1. 上传 hadoop 版本到主机一/opt/softwares
2. 解压 hadoop 包缩到 modules 目录，更改权限等（同第一章）
3. Hadoop 瘦身

   进入 hadoop 的 share 文件夹下，rm -rf ./doc/

```
[kfk@bigdata-pro01 etc]$ cd ../share
[kfk@bigdata-pro01 share]$ ls
doc  hadoop
[kfk@bigdata-pro01 share]$ rm -rf ./doc/
[kfk@bigdata-pro01 share]$ ls
hadoop
[kfk@bigdata-pro01 share]$  rm -rf ./doc/
```

   进入 ete/hadoop 文件夹下，rm -rf ./*.cmd

```
[kfk@bigdata-pro01 share]$ cd ../etc/hadoop/
[kfk@bigdata-pro01 hadoop]$ ll
total 124
-rw-r--r-- 1 kfk kfk  3589 Aug  6  2014 capacity-scheduler.xml
-rw-r--r-- 1 kfk kfk  1335 Aug  6  2014 configuration.xsl
-rw-r--r-- 1 kfk kfk   318 Aug  6  2014 container-executor.cfg
-rw-r--r-- 1 kfk kfk   774 Aug  6  2014 core-site.xml
-rw-r--r-- 1 kfk kfk  3589 Aug  6  2014 hadoop-env.cmd
-rw-r--r-- 1 kfk kfk  3443 Aug  6  2014 hadoop-env.sh
-rw-r--r-- 1 kfk kfk  1774 Aug  6  2014 hadoop-metrics2.properties
-rw-r--r-- 1 kfk kfk  2490 Aug  6  2014 hadoop-metrics.properties
-rw-r--r-- 1 kfk kfk  9201 Aug  6  2014 hadoop-policy.xml
-rw-r--r-- 1 kfk kfk   775 Aug  6  2014 hdfs-site.xml
-rw-r--r-- 1 kfk kfk  1449 Aug  6  2014 httpfs-env.sh
-rw-r--r-- 1 kfk kfk  1657 Aug  6  2014 httpfs-log4j.properties
-rw-r--r-- 1 kfk kfk    21 Aug  6  2014 httpfs-signature.secret
-rw-r--r-- 1 kfk kfk   620 Aug  6  2014 httpfs-site.xml
-rw-r--r-- 1 kfk kfk 11118 Aug  6  2014 log4j.properties
-rw-r--r-- 1 kfk kfk   918 Aug  6  2014 mapred-env.cmd
-rw-r--r-- 1 kfk kfk  1383 Aug  6  2014 mapred-env.sh
-rw-r--r-- 1 kfk kfk  4113 Aug  6  2014 mapred-queues.xml.template
-rw-r--r-- 1 kfk kfk   758 Aug  6  2014 mapred-site.xml.template
-rw-r--r-- 1 kfk kfk    10 Aug  6  2014 slaves
-rw-r--r-- 1 kfk kfk  2316 Aug  6  2014 ssl-client.xml.example
-rw-r--r-- 1 kfk kfk  2268 Aug  6  2014 ssl-server.xml.example
-rw-r--r-- 1 kfk kfk  2178 Aug  6  2014 yarn-env.cmd
-rw-r--r-- 1 kfk kfk  4567 Aug  6  2014 yarn-env.sh
-rw-r--r-- 1 kfk kfk   690 Aug  6  2014 yarn-site.xml
[kfk@bigdata-pro01 hadoop]$ rm -rf ./*.cmd
[kfk@bigdata-pro01 hadoop]$ ls
capacity-scheduler.xml  hadoop-env.sh               hdfs-site.xml           httpfs-site.xml             mapred-site.xml.template  yarn-env.sh
configuration.xsl       hadoop-metrics2.properties  httpfs-env.sh           log4j.properties            slaves                    yarn-site.xml
container-executor.cfg  hadoop-metrics.properties   httpfs-log4j.properties  mapred-env.sh               ssl-client.xml.example
core-site.xml           hadoop-policy.xml           httpfs-signature.secret  mapred-queues.xml.template  ssl-server.xml.example
[kfk@bigdata-pro01 hadoop]$
```

