Prepear

ubuntu-14.04-server-amd64.iso
http://releases.ubuntu.com/14.04/

hadoop 2.7.3
https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz

hbase 1.3.0
https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz

hive 2.1.1
https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz

zookeeper 3.4.9
https://archive.apache.org/dist/zookeeper/zookeeper-3.4.9/zookeeper-3.4.9.tar.gz

Download Resource

```
cd ~

wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
wget https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz
wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
wget https://archive.apache.org/dist/zookeeper/zookeeper-3.4.9/zookeeper-3.4.9.tar.gz
```

```
root@hadoop-master:~# wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
--2017-02-14 08:08:54--  https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 214092195 (204M) [application/x-gzip]
Saving to: 'hadoop-2.7.3.tar.gz'

100%[======================================================================>] 214,092,195 3.43MB/s   in 1m 58s

2017-02-14 08:10:53 (1.73 MB/s) - 'hadoop-2.7.3.tar.gz' saved [214092195/214092195]

root@hadoop-master:~# wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
--2017-02-14 08:27:31--  https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 149756462 (143M) [application/x-gzip]
Saving to: 'apache-hive-2.1.1-bin.tar.gz'

100%[======================================================================>] 149,756,462 4.99MB/s   in 38s

2017-02-14 08:28:10 (3.74 MB/s) - 'apache-hive-2.1.1-bin.tar.gz' saved [149756462/149756462]

root@hadoop-master:~#
```

Unzip

```
cd ~

tar -xvf hadoop-2.7.3.tar.gz
tar -xvf hbase-1.3.0-bin.tar.gz
tar -xvf apache-hive-2.1.1-bin.tar.gz
tar -xvf zookeeper-3.4.9.tar.gz
```

profile

modify profile

```
vim ~/.bashrc
```

append env settings

```
export SOFT_BASE_PATH=/root
# Hadoop Cluster
export HADOOP_BASE_PATH=/app/hadoop-cluster
export JAVA_HOME=$SOFT_BASE_PATH/jdk1.8.0_66
export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar

export HADOOP_HOME=$HADOOP_BASE_PATH/hadoop-2.7.2
export HADOOP_CONF_DIR=$HADOOP_BASE_PATH/hadoop-2.7.2/etc/hadoop
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

load profile

```
source ~/.bashrc
```

```
java -version
```

hadoop-env.sh，查找JAVA_HOME，并修改

```
cp $HADOOP_HOME/etc/hadoop/hadoop-env.sh $HADOOP_HOME/etc/hadoop/hadoop-env.
sh.bak
vim $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

```
export SOFT_BASE_PATH=/app/soft
export JAVA_HOME=$SOFT_BASE_PATH/jdk1.8.0_66
```

yarn-env.sh，查找JAVA_HOME，并修改

```
vim $HADOOP_HOME/etc/hadoop/yarn-env.sh
```

```
export SOFT_BASE_PATH=/app/soft
export JAVA_HOME=$SOFT_BASE_PATH/jdk1.8.0_66
```

core-site.xml

```
cp $HADOOP_HOME/etc/hadoop/core-site.xml $HADOOP_HOME/etc/hadoop/core-site.xml.bak
vim $HADOOP_HOME/etc/hadoop/core-site.xml
```

```
<configuration>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>file:/tmp/hadoop</value>
        <description>A base for other temporary directories.</description>
    </property>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://hadoop-ubuntu-01:9000</value>
    </property>
    <!-- file system properties -->
    <property>
        <name>fs.default.name</name>
        <value>hdfs://hadoop-ubuntu-01:9000</value>
    </property>
    <property>
        <name>io.file.buffer.size</name>
        <value>131072</value>
    </property>
    <property>
        <name>hadoop.proxyuser.hduser.hosts</name>
        <value>*</value>
    </property>
    <property>
        <name>hadoop.proxyuser.hduser.groups</name>
        <value>*</value>
    </property>
</configuration>
```

```
cp $HADOOP_HOME/etc/hadoop/hdfs-site.xml $HADOOP_HOME/etc/hadoop/hdfs-site.xml.bak
vim $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

```
<configuration>
    <property>
        <name>dfs.namenode.secondary.http-address</name>
        <value>hadoop-ubuntu-01:9001</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:/tmp/hadoop/name</value>
    </property>
    <property>
        <name>dfs.namenode.data.dir</name>
        <value>file:/tmp/hadoop/data</value>
    </property>
    <property>
        <name>dfs.replication</name>
        <value>3</value>
    </property>
    <property>
        <name>dfs.webhdfs.enabled</name>
        <value>true</value>
    </property>
</configuration>
```

