Apache　Hadoop 三种架构介绍（standAlone 环境介绍以及安装）

```
hadoop 文档
http://hadoop.apache.org/docs/
```

5.1、StandAlone 环境搭建

运行服务	服务器 IP
NameNode	192.168.52.100
SecondaryNameNode	192.168.52.100
DataNode	192.168.52.100
ResourceManager	192.168.52.100
NodeManager	192.168.52.100

第一步：下载 apache hadoop 并上传到服务器

下载链接：

http://archive.apache.org/dist/hadoop/common/hadoop-2.7.5/hadoop-2.7.5.tar.gz

解压命令

```
cd /export/softwares
tar -zxvf hadoop-2.7.5.tar.gz -C ../servers/
```

```
s[root@node04 softwares]# ll
total 415216
-rw-r--r--. 1 root root 216929574 Sep 27  2018 hadoop-2.7.5.tar.gz
-rw-r--r--. 1 root root 185516505 Sep 27  2018 jdk-8u141-linux-x64.tar.gz
-rw-r--r--. 1 root root  22724574 Sep 27  2018 zookeeper-3.4.9.tar.gz
You have new mail in /var/spool/mail/root
[root@node04 softwares]# tar -zxvf hadoop-2.7.5.tar.gz -C ../servers/
```

hadoop 安装包结构

hadoop-2.7.5/bin:一些 shell 脚本，供我们使用

hadoop-2.7.5/sbin:一些 shell 脚本，供我们使用

hadoop-2.7.5/etc/hadoop:所有的配置文件的路径

hadoop-2.7.5/lib/native:本地的 C 程序库

hadoop 六个核心配置文件的作用:

core-site.xml：核心配置文件，主要定义了我们文件访问的格式 hdfs://

hadoop-env.sh：主要配置我们的 java 路径

hdfs-site.xml：主要定义配置我们的 hdfs 的相关配置

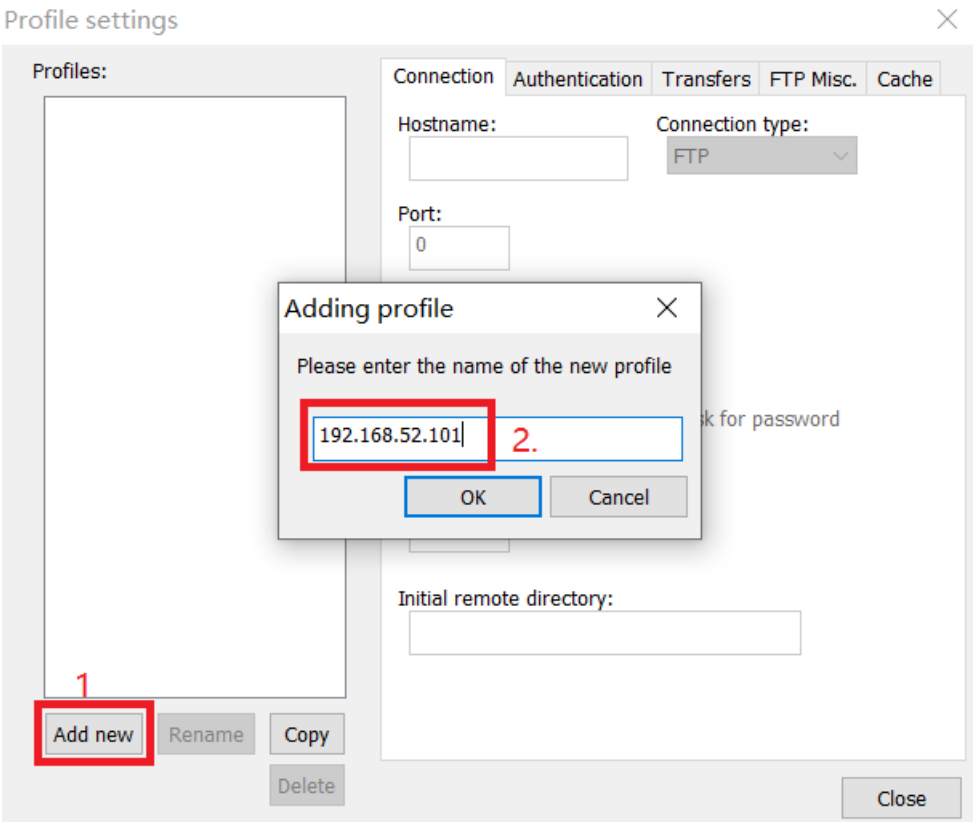
mapred-site.xml 主要定义我们的 mapreduce 相关的一些配置

