

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

```
1. cd ~
2.
3. wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3
   .tar.gz
4. wget https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz
5. wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.t
   ar.gz
6. 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

```
1. cd ~
2.
3. tar -xvf hadoop-2.7.3.tar.gz
4. tar -xvf hbase-1.3.0-bin.tar.gz
5. tar -xvf apache-hive-2.1.1-bin.tar.gz
6. tar -xvf zookeeper-3.4.9.tar.gz
```

profile

modify profile

```
1. vim ~/.bashrc
```

append env settings

```
1. export SOFT_BASE_PATH=/root
2. # Hadoop Cluster
3. export HADOOP_BASE_PATH=/app/hadoop-cluster
4. export JAVA_HOME=$SOFT_BASE_PATH/jdk1.8.0_66
5. export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
6.
7. export HADOOP_HOME=$HADOOP_BASE_PATH/hadoop-2.7.2
8. export HADOOP_CONF_DIR=$HADOOP_BASE_PATH/hadoop-2.7.2/etc/hadoop
9. export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

load profile

```
1. source ~/.bashrc
```

```
1. java -version
```

hadoop-env.sh , 查找JAVA_HOME , 并修改

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

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

yarn-env.sh , 查找JAVA_HOME , 并修改

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

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

core-site.xml

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

```
1. <configuration>
2.   <property>
3.     <name>hadoop.tmp.dir</name>
4.     <value>file:/tmp/hadoop</value>
5.     <description>A base for other temporary directories.</description>
6.   </property>
7.   <property>
8.     <name>fs.defaultFS</name>
9.     <value>hdfs://hadoop-ubuntu-01:9000</value>
10.  </property>
11.  <!-- file system properties -->
12.  <property>
13.    <name>fs.default.name</name>
14.    <value>hdfs://hadoop-ubuntu-01:9000</value>
15.  </property>
16.  <property>
17.    <name>io.file.buffer.size</name>
18.    <value>131072</value>
19.  </property>
20.  <property>
21.    <name>hadoop.proxyuser.hduser.hosts</name>
22.    <value>*</value>
23.  </property>
24.  <property>
25.    <name>hadoop.proxyuser.hduser.groups</name>
26.    <value>*</value>
27.  </property>
28. </configuration>
```

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

```
1. <configuration>
2.   <property>
3.     <name>dfs.namenode.secondary.http-address</name>
4.     <value>hadoop-ubuntu-01:9001</value>
5.   </property>
6.   <property>
7.     <name>dfs.namenode.name.dir</name>
8.     <value>file:/tmp/hadoop/name</value>
9.   </property>
10.  <property>
11.    <name>dfs.namenode.data.dir</name>
12.    <value>file:/tmp/hadoop/data</value>
13.  </property>
14.  <property>
15.    <name>dfs.replication</name>
16.    <value>3</value>
17.  </property>
18.  <property>
19.    <name>dfs.webhdfs.enabled</name>
20.    <value>true</value>
21.  </property>
22. </configuration>
```

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

```
1. <configuration>
2.   <property>
3.     <name>mapreduce.framework.name</name>
4.     <value>yarn</value>
5.   </property>
6.   <property>
7.     <name>mapreduce.jobhistory.address</name>
8.     <value>hadoop-ubuntu-01:10020</value>
9.   </property>
10.  <property>
11.    <name>mapreduce.jobhistory.webapp.address</name>
12.    <value>hadoop-ubuntu-01:19888</value>
13.  </property>
14. </configuration>
```

```
1. cp $HADOOP_HOME/etc/hadoop/yarn-site.xml $HADOOP_HOME/etc/hadoop/yarn-site.x
   ml.bak
2. vim $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

```

1. <configuration>
2.   <property>
3.     <name>yarn.nodemanager.aux-services</name>
4.     <value>mapreduce_shuffle</value>
5.   </property>
6.   <property>
7.     <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
8.     <value>org.apache.hadoop.mapred.ShuffleHandler</value>
9.   </property>
10.  <property>
11.    <name>yarn.resourcemanager.address</name>
12.    <value>hadoop-ubuntu-01:8032</value>
13.  </property>
14.  <property>
15.    <name>yarn.resourcemanager.scheduler.address</name>
16.    <value>hadoop-ubuntu-01:8030</value>
17.  </property>
18.  <property>
19.    <name>yarn.resourcemanager.resource-tracker.address</name>
20.    <value>hadoop-ubuntu-01:8031</value>
21.  </property>
22.  <property>
23.    <name>yarn.resourcemanager.admin.address</name>
24.    <value>hadoop-ubuntu-01:8033</value>
25.  </property>
26.  <property>
27.    <name>yarn.resourcemanager.webapp.address</name>
28.    <value>hadoop-ubuntu-01:8088</value>
29.  </property>
30. </configuration>

```

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