```
cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template $HADOOP_HOME/etc/hadoop/mapred-site.xml
vim $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.address</name>
        <value>hadoop-ubuntu-01:10020</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>hadoop-ubuntu-01:19888</value>
    </property>
</configuration>
```

```
cp $HADOOP_HOME/etc/hadoop/yarn-site.xml $HADOOP_HOME/etc/hadoop/yarn-site.xml.bak
vim $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

```
<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-ubuntu-01:8032</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.address</name>
        <value>hadoop-ubuntu-01:8030</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>hadoop-ubuntu-01:8031</value>
    </property>
    <property>
        <name>yarn.resourcemanager.admin.address</name>
        <value>hadoop-ubuntu-01:8033</value>
    </property>
    <property>
        <name>yarn.resourcemanager.webapp.address</name>
        <value>hadoop-ubuntu-01:8088</value>
    </property>
</configuration>
```

修改 hadoop/etc/hadoop/masters 文件，默认如果没有那么就得新建一个，在里面加上一行：

```
cp $HADOOP_HOME/etc/hadoop/masters $HADOOP_HOME/etc/hadoop/masters.bak
vim $HADOOP_HOME/etc/hadoop/masters
```

```
hadoop-ubuntu-01
```

```
cp $HADOOP_HOME/etc/hadoop/slaves $HADOOP_HOME/etc/hadoop/slaves.bak
vim $HADOOP_HOME/etc/hadoop/slaves
```

```
hadoop-ubuntu-02
hadoop-ubuntu-03
```

格式化HDFS系统

```
$HADOOP_HOME/bin/hadoop namenode -format
```

```
scp -r /app/hadoop-cluster root@hadoop-ubuntu-02:/app/
scp -r /app/hadoop-cluster root@hadoop-ubuntu-03:/app/
```

启动dfs，yarn

```
$HADOOP_HOME/sbin/start-dfs.sh
# Master: NameNode, SecondaryNameNode
# Slave: DataNode
jps
$HADOOP_HOME/sbin/start-yarn.sh
# Master: ResourceManager
# Slave: NodeManager
jps
```

查看hdfs的运行状态（On Master）

```
$HADOOP_HOME/bin/hdfs dfsadmin -report
netstat -tnulp | grep java
```