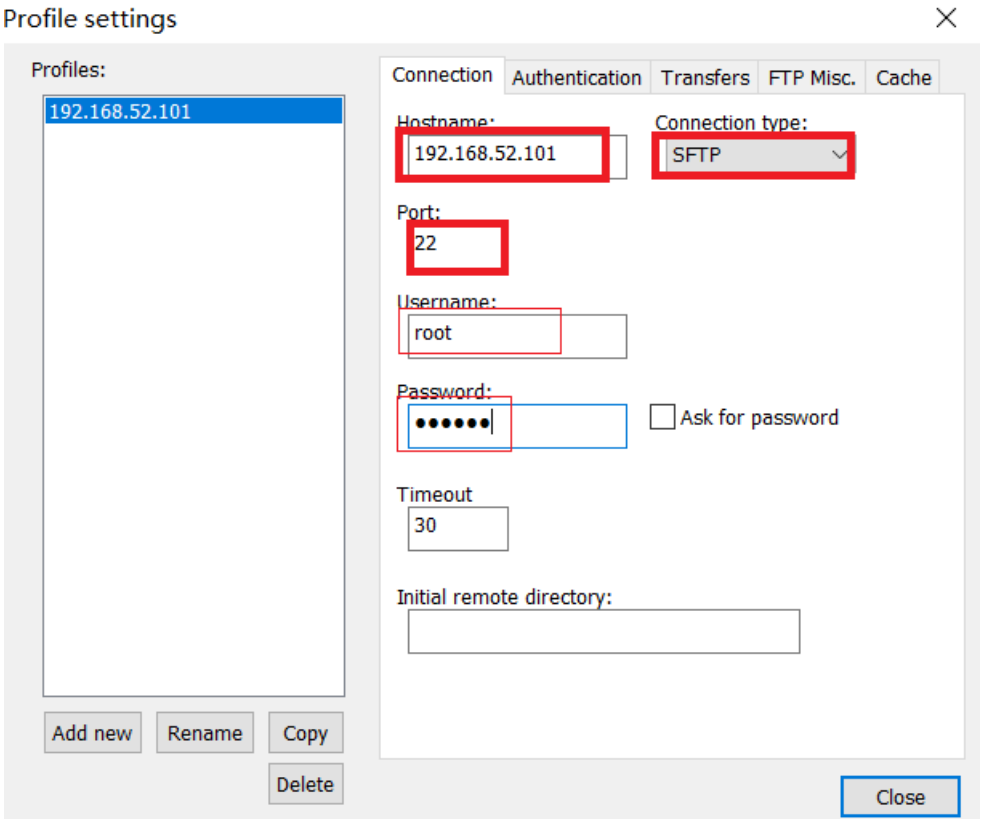
slaves：控制我们的从节点在哪里 datanode nodemanager 在哪些机器上

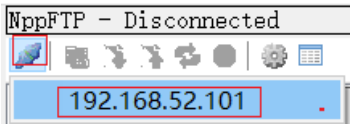
yarn-site.xml：配置我们的 resourcemanager 资源调度

