlou:shiyanlou /app)。

Hadoop 搭建环境：

- 虚拟机操作系统：CentOS6.6 64 位，单核，1G 内存
- JDK：1.7.0_55 64 位
- Hadoop：2.2.0 64 位（该部署包为第 2 个实验所编译完成）

2 部署 Hadoop2.X

2.1 配置 Hadoop 环境

在 Apache 网站上提供 Hadoop2.X 安装包只支持 32 位操作系统安装，在 64 位服务器安装会出现 3.1 的错误异常。这里我们使用上一步骤编译好的 hadoop-2.2.0-bin.tar.gz 文件作为安装包（也可以在/home/shiyanlou/install-pack 目录中找到 hadoop-2.2.0.tar.gz 安装包）

2.1.1 下载并解压 hadoop 安装包

解压缩并移动到/app 目录下

```
cd /home/shiyanlou/install-pack
```

```
tar -xzf hadoop-2.2.0.tar.gz
```

```
mv hadoop-2.2.0 /app
```

```
[shiyanolou@b393a04554e1 ~]$ cd /home/shiyanolou/install-pack
[shiyanolou@b393a04554e1 install-pack]$ tar -xzf hadoop-2.2.0.tar.gz
[shiyanolou@b393a04554e1 install-pack]$ ls
apache-maven-3.0.5-bin.tar.gz          jdk-7u55-linux-x64.tar.gz
chukwa-0.6.0.tar.gz                   mahout-distribution-0.6.tar.gz
eclipse-jee-luna-SR1-linux-gtk-x86_64.tar.gz MySQL-client-5.6.21-1.el6.x86_64.rpm
flume-1.5.2-bin.tar.gz                 mysql-connector-java-5.1.22-bin.jar
hadoop-1.1.2-bin.tar.gz                 MySQL-devel-5.6.21-1.el6.x86_64.rpm
hadoop-2.2.0                           MySQL-server-5.6.21-1.el6.x86_64.rpm
hadoop-2.2.0.tar.gz                    pig-0.13.0.tar.gz
hadoop-eclipse-plugin-1.1.2.jar         protobuf-2.5.0.tar.gz
hbase-0.96.2-hadoop1-bin.tar.gz        sqoop-1.4.5.bin__hadoop-1.0.0.tar.gz
hive-0.13.0-bin.tar.gz
[shiyanolou@b393a04554e1 install-pack]$ mv hadoop-2.2.0 /app
[shiyanolou@b393a04554e1 install-pack]$ ls /app
compile  hadoop-1.1.2  hadoop-2.2.0  lib
[shiyanolou@b393a04554e1 install-pack]$
```

2.1.2 在 Hadoop 目录下创建子目录

在 hadoop-2.2.0 目录下创建 tmp、name 和 data 目录

```
cd /app/hadoop-2.2.0
```

```
mkdir tmp
```

```
mkdir hdfs
```

```
mkdir hdfs/name
```

```
mkdir hdfs/data
```

```
[shiyanolou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0
[shiyanolou@b393a04554e1 hadoop-2.2.0]$ mkdir tmp
[shiyanolou@b393a04554e1 hadoop-2.2.0]$ mkdir hdfs
[shiyanolou@b393a04554e1 hadoop-2.2.0]$ mkdir hdfs/name
[shiyanolou@b393a04554e1 hadoop-2.2.0]$ mkdir hdfs/data
[shiyanolou@b393a04554e1 hadoop-2.2.0]$ ls
bin  hdfs  lib  LICENSE.txt  README.txt  share
etc  include  libexec  NOTICE.txt  sbin  tmp
```

2.1.3 配置 hadoop-env.sh

1. 打开配置文件 hadoop-env.sh

```
cd /app/hadoop-2.2.0/etc/hadoop
```

```
sudo vi hadoop-env.sh
```

```
[shiyanolou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0/etc/hadoop
[shiyanolou@b393a04554e1 hadoop]$ ls
capacity-scheduler.xml  hdfs-site.xml  mapred-site.xml.template
configuration.xml      httpfs-env.sh  slaves
container-executor.cfg  httpfs-log4j.properties  ssl-client.xml.example
core-site.xml           httpfs-signature.secret  ssl-server.xml.example
hadoop-env.cmd          httpfs-site.xml  yarn-env.cmd
hadoop-env.sh           log4j.properties     yarn-env.sh
hadoop-metrics2.properties  mapred-env.cmd  yarn-site.xml
hadoop-metrics.properties  mapred-env.sh
hadoop-policy.xml        mapred-queues.xml.template
[shiyanolou@b393a04554e1 hadoop]$ sudo vi hadoop-env.sh
```