ResourceManager web UI

http://192.168.101.161:8088/

NodeManager Web UI

http://192.168.101.161:8042/

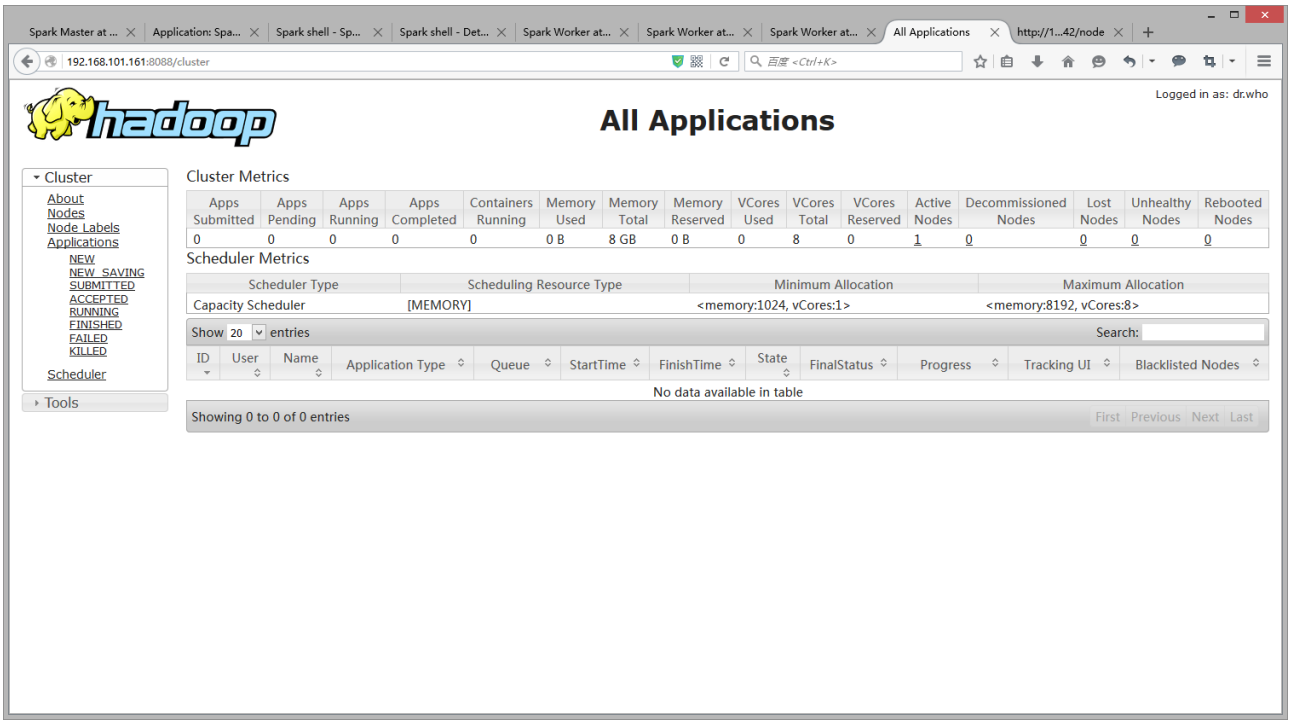
http://192.168.101.162:8042/

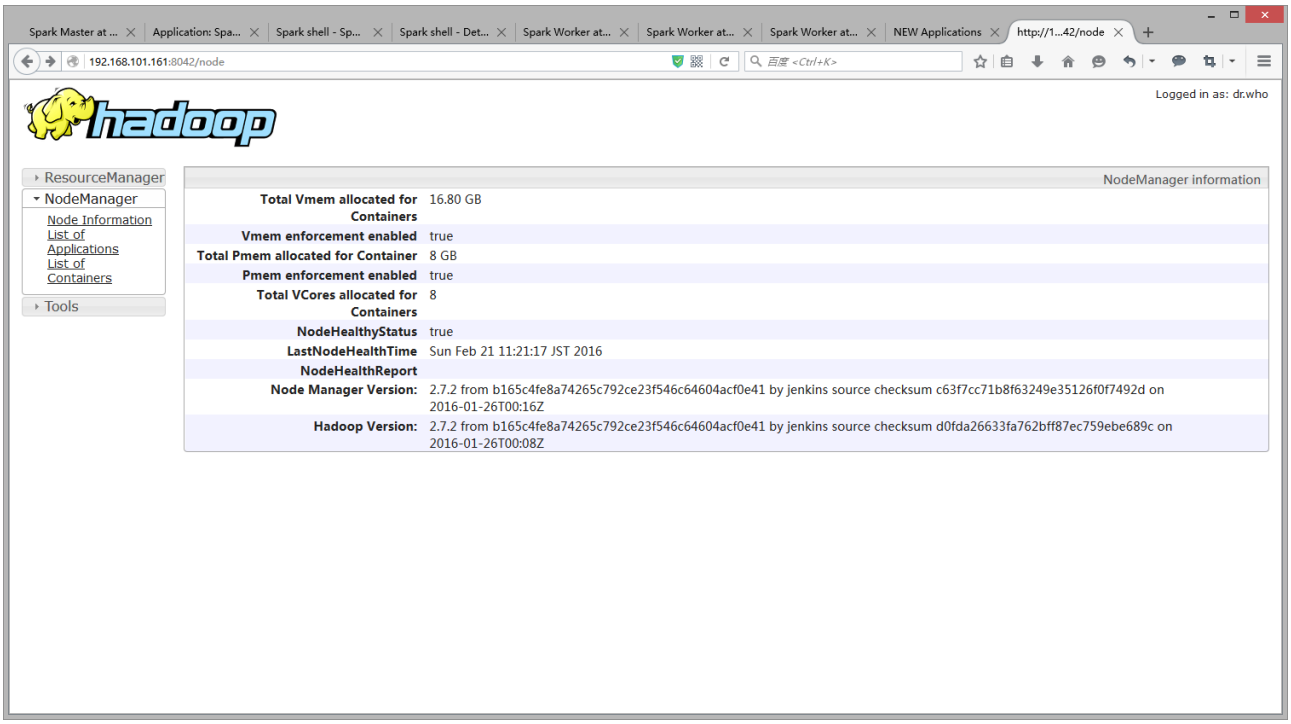
http://192.168.101.163:8042/

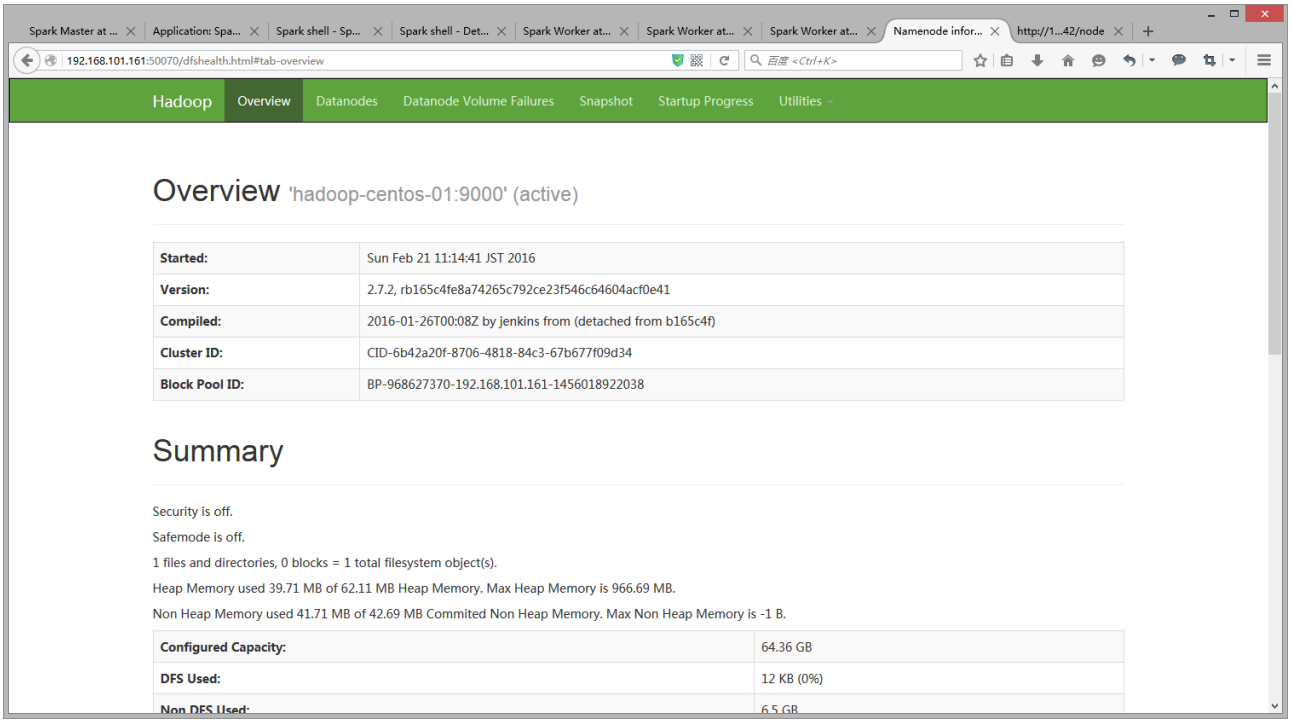
NameNode Web UI

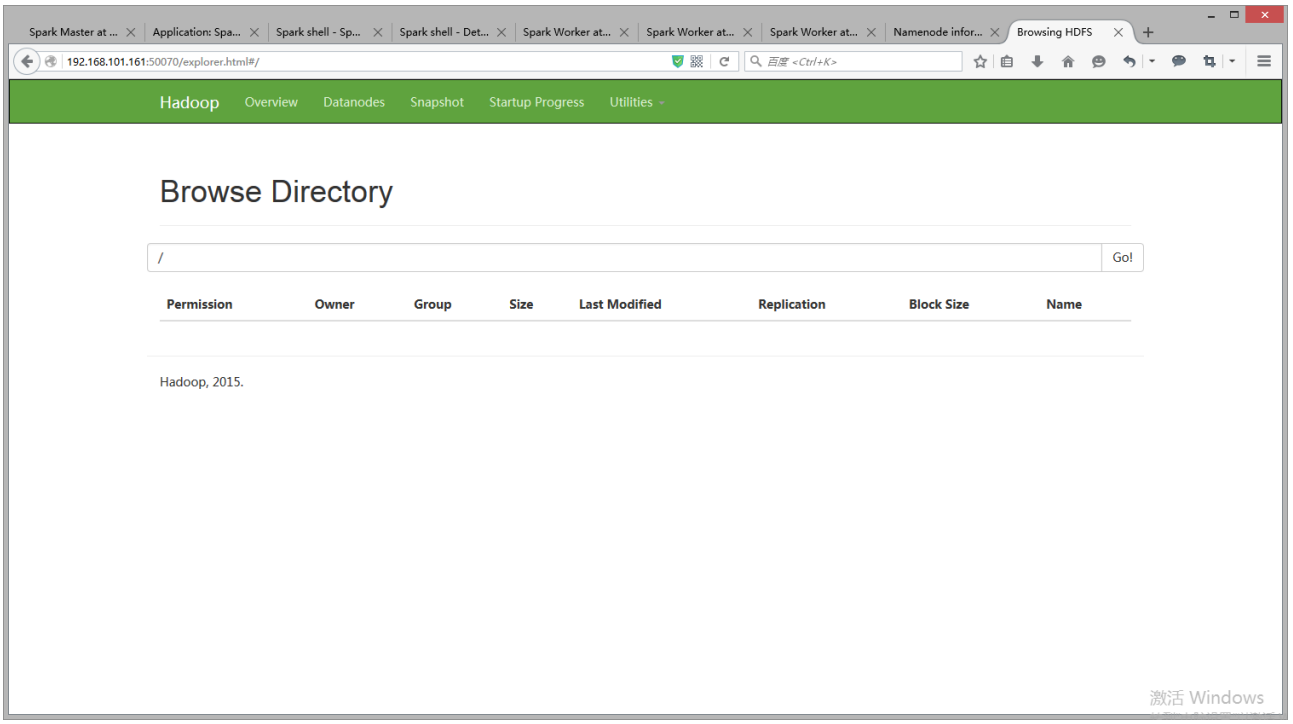
http://192.168.101.161:50070/

http://192.168.101.161:50070/explorer.html#/




Spark Master at ...
Application: Spa...
Spark shell - Sp...
Spark shell - Det...
Spark Worker at...
Spark Worker at...
Spark Worker at...
NEW Applications
http://1...42/node
192.168.101.161:8042/node
Logged in as: dr.who
hadoop
ResourceManager
NodeManager
Node Information
List of Applications
List of Containers
Tools
NodeManager information
Total Vmem allocated for Containers 16.80 GB
Vmem enforcement enabled true
Total Pmem allocated for Container 8 GB
Pmem enforcement enabled true