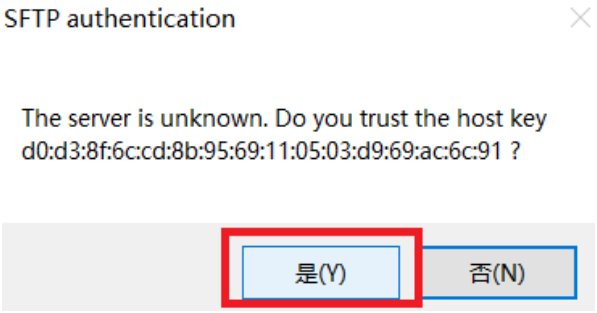
第二步：修改配置文件

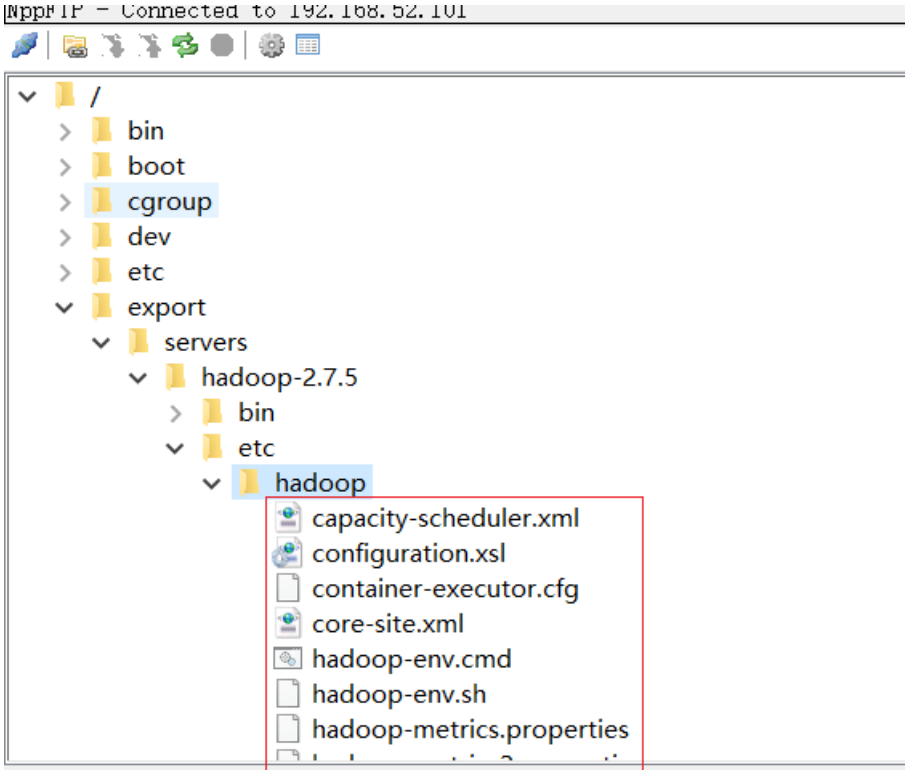
打开 notepad++

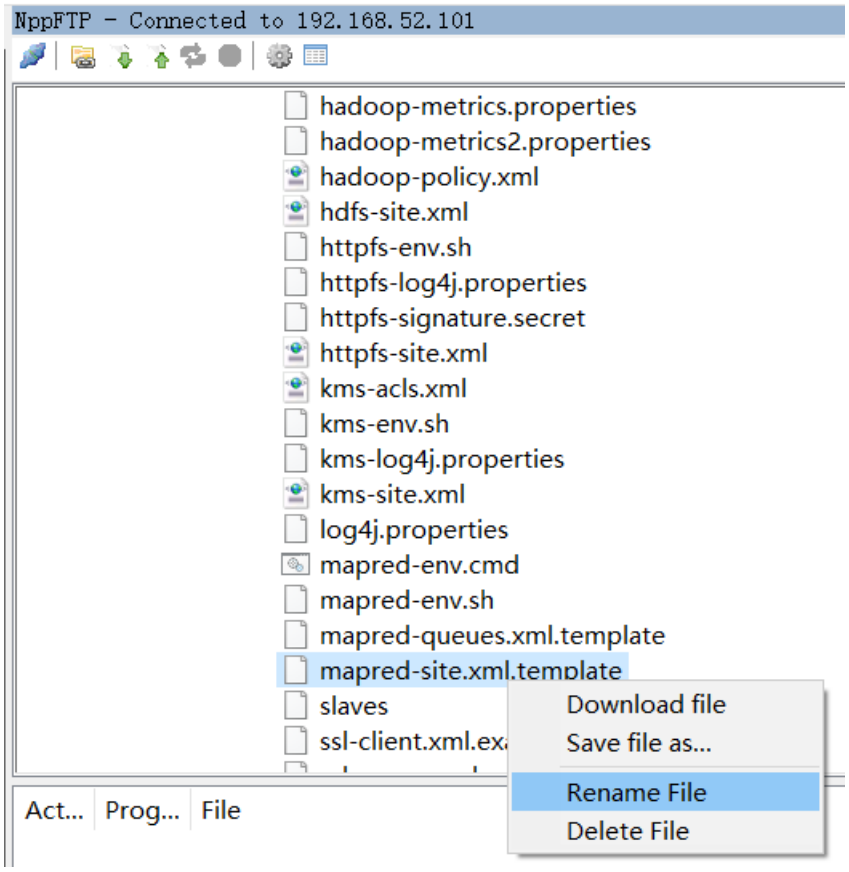

Profile settings
Profiles:
Connection
Authentication
Transfers
FTP Misc.
Cache
Hostname:
Connection type:
FTP
Port:
0
Adding profile
Please enter the name of the new profile
192.168.52.101
2.
OK
Cancel
for password
Initial remote directory:
1
Add new
Rename
Copy
Delete
Close

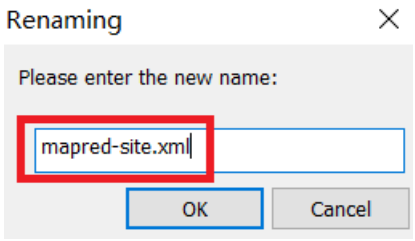

Profile settings
Profiles:
192.168.52.101
Connection
Authentication
Transfers
FTP Misc.
Cache
Hostname:
192.168.52.101
Connection type:
SFTP
Port:
22
Username:
root
Password:
Ask for password
Timeout
30
Initial remote directory:
Add new
Rename
Copy
Delete
Close