2. 加入配置内容，设置了 hadoop 中 jdk 和 hadoop/bin 路径

```
export HADOOP_CONF_DIR=/app/hadoop2.2.0/etc/hadoop
```

```
export JAVA_HOME=/app/lib/jdk1.7.0_55
export PATH=$PATH:/app/hadoop-2.2.0/bin
```

```
# A string representing this instance of hadoop. $USER by default.
export HADOOP_IDENT_STRING=$USER

export HADOOP_CONF_DIR=/app/hadoop-2.2.0/etc/hadoop
export JAVA_HOME=/app/lib/jdk1.7.0_55
export PATH=$PATH:/app/hadoop-2.2.0/bin
```

3. 编译配置文件 `hadoop-env.sh` , 并确认生效

```
source hadoop-env.sh
hadoop version
```

```
[shiyanolou@b393a04554e1 hadoop]$ source hadoop-env.sh
[shiyanolou@b393a04554e1 hadoop]$ hadoop version
Hadoop 2.2.0
Subversion Unknown -r 1683027
Compiled by shiyanolou on 2015-06-02T02:30Z
Compiled with protoc 2.5.0
From source with checksum 79e53ce7994d1628b240f09af91e1af4
This command was run using /app/hadoop-2.2.0/share/hadoop/common/hadoop-common-2.2.0.jar
```

2.1.4 配置 yarn-env.sh

打开配置文件 `yarn-env.sh` , 设置了 `hadoop` 中 `jdk` 路径 , 配置完毕后使用 `source yarn-env.sh` 编译该文件

```
export JAVA_HOME=/app/lib/jdk1.7.0_55
```

```
YARN_OPTS="$YARN_OPTS -Djava.library.path=$JAVA_LIBRARY_PATH"
fi
YARN_OPTS="$YARN_OPTS -Dyarn.policy.file=$YARN_POLICYFILE"
export JAVA_HOME=/app/lib/jdk1.7.0_55
```

2.1.5 配置 core-site.xml

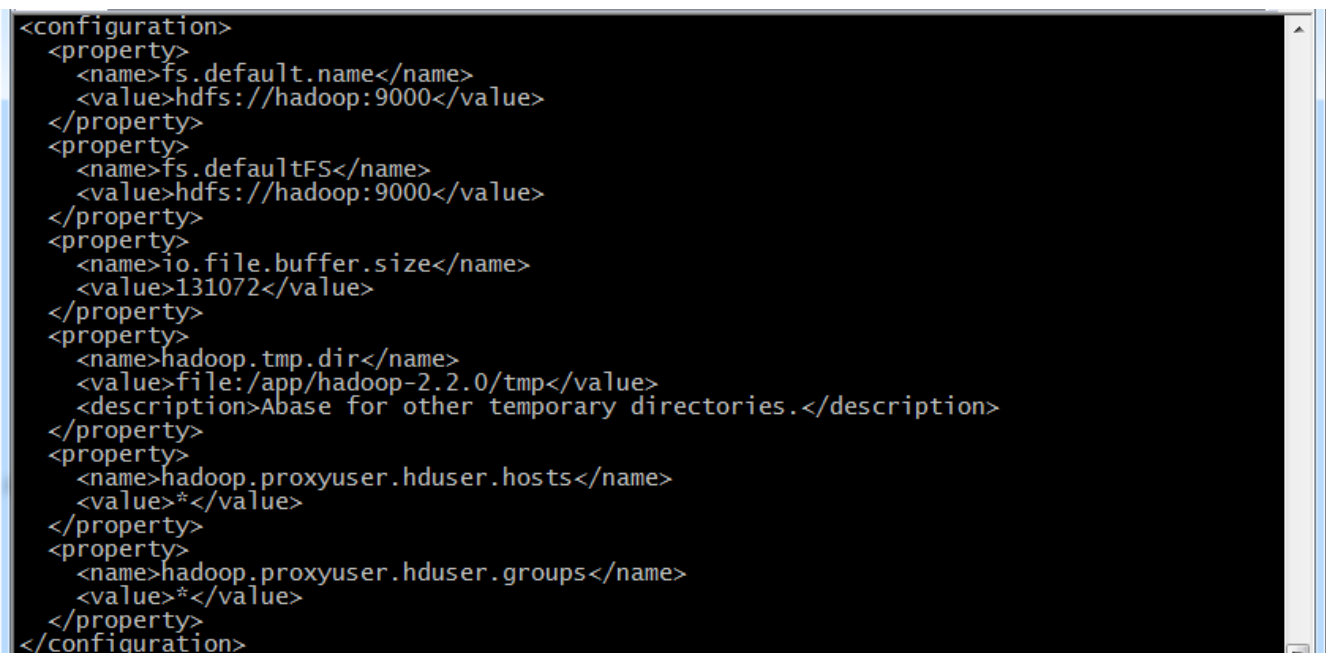
1. 使用如下命令打开 `core-site.xml` 配置文件

```
cd /app/hadoop-2.2.0/etc/hadoop
sudo vi core-site.xml
```