4. 连接 notepad 配置 hadoop
5. 格式化：

   进入 hadoop-2.5.0 目录下

   bin/hdfs namenode -format
6. 启动节点：

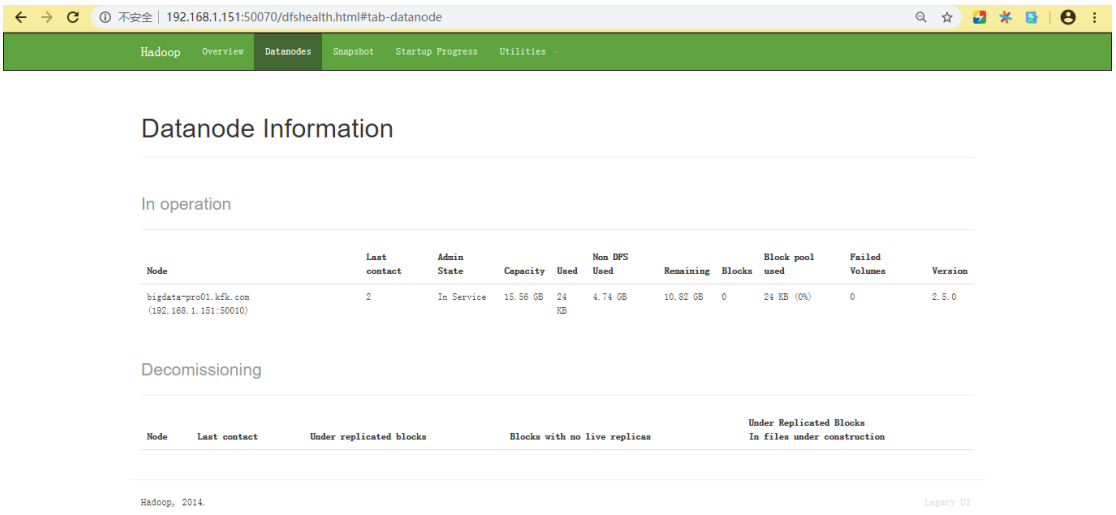
   sbin/hadoop-daemon.sh start namenode

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ ls
bin  etc  include  lib  libexec  sbin  share
[kfk@bigdata-pro01 hadoop-2.5.0]$ sbin/hadoop-daemon.sh start namenode
starting namenode, logging to /opt/modules/hadoop-2.5.0/logs/hadoop-kfk-namenode-bigdata-pro01.kfk.com.out
[kfk@bigdata-pro01 hadoop-2.5.0]$ jps
27711 Jps
27676 NameNode
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