```

1. cp $HADOOP_HOME/etc/hadoop/masters $HADOOP_HOME/etc/hadoop/masters.bak
2. vim $HADOOP_HOME/etc/hadoop/masters

```

```

1. hadoop-ubuntu-01

```

```

1. cp $HADOOP_HOME/etc/hadoop/slaves $HADOOP_HOME/etc/hadoop/slaves.bak
2. vim $HADOOP_HOME/etc/hadoop/slaves

```

```

1. hadoop-ubuntu-02
2. hadoop-ubuntu-03

```

格式化HDFS系统

```

1. $HADOOP_HOME/bin/hadoop namenode -format

```

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

启动dfs , yarn

1. `$HADOOP_HOME/sbin/start-dfs.sh`
2. `# Master: NameNode, SecondaryNameNode`
3. `# Slave: DataNode`
4. `jps`
5. `$HADOOP_HOME/sbin/start-yarn.sh`
6. `# Master: ResourceManager`
7. `# Slave: NodeManager`
8. `jps`

查看hdfs的运行状态 (On Master)

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

ResourceManager web UI

<http://192.168.101.161:8088/>

NodeManager Web UI

<http://192.168.101.161:8042/>

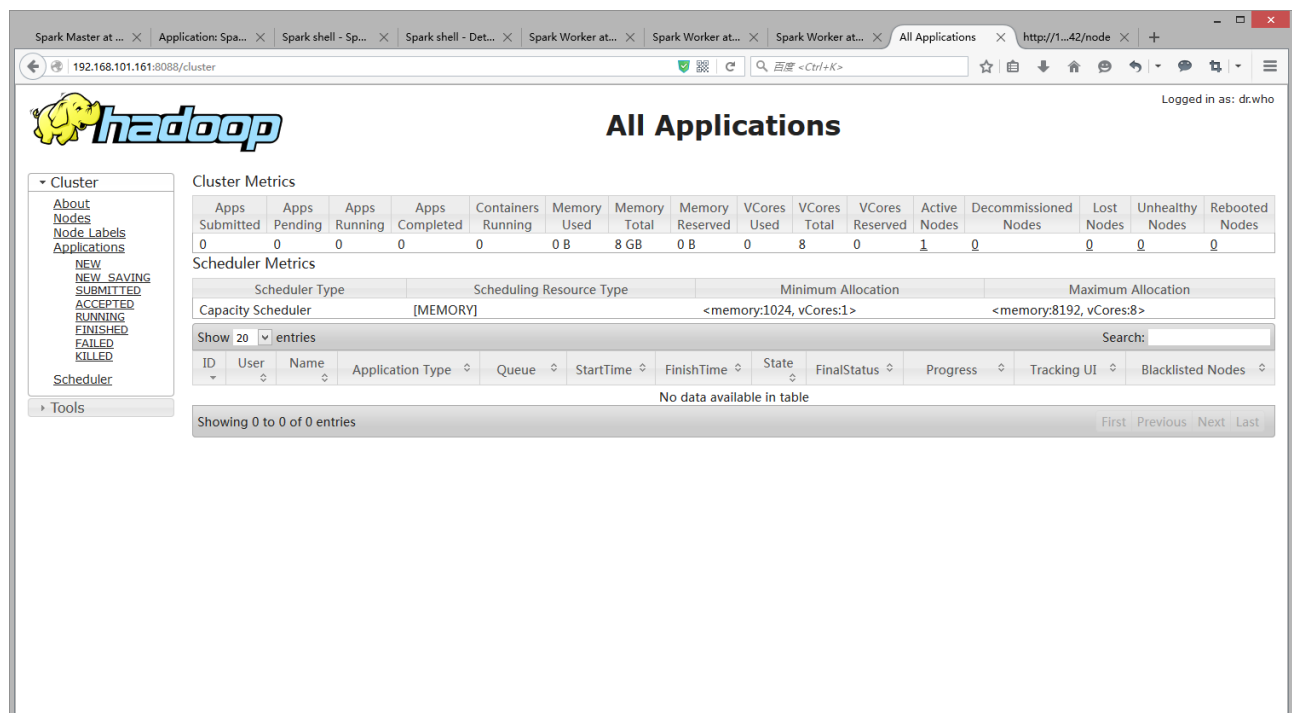
<http://192.168.101.162:8042/>

<http://192.168.101.163:8042/>

NameNode Web UI

<http://192.168.101.161:50070/>