2. 在配置文件中 , 按照如下内容进行配置

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://hadoop:9000</value>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop:9000</value>
  </property>
```

```
<property>
  <name>io.file.buffer.size</name>
  <value>131072</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>file:/app/hadoop-2.2.0/tmp</value>
  <description>Abase for other temporary directories.</description>
</property>
<property>
  <name>hadoop.proxyuser.hduser.hosts</name>
  <value>*</value>
</property>
<property>
  <name>hadoop.proxyuser.hduser.groups</name>
  <value>*</value>
</property>
</configuration>
```



```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://hadoop:9000</value>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop:9000</value>
  </property>
  <property>
    <name>io.file.buffer.size</name>
    <value>131072</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/app/hadoop-2.2.0/tmp</value>
    <description>Abase for other temporary directories.</description>
  </property>
  <property>
    <name>hadoop.proxyuser.hduser.hosts</name>
    <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.hduser.groups</name>
    <value>*</value>
  </property>
</configuration>
```

2.1.6 配置 hdfs-site.xml

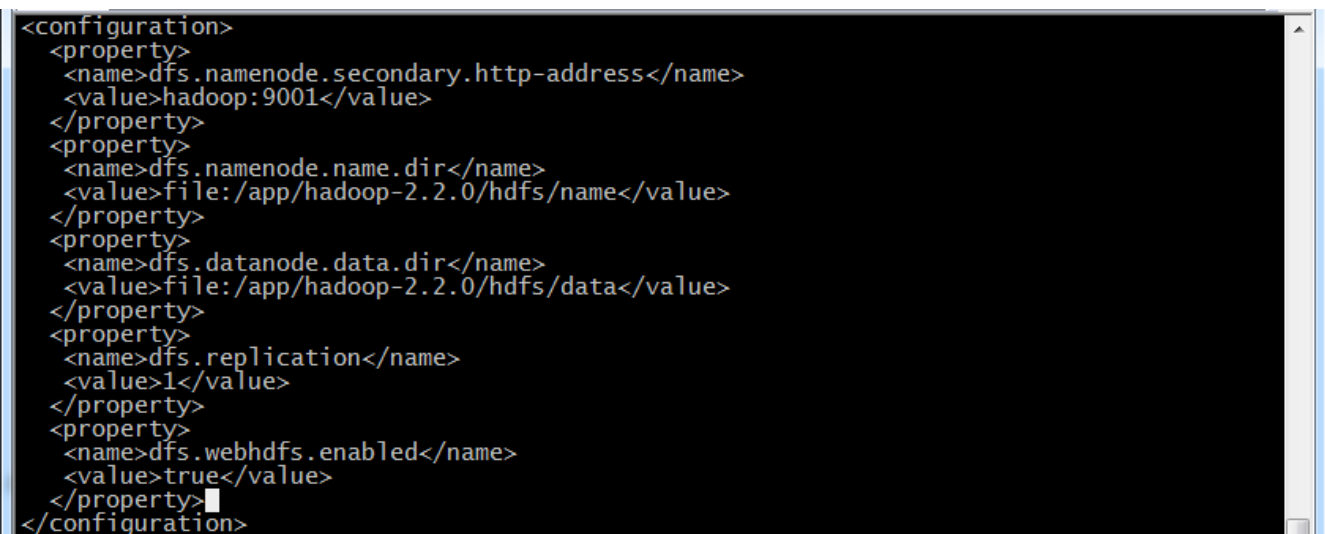
1. 使用如下命令打开 hdfs-site.xml 配置文件