Total VCores allocated for Containers 8
NodeHealthyStatus true
LastNodeHealthTime Sun Feb 21 11:21:17 JST 2016
NodeHealthReport
Node Manager Version: 2.7.2 from b165c4fe8a74265c792ce23f546c64604acf0e41 by jenkins source checksum c63f7cc71b8f63249e35126f0f7492d on 2016-01-26T00:16Z
Hadoop Version: 2.7.2 from b165c4fe8a74265c792ce23f546c64604acf0e41 by jenkins source checksum d0fda26633fa762bff87ec759ebe689c on 2016-01-26T00:08Z


Spark Master at ...
Application: Spa...
Spark shell - Sp...
Spark shell - Det...
Spark Worker at...
Spark Worker at...
Spark Worker at...
Namenode infor...
http://1...42/node
192.168.101.161:50070/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'hadoop-centos-01:9000' (active)
Started: Sun Feb 21 11:14:41 JST 2016
Version: 2.7.2, rb165c4fe8a74265c792ce23f546c64604acf0e41
Compiled: 2016-01-26T00:08Z by jenkins from (detached from b165c4f)
Cluster ID: CID-6b42a20f-8706-4818-84c3-67b677f09d34
Block Pool ID: BP-968627370-192.168.101.161-1456018922038
Summary
Security is off.
Safemode is off.
1 files and directories, 0 blocks = 1 total filesystem object(s).
Heap Memory used 39.71 MB of 62.11 MB Heap Memory. Max Heap Memory is 966.69 MB.
Non Heap Memory used 41.71 MB of 42.69 MB Commited Non Heap Memory. Max Non Heap Memory is -1 B.
Configured Capacity: 64.36 GB
DFS Used: 12 KB (0%)
Non DFS Used: 6.5 GB




```
1. $HADOOP_HOME/bin/hadoop fs -ls /
2. $HADOOP_HOME/bin/hadoop fs -mkdir /input
3. $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/README.txt /input
4. $HADOOP_HOME/bin/hadoop fs -put /app/backup/sell.txt /tmp
5. $HADOOP_HOME/bin/hadoop fs -ls /input
6. $HADOOP_HOME/bin/hadoop fs -get /tmp/sell.txt /tmp/sell.txt
7. $HADOOP_HOME/bin/hadoop jar /tmp/MultipleThread.jar com.study.hadoop.hw02.FileSystemFileToHDFS /tmp/sell.txt /input/sellnew.txt
```