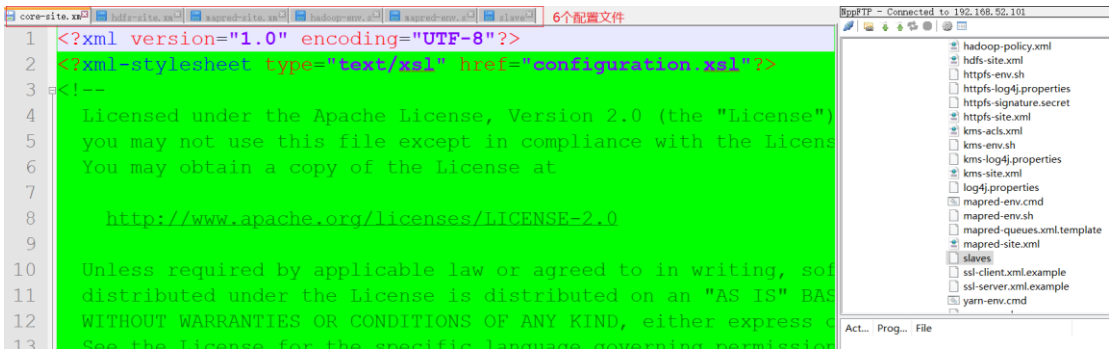

NppFTP - Disconnected
192.168.52.101


SFTP authentication
The server is unknown. Do you trust the host key
d0:d3:8f:6c:cd:8b:95:69:11:05:03:d9:69:ac:6c:91 ?
是(Y)
否(N)


NppFTP - Connected to 192.168.52.101
/
bin
boot
cgroup
dev
etc
export
servers
hadoop-2.7.5
bin
etc
hadoop
capacity-scheduler.xml
configuration.xsl
container-executor.cfg
core-site.xml
hadoop-env.cmd
hadoop-env.sh
hadoop-metrics.properties







修改 core-site.xml

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim  core-site.xml
```

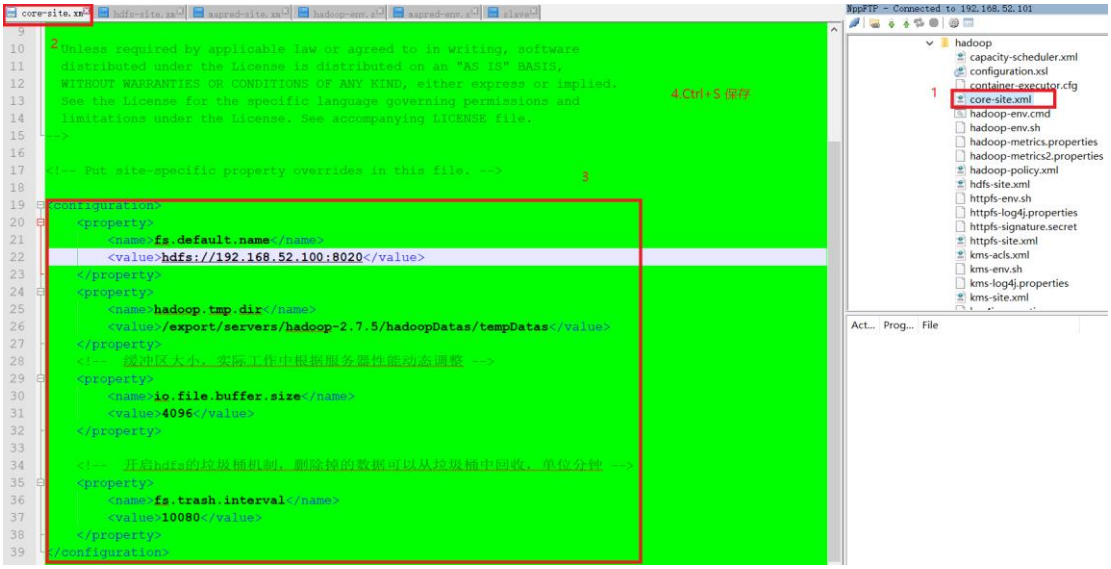
http://archive.cloudera.com/cdh5/cdh/5/hadoop-2.6.0-cdh5.14.0/

fs.defaultFS	file:///

MapFile consist of two files - data file (tuples) and index file (keys). For every io.map.index.interval records written in the data file, an entry (record-key, data-file-position) is written in the index file. This is to allow for doing binary search later within the index file to look up records by their keys and get their closest positions in the data file.

定义文件系统的实现 file:/// 本地文件系统 hdfs://分布式文件系统