```
cd /app/hadoop-2.2.0/etc/hadoop
```

sudo vi hdfs-site.xml

2. 在配置文件中，按照如下内容进行配置

```
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>hadoop:9001</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/app/hadoop-2.2.0/hdfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/app/hadoop-2.2.0/hdfs/data</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.webhdfs.enabled</name>
    <value>true</value>
  </property>
</configuration>
```



```
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>hadoop:9001</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/app/hadoop-2.2.0/hdfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/app/hadoop-2.2.0/hdfs/data</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.webhdfs.enabled</name>
    <value>true</value>
  </property>
</configuration>
```

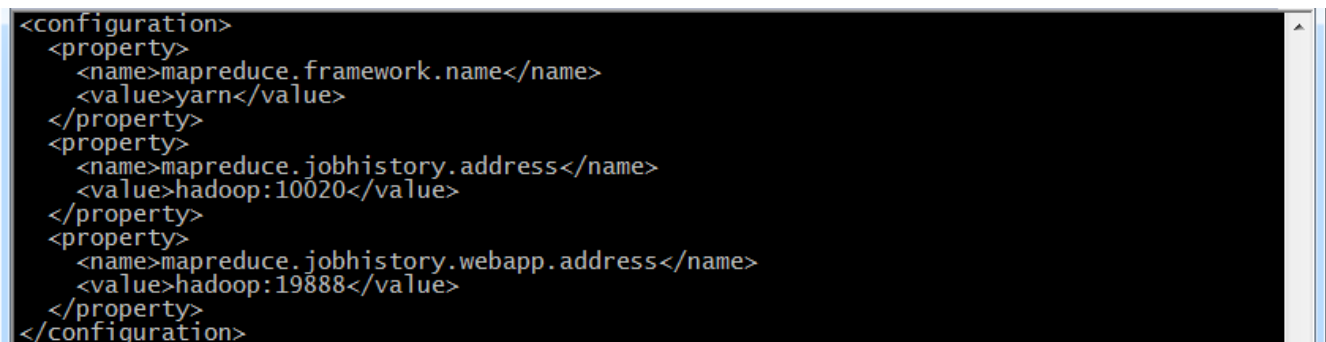
2.1.7 配置 mapred-site.xml

1. 默认情况下不存在 mapred-site.xml 文件，可以从模板拷贝一份，并使用如下命令打开 mapred-site.xml 配置文件

```
cd /app/hadoop-2.2.0/etc/hadoop
cp mapred-site.xml.template mapred-site.xml
sudo vi mapred-site.xml
```

2. 在配置文件中，按照如下内容进行配置

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>hadoop:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>hadoop:19888</value>
  </property>
</configuration>
```



```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>hadoop:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>hadoop:19888</value>
  </property>
</configuration>
```

2.1.8 配置 yarn-site.xml

1. 使用如下命令打开 yarn-site.xml 配置文件

```
cd /app/hadoop-2.2.0/etc/hadoop
sudo vi yarn-site.xml
```

2. 在配置文件中，按照如下内容进行配置

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>hadoop:8032</value>
  </property>
  <property>
    <name>yarn.resourcemanager.scheduler.address</name>
    <value>hadoop:8030</value>
  </property>
  <property>
    <name>yarn.resourcemanager.resource-tracker.address</name>
    <value>hadoop:8031</value>
  </property>
  <property>
    <name>yarn.resourcemanager.admin.address</name>
    <value>hadoop:8033</value>
  </property>
  <property>
    <name>yarn.resourcemanager.webapp.address</name>
    <value>hadoop:8088</value>
  </property>
</configuration>
```