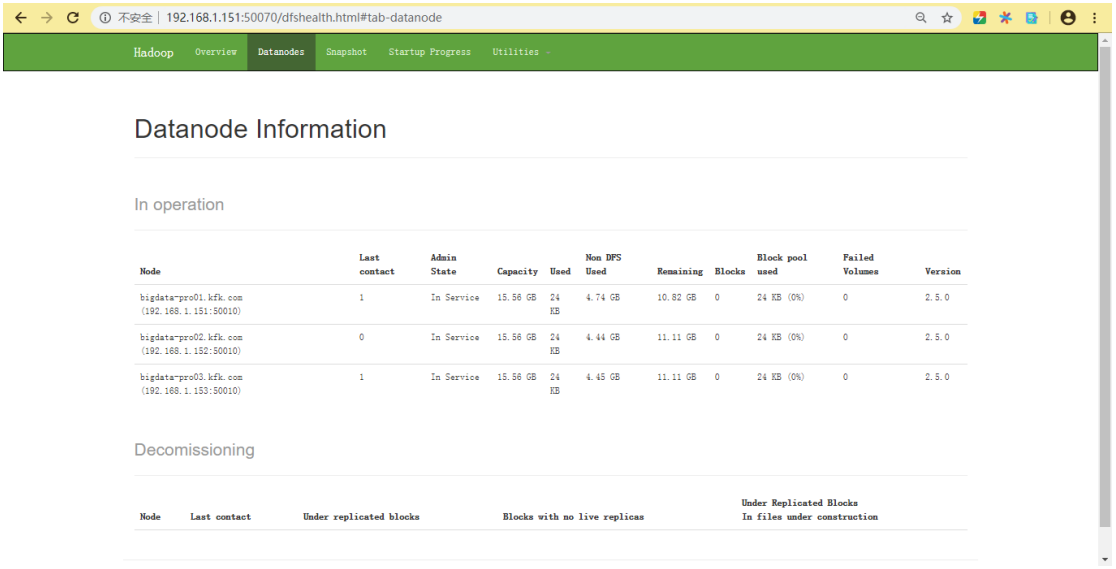
   sbin/hadoop-daemon.sh start datanode

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ sbin/hadoop-daemon.sh start datanode
starting datanode, logging to /opt/modules/hadoop-2.5.0/logs/hadoop-kfk-datanode-bigdata-pro01.kfk.com.out
[kfk@bigdata-pro01 hadoop-2.5.0]$ jps
27828 Jps
27676 NameNode
27764 DataNode
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

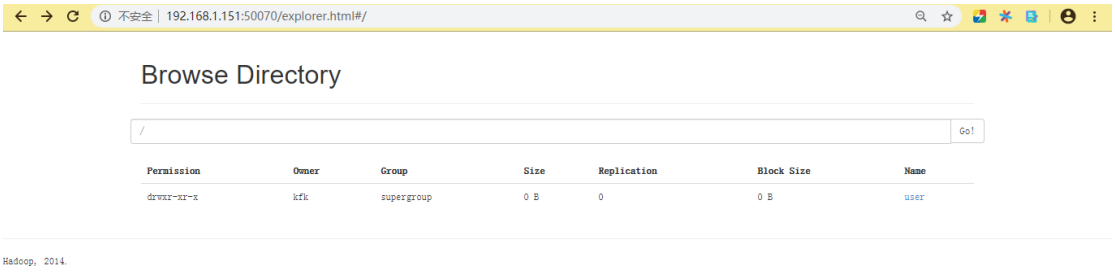
   192.168.1.151:50070



7. 将配置好的 hadoop 发给其他两台主机：
   scp -r hadoop-2.5.0/ kfk@bigdata-pro02.kfk.com:/opt/modules/
   scp -r hadoop-2.5.0/ kfk@bigdata-pro03.kfk.com:/opt/modules/
8. 启动 2，3 主机的 datanode：
   都进入 hadoop-2.5.0 目录下，sbin/Hadoop-daemon.sh start datanode