```
<configuration>
    <property>
        <name>fs.default.name</name>
        <value>hdfs://192.168.52.100:8020</value>
    </property>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>/export/servers/hadoop-
2.7.5/hadoopDatas/tempDatas</value>
    </property>
    <!--  缓冲区大小，实际工作中根据服务器性能动态调整 -->
    <property>
        <name>io.file.buffer.size</name>
        <value>4096</value>
    </property>

    <!--  开启 hdfs 的垃圾桶机制，删除掉的数据可以从垃圾桶中回收，单
位分钟 -->
    <property>
        <name>fs.trash.interval</name>
        <value>10080</value>
    </property>
</configuration>
```



修改 hdfs-site.xml

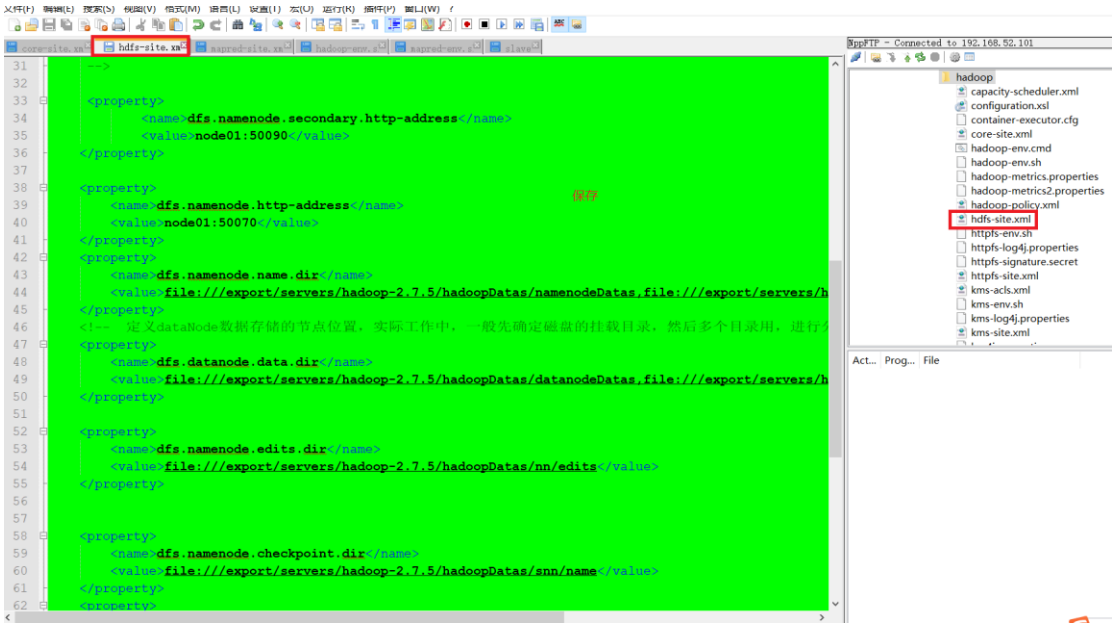
第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim hdfs-site.xml
```

```
<configuration>
      <property>
            <name>dfs.namenode.secondary.http-address</name>
            <value>node01:50090</value>
    </property>

    <property>
        <name>dfs.namenode.http-address</name>
        <value>node01:50070</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:///export/servers/hadoop-
2.7.5/hadoopDatas/namenodeDatas,file:///export/servers/hadoop-
2.7.5/hadoopDatas/namenodeDatas2</value>
    </property>
    <!--   定义 dataNode 数据存储的节点位置，实际工作中，一般先确定
```

```
磁盘的挂载目录，然后多个目录用，进行分割  -->
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:///export/servers/hadoop-
2.7.5/hadoopDatas/datanodeDatas,file:///export/servers/hadoop-
2.7.5/hadoopDatas/datanodeDatas2</value>
    </property>

    <property>
        <name>dfs.namenode.edits.dir</name>
        <value>file:///export/servers/hadoop-
2.7.5/hadoopDatas/nn/edits</value>
    </property>



    <property>
        <name>dfs.namenode.checkpoint.dir</name>
        <value>file:///export/servers/hadoop-
2.7.5/hadoopDatas/snn/name</value>
    </property>
    <property>
        <name>dfs.namenode.checkpoint.edits.dir</name>
        <value>file:///export/servers/hadoop-
2.7.5/hadoopDatas/dfs/snn/edits</value>
    </property>

    <property>
        <name>dfs.replication</name>
        <value>3</value>
    </property>

    <property>
        <name>dfs.permissions</name>
        <value>false</value>
    </property>

<property>
        <name>dfs.blocksize</name>
```

```
        <value>134217728</value>
    </property>
</configuration>
```