```

<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>hadoop:8032</value>
  </property>
  <property>
    <name>yarn.resourcemanager.scheduler.address</name>
    <value>hadoop:8030</value>
  </property>
  <property>
    <name>yarn.resourcemanager.resource-tracker.address</name>
    <value>hadoop:8031</value>
  </property>
  <property>
    <name>yarn.resourcemanager.admin.address</name>
    <value>hadoop:8033</value>
  </property>
  <property>
    <name>yarn.resourcemanager.webapp.address</name>
    <value>hadoop:8088</value>
  </property>
</configuration>

```

2.1.9 配置 slaves 文件

在 slaves 配置文件中设置从节点，这里设置为 hadoop，与 Hadoop1.X 区别的是 Hadoop2.X 不需要设置 Master

```
cd /app/hadoop-2.2.0/etc/hadoop
```

```
vi slaves
```

```

[shiyantou@b393a04554e1 hadoop]$ vi slaves
hadoop
~

```

2.1.10 格式化 namenode

```
cd /app/hadoop-2.2.0/bin
```

```
./hdfs namenode -format
```

```

[shiyantou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0/bin
[shiyantou@b393a04554e1 bin]$ ls
container-executor  hadoop.cmd  hdfs.cmd  mapred.cmd  test-container-executor  yarn.cmd
hadoop             hdfs       mapred     rcc         yarn
[shiyantou@b393a04554e1 bin]$ ./hdfs namenode -format
2015-06-02 07:10:17,177 INFO [main] namenode.NameNode (StringUtils.java:startupShutdownMessage(601)) -
STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = b393a04554e1/192.168.42.8
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.2.0
STARTUP_MSG: classpath = /etc/hadoop:/app/hadoop-2.2.0/share/hadoop/common/lib/asm-3.2.jar:/app/hadoo
p-2.2.0/share/hadoop/common/lib/junit-4.8.2.jar:/app/hadoop-2.2.0/share/hadoop/common/lib/jackson-mappe
r-asl-1.8.8.jar:/app/hadoop-2.2.0/share/hadoop/common/lib/hadoop-auth-2.2.0.jar:/app/hadoop-2.2.0/share
/hadoop/common/lib/zookeeper-3.4.5.jar:/app/hadoop-2.2.0/share/hadoop/common/lib/paranamer-2.3.jar:/app
/hadoop-2.2.0/share/hadoop/common/lib/jets3t-0.6.1.jar:/app/hadoop-2.2.0/share/hadoop/common/lib/jsr305
-1.3.9.jar:/app/hadoop-2.2.0/share/hadoop/common/lib/loq4j-1.2.17.jar:/app/hadoop-2.2.0/share/hadoop/co

```



```
2015-06-02 07:10:19,102 INFO [main] util.GSet (LightWeightGSet.java:computeCapacity(330)) - 0.0299999999329447746% max memory = 889 MB
2015-06-02 07:10:19,102 INFO [main] util.GSet (LightWeightGSet.java:computeCapacity(332)) - capacity = 2^15 = 32768 entries
2015-06-02 07:10:19,313 INFO [main] common.Storage (NNStorage.java:format(527)) - Storage directory /tmp/hadoop-shiyanlou/dfs/name has been successfully formatted.
2015-06-02 07:10:19,351 INFO [FSImageSaver for /tmp/hadoop-shiyanlou/dfs/name of type IMAGE_AND_EDITS] namenode.FSImage (FSImageFormat.java:save(976)) - Saving image file /tmp/hadoop-shiyanlou/dfs/name/current/fsimage.ckpt_00000000000000000000 using no compression
2015-06-02 07:10:19,403 INFO [FSImageSaver for /tmp/hadoop-shiyanlou/dfs/name of type IMAGE_AND_EDITS] namenode.FSImage (FSImageFormat.java:save(1004)) - Image file /tmp/hadoop-shiyanlou/dfs/name/current/fsimage.ckpt_00000000000000000000 of size 201 bytes saved in 0 seconds.
2015-06-02 07:10:19,445 INFO [main] namenode.NNStorageRetentionManager (NNStorageRetentionManager.java:getImageTxIdToRetain(177)) - Going to retain 1 images with txid >= 0
2015-06-02 07:10:19,452 INFO [main] util.ExitUtil (ExitUtil.java:terminate(124)) - Exiting with status 0
2015-06-02 07:10:19,462 INFO [Thread-1] namenode.NameNode (StringUtils.java:run(627)) - SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at b393a04554e1/192.168.42.8
SHUTDOWN_MSG: Shutting down NameNode at b393a04554e1/192.168.42.8
*****/
[shiyanlou@b393a04554e1 bin]$
```

2.2 启动 Hadoop

2.2.1 启动 hdfs