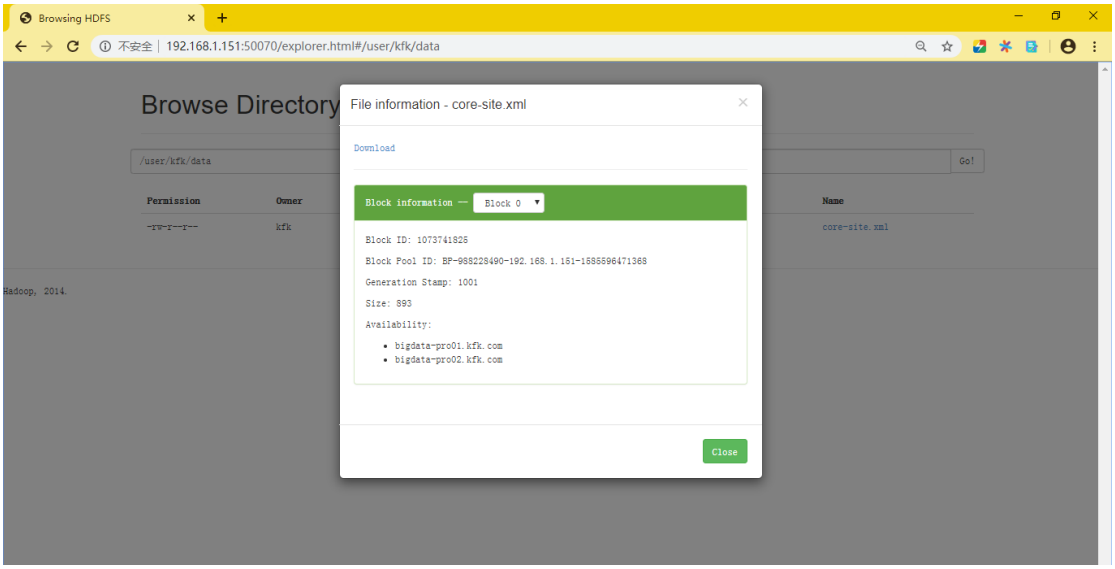


9. 测试上传文件：
   主机一，进入 hadoop-2.5.0 目录下
   创建 user 目录，bin/hdfs dfs -mkdir -p /user/kfk/data/



上传文件，

bin/hdfs dfs -put /opt/modules/hadoop-2.5.0/etc/hadoop/core-site.xml /user/kfk/data



读取文件，bin/hdfs dfs -text /user/kfk/data/core-site.xml

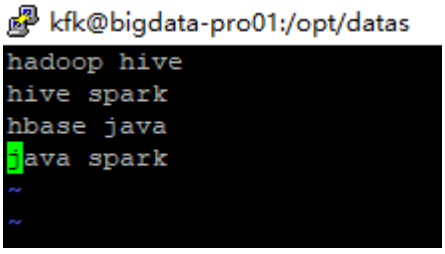
10. 配置 yarn
11. 分发 hadoop 给 2，3 主机：

    scp -r ./* kfk@bigdata-pro02.kfk.com:/opt/modules/hadoop-2.5.0/etc/hadoop/

    scp -r ./* kfk@bigdata-pro03.kfk.com:/opt/modules/hadoop-2.5.0/etc/hadoop/
12. 创建一个数据，进入 datas 目录下

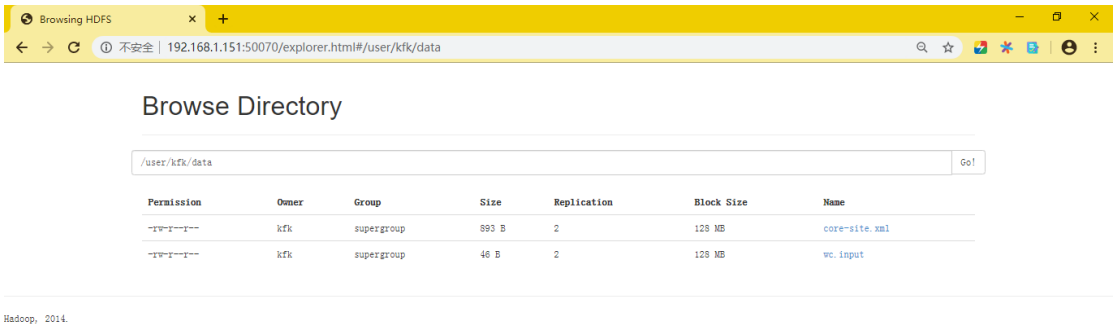
    touch wc.input

    vi wc.input

13. 进入 hadoop-2.5.0 目录，上传文件：

    bin/hdfs dfs -put /opt/datas/wc.input /user/kfk/data



14. 启动主机 1 的 resourcemanager:
    Hadoop-2.5.0 目录下
    sbin/yarn-daemon.sh start resourcemanager
15. 启动主机 1 的 nodemanager:
    sbin/yarn-daemon.sh start nodemanager