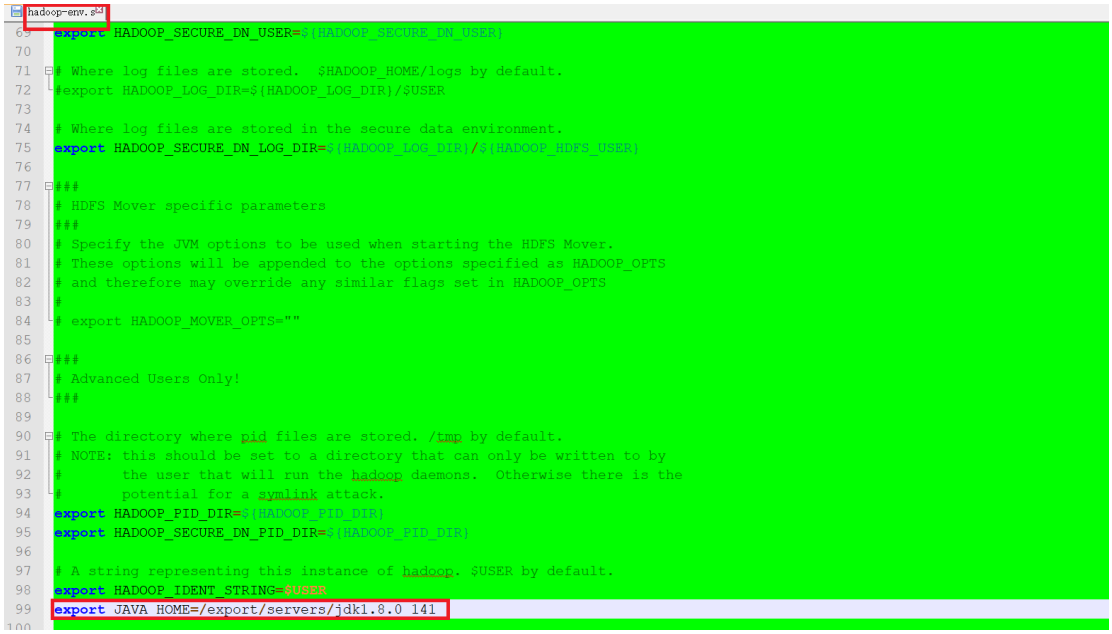


修改 hadoop-env.sh

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim  hadoop-env.sh
```

```
vim   hadoop-env.sh
export JAVA_HOME=/export/servers/jdk1.8.0_141
```



修改 mapred-site.xml

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim  mapred-site.xml
```

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

    <property>
        <name>mapreduce.job.ubertask.enable</name>
        <value>true</value>
    </property>

    <property>
        <name>mapreduce.jobhistory.address</name>
        <value>node01:10020</value>
    </property>

    <property>
```

```
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>node01:19888</value>
    </property>
</configuration>
```

修改 yarn-site.xml

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim  yarn-site.xml
```

```
<configuration>
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>node01</value>
    </property>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>

    <property>
        <name>yarn.log-aggregation-enable</name>
        <value>true</value>
    </property>
```

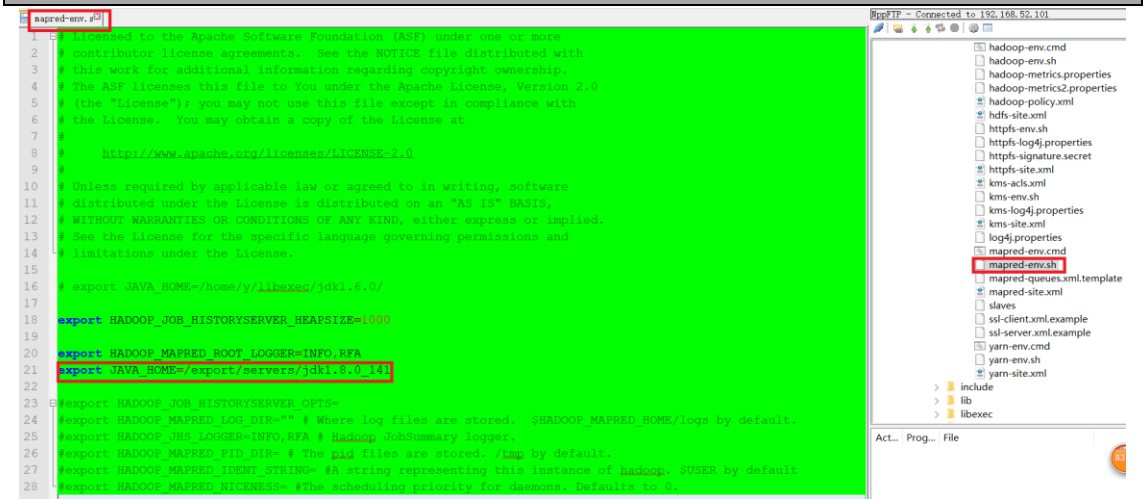
```
    <property>
        <name>yarn.log-aggregation.retain-seconds</name>
        <value>604800</value>
    </property>
</configuration>
```