```
cd /app/hadoop-2.2.0/sbin
./start-dfs.sh
```

```
[shiyanlou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0/sbin
[shiyanlou@b393a04554e1 sbin]$ ls
distribute-exclude.sh  refresh-namenodes.sh  start-secure-dns.sh  stop-dfs.sh
hadoop-daemon.sh      slaves.sh              start-yarn.cmd       stop-secure-dns.sh
hadoop-daemons.sh    start-all.cmd         start-yarn.sh        stop-yarn.cmd
hdfs-config.cmd       start-all.sh          stop-all.cmd        stop-yarn.sh
hdfs-config.sh        start-balancer.sh     stop-all.sh         yarn-daemon.sh
httpfs.sh             start-dfs.cmd         stop-balancer.sh    yarn-daemons.sh
mr-jobhistory-daemon.sh start-dfs.sh          stop-dfs.cmd
[shiyanlou@b393a04554e1 sbin]$ ./start-dfs.sh
Starting namenodes on [hadoop]
hadoop: starting namenode, logging to /app/hadoop-2.2.0/logs/hadoop-shiyanlou-namenode-b393a04554e1.out
hadoop: starting datanode, logging to /app/hadoop-2.2.0/logs/hadoop-shiyanlou-datanode-b393a04554e1.out
Starting secondary namenodes [hadoop]
hadoop: starting secondarynamenode, logging to /app/hadoop-2.2.0/logs/hadoop-shiyanlou-secondarynamenode-b393a04554e1.out
[shiyanlou@b393a04554e1 sbin]$
```

2.2.2 验证当前进行

使用 jps 命令查看运行进程，此时在 hadoop 上面运行的进程有：namenode、secondarynamenode 和 datanode 三个进程

```
[shiyanlou@b393a04554e1 sbin]$ jps
30994 Jps
30885 SecondaryNameNode
30734 DataNode
30618 NameNode
[shiyanlou@b393a04554e1 sbin]$
```

2.2.3 启动 yarn

```
cd /app/hadoop-2.2.0/sbin
./start-yarn.sh
```

```
[shiyanolou@b393a04554e1 sbin]$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /app/hadoop-2.2.0/logs/yarn-shiyanolou-resourcemanager-b393a04554e1.out
hadoop: starting nodemanager, logging to /app/hadoop-2.2.0/logs/yarn-shiyanolou-nodemanager-b393a04554e1.out
[shiyanolou@b393a04554e1 sbin]$
```

2.2.4 验证当前进行

使用 `jps` 命令查看运行进程，此时在 `hadoop` 上运行的进程除了：`namenode`、`secondarynamenode` 和 `datanode`，增加了 `resourcemanager` 和 `nodemanager` 两个进程：

```
[shiyanolou@b393a04554e1 sbin]$ jps
31420 Jps
31046 ResourceManager
30885 SecondaryNameNode
30734 DataNode
31140 NodeManager
30618 NameNode
[shiyanolou@b393a04554e1 sbin]$
```

2.3 测试 Hadoop

2.3.1 创建测试目录

```
cd /app/hadoop-2.2.0/bin
./hadoop fs -mkdir -p /class3/input
```

```
[shiyanolou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0/bin
[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -mkdir -p /class3/input
[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -ls /class3
Found 1 items
drwxr-xr-x - shiyanolou supergroup 0 2015-06-02 08:28 /class3/input
[shiyanolou@b393a04554e1 bin]$
```

2.3.2 准备测试数据