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ jps
27676 NameNode
28447 Jps
28417 NodeManager
27764 DataNode
28164 ResourceManager
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

16. 启动主机 2 的 nodemanager:
    sbin/yarn-daemon.sh start nodemanager

```
[kfk@bigdata-pro02 hadoop-2.5.0]$ jps
27847 Jps
27817 NodeManager
23460 DataNode
[kfk@bigdata-pro02 hadoop-2.5.0]$
```

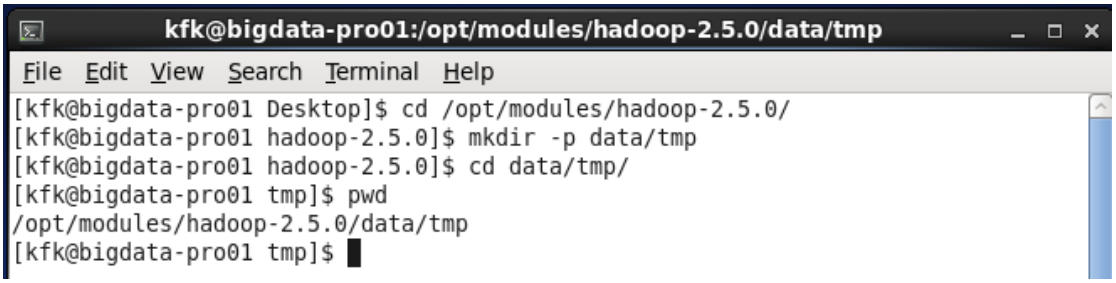
17. 启动主机 3 的 nodemanager:
    sbin/yarn-daemon.sh start nodemanager

```
[kfk@bigdata-pro03 hadoop-2.5.0]$ jps
27597 DataNode
27843 Jps
27812 NodeManager
[kfk@bigdata-pro03 hadoop-2.5.0]$
```

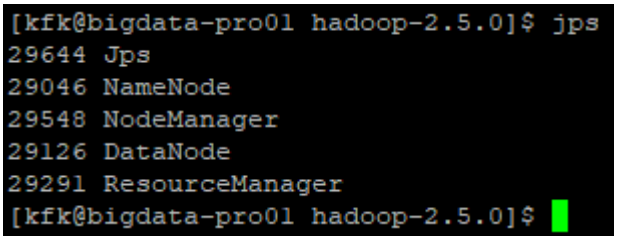
18. 进入 yarn 页面，http://192.168.1.151:8088/
19. 启动主机一的日志文件：
    sbin/mr-jobhistory-daemon.sh start historyserver

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ sbin/mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /opt/modules/hadoop-2.5.0/logs/mapred-kfk-historyserver-bigdata-pro01.kfk.com.out
[kfk@bigdata-pro01 hadoop-2.5.0]$ jps
27676 NameNode
28597 Jps
28568 JobHistoryServer
28417 NodeManager
27764 DataNode
28164 ResourceManager
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