修改 mapred-env.sh

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop

vim  mapred-env.sh
```

```
export JAVA_HOME=/export/servers/jdk1.8.0_141
```



修改 slaves

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/etc/hadoop
vim   slaves
```

```
localhost
```

第三步：启动集群

要启动 Hadoop 集群，需要启动 HDFS 和 YARN 两个模块。

注意： 首次启动 HDFS 时，必须对其进行格式化操作。 本质上是一些清理和准备工作，因为此时的 HDFS 在物理上还是不存在的。

```
cd /export/servers/hadoop-2.7.5/bin
hdfs namenode -format
```

```
[root@node04 bin]# hdfs namenode -format
19/07/09 11:27:18 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = node04.hadoop.com/192.168.52.101
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.7.5
STARTUP_MSG:   classpath = /export/servers/hadoop-2.7.5/etc/hadoop:/export/servers/hadoop-2.7.5/share/hadoop
lib/commons-logging-1.1.3.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/commons-beanutils-1.7.0.
rt/servers/hadoop-2.7.5/share/hadoop/common/lib/httpcore-4.2.5.jar:/export/servers/hadoop-2.7.5/share/hadoop
lib/apacheds-kerberos-codec-2.0.0-M15.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/asm-3.2.jar:/
ervers/hadoop-2.7.5/share/hadoop/common/lib/commons-codec-1.4.jar:/export/servers/hadoop-2.7.5/share/hadoop/
ib/jackson-core-asl-1.9.13.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/jsp-api-2.1.jar:/export/
hadoop-2.7.5/share/hadoop/common/lib/junit-4.11.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/je
6.1.26.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/jets3t-0.9.0.jar:/export/servers/hadoop-2.7.
hadoop/common/lib/slf4j-log4j12-1.7.10.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/api-util-1.0
ar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/export/servers/hadoop
hare/hadoop/common/lib/hadoop-auth-2.7.5.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/xmlenc-0.
xport/servers/hadoop-2.7.5/share/hadoop/common/lib/jetty-sslengine-6.1.26.jar:/export/servers/hadoop-2.7.5/s
oop/common/lib/jetty-6.1.26.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/jersey-server-1.9.jar:/
ervers/hadoop-2.7.5/share/hadoop/common/lib/zookeeper-3.4.6.jar:/export/servers/hadoop-2.7.5/share/hadoop/co
/hadoop-annotations-2.7.5.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/jersey-core-1.9.jar:/expo
rs/hadoop-2.7.5/share/hadoop/common/lib/commons-digester-1.8.jar:/export/servers/hadoop-2.7.5/share/hadoop/c
b/jsch-0.1.54.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/apacheds-i18n-2.0.0-M15.jar:/export/s
adoop-2.7.5/share/hadoop/common/lib/commons-compress-1.4.1.jar:/export/servers/hadoop-2.7.5/share/hadoop/com
commons-cli-1.2.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/paranamer-2.3.jar:/export/servers/h
7.5/share/hadoop/common/lib/stax-api-1.0-2.jar:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/log4j-1.
:/export/servers/hadoop-2.7.5/share/hadoop/common/lib/commons-io-2.4.jar:/export/servers/hadoop-2.7.5/share/
```

启动命令:

创建数据存放文件夹

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/tempDatas

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/namenodeDatas

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/namenodeDatas2

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/datanodeDatas

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/datanodeDatas2

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/nn/edits

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/snn/name

mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/dfs/snn/edits
```

```
[root@node04 bin]# cd  /export/servers/hadoop-2.7.5
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/tempDatas
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/namenodeDatas
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/namenodeDatas2
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/datanodeDatas
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/datanodeDatas2
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/nn/edits
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/snn/name
[root@node04 hadoop-2.7.5]# mkdir -p /export/servers/hadoop-2.7.5/hadoopDatas/dfs/snn/edits
```

准备启动

第一台机器执行以下命令