```
./hadoop fs -copyFromLocal ../etc/hadoop/* /class3/input
```

```
[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -copyFromLocal ../etc/hadoop/* /class3/input
[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -ls /class3/input
Found 26 items
-rw-r--r-- 1 shiyanolou supergroup 3560 2015-06-02 14:45 /class3/input/capacity-scheduler.xml
-rw-r--r-- 1 shiyanolou supergroup 1335 2015-06-02 14:45 /class3/input/configuration.xml
-rw-r--r-- 1 shiyanolou supergroup 318 2015-06-02 14:45 /class3/input/container-executor.cfg
-rw-r--r-- 1 shiyanolou supergroup 1423 2015-06-02 14:45 /class3/input/core-site.xml
-rw-r--r-- 1 shiyanolou supergroup 3589 2015-06-02 14:45 /class3/input/hadoop-env.cmd
-rw-r--r-- 1 shiyanolou supergroup 3577 2015-06-02 14:45 /class3/input/hadoop-env.sh
-rw-r--r-- 1 shiyanolou supergroup 2490 2015-06-02 14:45 /class3/input/hadoop-metrics.properties
```

2.3.3 运行 wordcount 例子

```
cd /app/hadoop-2.2.0/bin
./hadoop jar ../share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar
wordcount /class3/input /class3/output
```

```

[shiyanolou@b393a04554e1 ~]$ cd /app/hadoop-2.2.0/bin
[shiyanolou@b393a04554e1 bin]$ ./hadoop jar ../share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar wordcount /class3/input /class3/output
15/06/03 02:00:26 INFO client.RMProxy: Connecting to ResourceManager at hadoop/192.168.42.8:8032
15/06/03 02:00:27 INFO input.FileInputFormat: Total input paths to process : 26
15/06/03 02:00:27 INFO mapreduce.JobSubmitter: number of splits:26
15/06/03 02:00:27 INFO Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.output.value.class is deprecated. Instead, use mapreduce.job.output.value.class
15/06/03 02:00:27 INFO Configuration.deprecation: mapreduce.combine.class is deprecated. Instead, use mapreduce.job.combine.class
15/06/03 02:00:27 INFO Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
15/06/03 02:00:27 INFO Configuration.deprecation: mapreduce.reduce.class is deprecated. Instead, use mapreduce.job.reduce.class
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.output.key.class is deprecated. Instead, use mapreduce.job.output.key.class
15/06/03 02:00:27 INFO Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
15/06/03 02:00:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1433295148948_0005
15/06/03 02:00:27 INFO impl.YarnClientImpl: Submitted application application_1433295148948_0005 to ResourceManager at hadoop/192.168.42.8:8032
15/06/03 02:00:28 INFO mapreduce.Job: The url to track the job: http://hadoop:8088/proxy/application_1433295148948_0005/
15/06/03 02:00:28 INFO mapreduce.Job: Running job: job_1433295148948_0005
15/06/03 02:00:36 INFO mapreduce.Job: Job job_1433295148948_0005 running in uber mode : false
15/06/03 02:00:36 INFO mapreduce.Job: map 0% reduce 0%
15/06/03 02:00:55 INFO mapreduce.Job: map 12% reduce 0%
15/06/03 02:00:56 INFO mapreduce.Job: map 23% reduce 0%
15/06/03 02:01:12 INFO mapreduce.Job: map 27% reduce 0%
15/06/03 02:01:13 INFO mapreduce.Job: map 35% reduce 0%
15/06/03 02:01:14 INFO mapreduce.Job: map 38% reduce 0%
15/06/03 02:01:15 INFO mapreduce.Job: map 42% reduce 0%
15/06/03 02:01:20 INFO mapreduce.Job: map 42% reduce 14%
15/06/03 02:01:27 INFO mapreduce.Job: map 46% reduce 14%
15/06/03 02:01:28 INFO mapreduce.Job: map 50% reduce 14%
15/06/03 02:01:29 INFO mapreduce.Job: map 54% reduce 17%
15/06/03 02:01:30 INFO mapreduce.Job: map 62% reduce 17%
15/06/03 02:01:32 INFO mapreduce.Job: map 62% reduce 21%

```