20. 修改配置文件
    新建文件夹 tmp
    /opt/modules/hadoop-2.5.0/data/tmp



21. 删除 2，3 主机上的 hadoop-2.5.0 文件夹，等 1 重新分发
    进入主机 1 的 modules 目录下，分发 hadoop 给 2，3 主机：
    scp -r hadoop-2.5.0/ kfk@bigdata-pro02.kfk.com:/opt/modules/
    scp -r hadoop-2.5.0/ kfk@bigdata-pro03.kfk.com:/opt/modules/
22. 停掉主机一的服务：(主机 2，3 重启)
    sbin/yarn-daemon.sh stop resourcemanager
    sbin/yarn-daemon.sh stop nodemanager
    sbin/mr-jobhistory-daemon.sh stop historyserver
    sbin/hadoop-daemon.sh stop namenode
    sbin/hadoop-daemon.sh stop datanode
23. 重新格式化主机一：
    bin/hdfs namenode -format
24. 重新启动主机一的 hdfs 和 yarn：
    sbin/hadoop-daemon.sh start namenode
    sbin/hadoop-daemon.sh start datanode
25. 启动 2，3 的服务：
    Sbin/Hadoop-daemon.sh start datanode
26. 主机一创建文件和启动服务：
    bin/hdfs dfs -mkdir -p /user/kfk/data/
    sbin/yarn-daemon.sh start resourcemanager
    sbin/yarn-daemon.sh start nodemanager



27. 启动 2，3 的服务：
    sbin/yarn-daemon.sh start nodemanager
    sbin/yarn-daemon.sh start nodemanager

```
[kfk@bigdata-pro02 hadoop-2.5.0]$ jps
2370 DataNode
2484 NodeManager
2586 Jps
[kfk@bigdata-pro02 hadoop-2.5.0]$
```

28. 主机一启动 log 日志：

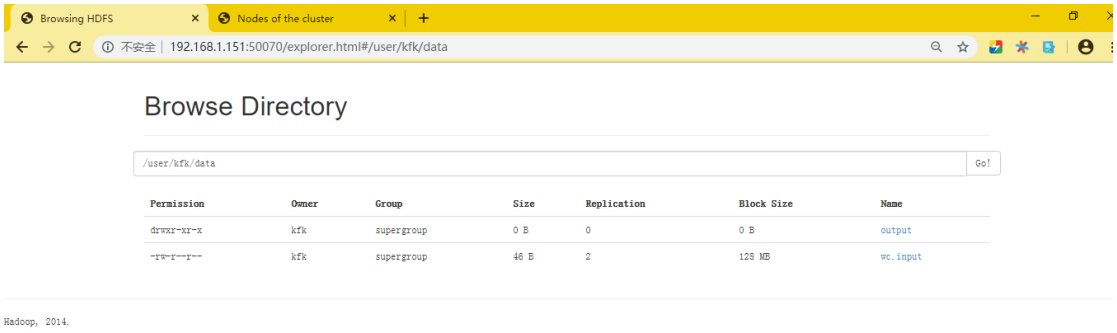
    sbin/mr-jobhistory-daemon.sh start historyserver

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ jps
29046 NameNode
29704 JobHistoryServer
29735 Jps
29548 NodeManager
29126 DataNode
29291 ResourceManager
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

29. Mapreduce 运行：

    上传数据：bin/hdfs dfs -put /opt/datas/wc.input /user/kfk/data/

    创建输出目录：bin/hdfs dfs -mkdir -p /user/kfk/data/output/



    运行 wordcount:

    bin/yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.0.jar wordcount /user/kfk/data/wc.input /user/kfk/data/output/1

    查看结果：

    bin/hdfs dfs -text /user/kfk/data/output/1/par*

```
[kfk@bigdata-pro01 hadoop-2.5.0]$ bin/hdfs dfs -text /user/kfk/data/output/1/par*
20/03/30 17:23:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop  1
hbase   1
hive    2
java    2
spark   2
[kfk@bigdata-pro01 hadoop-2.5.0]$
```

30. Ssh 配置：

    cd ~

    cd .ssh/

    rm -rf known_hosts

ssh-keygen -t rsa(一直回车)
ssh-copy-id bigdata-pro01.kfk.com
ssh-copy-id bigdata-pro02.kfk.com
ssh-copy-id bigdata-pro03.kfk.com
ssh bigdata-pro02.kfk.com
主机 1hadoop-2.5.0 目录下：
sbin/start-dfs.sh
sbin/start-yarn.sh
(jobHistoryserver 手动开的)