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



The screenshot displays the Hadoop All Applications web UI. The page includes a sidebar with navigation links for Cluster, About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, and Scheduler. The main content area shows Cluster Metrics, Scheduler Metrics, and a table of applications. The table is currently empty, showing 'No data available in table'.

Cluster Metrics															
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Scheduler Metrics			
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
No data available in table											

Showing 0 to 0 of 0 entries

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

hadoop

Logged in as: dr.who

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

🔍

🔍 百度 <Ctrl+K>

☆

📄

⬇️

🏠

💬

↶

↷

🔗

⌵

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.

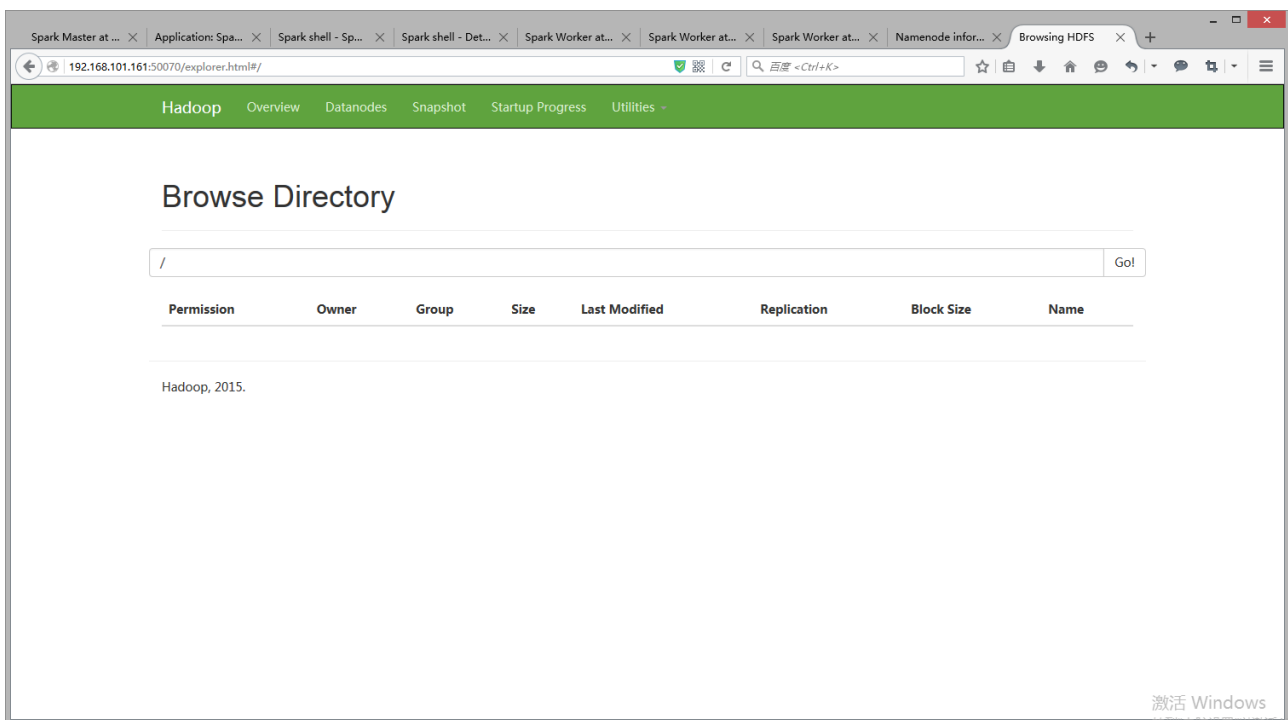
Safemode is off.

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 Committed 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`