```

15/06/03 02:01:57 INFO mapreduce.Job: map 92% reduce 28%
15/06/03 02:01:59 INFO mapreduce.Job: map 100% reduce 32%
15/06/03 02:02:02 INFO mapreduce.Job: map 100% reduce 100%
15/06/03 02:02:03 INFO mapreduce.Job: Job job_1433295148948_0005 completed successfully
15/06/03 02:02:03 INFO mapreduce.Job: Counters: 43
File System Counters
  FILE: Number of bytes read=60766
  FILE: Number of bytes written=2262131
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=64745
  HDFS: Number of bytes written=28180
  HDFS: Number of read operations=81
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=26
  Total time spent by all maps in occupied slots (ms)=395615
  Total time spent by all reduces in occupied slots (ms)=63888
Map-Reduce Framework
  Map input records=1635
  Map output records=6474
  Map output bytes=85160
  Map output materialized bytes=60916

```

2.3.4 查看结果

使用如下命令查看运行结果：

```
./hadoop fs -ls /class3/output/
```

```
./hadoop fs -cat /class3/output/part-r-00000 | less
```

```

[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -ls /class3/output/
Found 2 items
-rw-r--r-- 1 shiyanolou supergroup 0 2015-06-03 02:02 /class3/output/_SUCCESS
-rw-r--r-- 1 shiyanolou supergroup 28180 2015-06-03 02:02 /class3/output/part-r-00000
[shiyanolou@b393a04554e1 bin]$ ./hadoop fs -cat /class3/output/part-r-00000 | less
!= 3
"" 6
"" 4
"$HADOOP_CLASSPATH" 1
"$JAVA_HOME" 2
"$YARN_HEAPSIZE" 1
"$YARN_LOGFILE" 1
"$YARN_LOG_DIR" 1
"$YARN_POLICYFILE" 1
"*" 17

```

3 问题解决

3.1 CentOS 64bit 安装 Hadoop2.2.0 中出现文件编译位数异常

在安装 hadoop2.2.0 过程中出现如下异常：**Unable to load native-hadoop library for your platform... using builtin-java classes where applicable**

```
[hadoop@hadoop1 /usr/local/hadoop-2.2.0/sbin]$ ls
distribute-exclude.sh  refresh-namenodes.sh  start-secure-dns.sh  stop-dfs.sh
hadoop-daemon.sh      slaves.sh              start-yarn.cmd       stop-secure-dns.sh
hadoop-daemons.sh    start-all.cmd        start-yarn.sh        stop-yarn.cmd
hdfs-config.cmd       start-all.sh         stop-all.cmd        stop-yarn.sh
hdfs-config.sh        start-balancer.sh    stop-all.sh         yarn-daemon.sh
httpfs.sh             start-dfs.cmd        stop-balancer.sh    yarn-daemons.sh
mr-jobhistory-daemon.sh start-dfs.sh         stop-dfs.cmd
[hadoop@hadoop1 /usr/local/hadoop-2.2.0/sbin]$ ./start-dfs.sh
14/09/24 10:17:15 WARN util.NativeCodeLoader: unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [Java HotSpot(TM) 64-Bit Server VM warning: You have loaded library /usr/local/hadoop-2.2.0/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
hadoop1]
```

通过分析是由于 lib/native 目录中有些文件是在 32 位编译，无法适应 CentOS 64 位环境造成

```
[hadoop@hadoop1 /usr/local/hadoop-2.2.0]$ ls
bin  etc  lib  LICENSE.txt  name  README.txt  share
data include libexec logs  NOTICE.txt sbin  tmp
[hadoop@hadoop1 /usr/local/hadoop-2.2.0]$ cd lib/native/
[hadoop@hadoop1 /usr/local/hadoop-2.2.0/lib/native]$ ls
libhadoop.a  libhadoop.so  libhadooputils.a  libhdfs.so
libhadooppipes.a  libhadoop.so.1.0.0  libhdfs.a  libhdfs.so.0.0.0
[hadoop@hadoop1 /usr/local/hadoop-2.2.0/lib/native]$ file ./libhadoop.so.1.0.0
./libhadoop.so.1.0.0: ELF 32-bit LSB shared object, Intel 80386, version 1 (SYSV), dynamically
linked, not stripped
[hadoop@hadoop1 /usr/local/hadoop-2.2.0/lib/native]$
```

有两种办法解决：

- 重新编译 hadoop，然后重新部署
- 暂时办法是修改配置，忽略有问题的文件

```
export HADOOP_COMMON_LIB_NATIVE_DIR=/home/grid/hadoop-2.2.0/lib/native
export HADOOP_OPTS="-Djava.library.path=/home/grid/hadoop-2.2.0/lib"
```