```
cd  /export/servers/hadoop-2.7.5/

sbin/start-dfs.sh
```

```
[root@node04 hadoop-2.7.5]# cd  /export/servers/hadoop-2.7.5/
[root@node04 hadoop-2.7.5]# sbin/start-dfs.sh
```

```
[root@node04 hadoop-2.7.5]# jps
25558 DataNode
25431 NameNode
2391 QuorumPeerMain
25833 Jps
25724 SecondaryNameNode
```

```
sbin/start-yarn.sh
```

```
[root@node04 hadoop-2.7.5]# sbin/start-yarn.sh
```

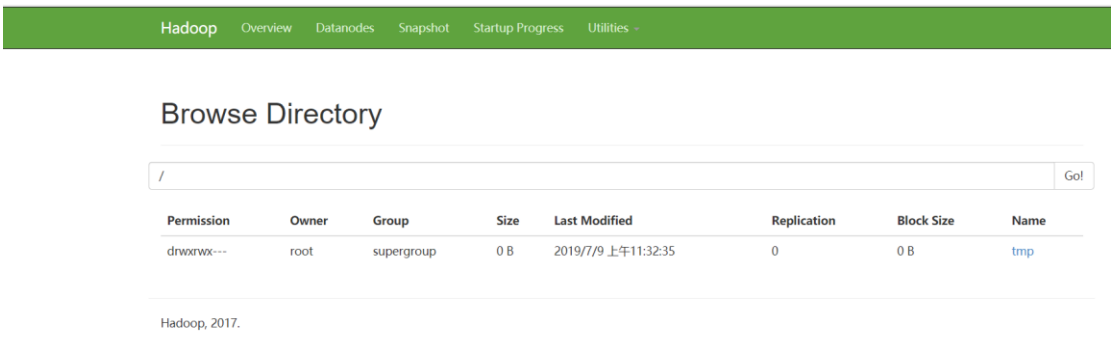
```
[root@node04 hadoop-2.7.5]# jps
25558 DataNode
25431 NameNode
2391 QuorumPeerMain
26200 Jps
25992 NodeManager
25897 ResourceManager
25724 SecondaryNameNode
```

```
sbin/mr-jobhistory-daemon.sh start historyserver
```

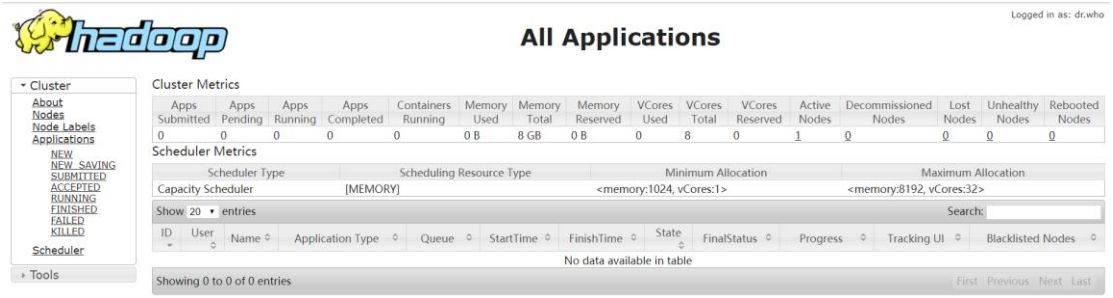
```
[root@node04 hadoop-2.7.5]# sbin/mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /export/servers/hadoop-2.7.5/logs/mapred-root-h
You have new mail in /var/spool/mail/root
[root@node04 hadoop-2.7.5]# jps
25558 DataNode
25431 NameNode
2391 QuorumPeerMain
25992 NodeManager
26329 JobHistoryServer
25897 ResourceManager
25724 SecondaryNameNode
26366 Jps
[root@node04 hadoop-2.7.5]#
```

三个端口查看界面

192.168.52.100:50070/explorer.html#/ 查看 hdfs 绿色的!



192.168.52.100:8088/cluster 查看 yarn 集群



192.168.52.100:19888/jobhistory 查看历史完成的任务

Logged in as: dr.who

JobHistory

- Application
  - About
  - Jobs
- Tools

Retired Jobs

Show 20 entries Search:

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
No data available in table											
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Complete	Reduces Total	Reduces Compl

Showing 0 to 0 of 0 entries First Previous Next Last

Logged in as: dr.who

JobHistory

Application
About
Jobs
Tools

Retired Jobs

Show 20 entries Search:

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
No data available in table											
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Complete	Reduces Total	Reduces Compl

Showing 0 to 0 of 0 entries First Previous Next Last