**5-步骤4-Hadoop的安装与配置**

**一、Hadoop安装包解压**

复制并解压Hadoop软件包，命令如下：

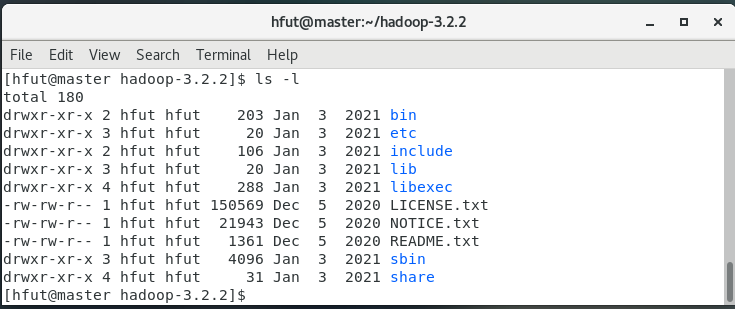
[hfut@master ~]$ cp ~/5-resources/hadoop-3.2.2.tar.gz ~/

[hfut@master ~]$ tar -zxvf hadoop-3.2.2.tar.gz

[hfut@master ~]$ rm hadoop-3.2.2.tar.gz

[hfut@master ~]$ cd hadoop-3.2.2

ls -l看到如下图的内容，表示解压成功：



**二、配置**

**（1）配置环境变量hadoop-env.sh**

环境变量文件中，只需要配置JDK的路径。

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/hadoop-env.sh

在文件中找到下面的一行代码：

#export JAVA\_HOME =

将这行代码修改为下面的代码：

export JAVA\_HOME=/usr/java/jdk1.8.0\_301/

在文件中找到下面的一行代码：

#export HADOOP\_HOME =

将这行代码修改为下面的代码：

export HADOOP\_HOME=/home/hfut/hadoop-3.2.2/

然后保存文件。

**（2）配置核心组件core-site.xml**

使用gedit编辑：

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/core-site.xml

用下面的代码替换core-site.xml中的内容：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/hfut/hadoopdata</value>

</property>

</configuration>

**（3）配置文件系统hdfs-site.xml**

使用gedit编辑：

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/hdfs-site.xml

用下面的代码替换hdfs-site.xml中的内容：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

**（4）配置文件系统yarn-site.xml**

使用gedit编辑：

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/yarn-site.xml

用下面的代码替换yarn-site.xml中的内容：

<?xml version="1.0"?>

<!--

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License. See accompanying LICENSE file.

-->

<configuration>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>master</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

**（5）配置计算框架mapred-site.xml**

使用gedit编辑：

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/mapred-site.xml

用下面的代码替换mapred-site.xml中的内容：

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>mapreduce.jobhistory.address</name>

<value>master:10020</value>

</property>

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>master:19888</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/hfut/hadoop-3.2.2</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/hfut/hadoop-3.2.2</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP\_MAPRED\_HOME=/home/hfut/hadoop-3.2.2</value>

</property>

</configuration>

**（6）在master节点配置workers文件**

使用gedit编辑：

[hfut@master Hadoop-3.2.2]$ gedit ~/hadoop-3.2.2/etc/hadoop/workers

用下面的代码替换workers中的内容：

slave

**三、复制到从节点**

使用下面的命令将已经配置完成的Hadoop复制到从节点HadoopSlave上：

[hfut@master Hadoop-3.2.2]$ cd

[hfut@master ~]$ scp -r ~/hadoop-3.2.2/ hfut@slave:~/

**四、启动Hadoop集群**

下面所有的操作都要在hfut用户下，切换到hfut用户的命令是：

su hfut

密码是hfut

**（1）配置Hadoop启动的系统环境变量**

该节的配置需要同时在两个节点（HadoopMaster和HadoopSlave）上进行操作，操作命令如下：

[hfut@master Hadoop-3.2.2]$ cd

[hfut@master ~]$ gedit ~/.bash\_profile

将下面的代码追加到.bash\_profile末尾：

#HADOOP

export HADOOP\_HOME=/home/hfut/hadoop-3.2.2

export PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$PATH

然后执行命令：

[hfut@master ~]$ source ~/.bash\_profile

**（2）创建数据目录**

该节的配置需要同时在两个节点（HadoopMaster和HadoopSlave）上进行操作，操作命令如下：

在hfut的用户主目录下，创建数据目录，命令如下：

[hfut@master ~]$ mkdir /home/hfut/hadoopdata

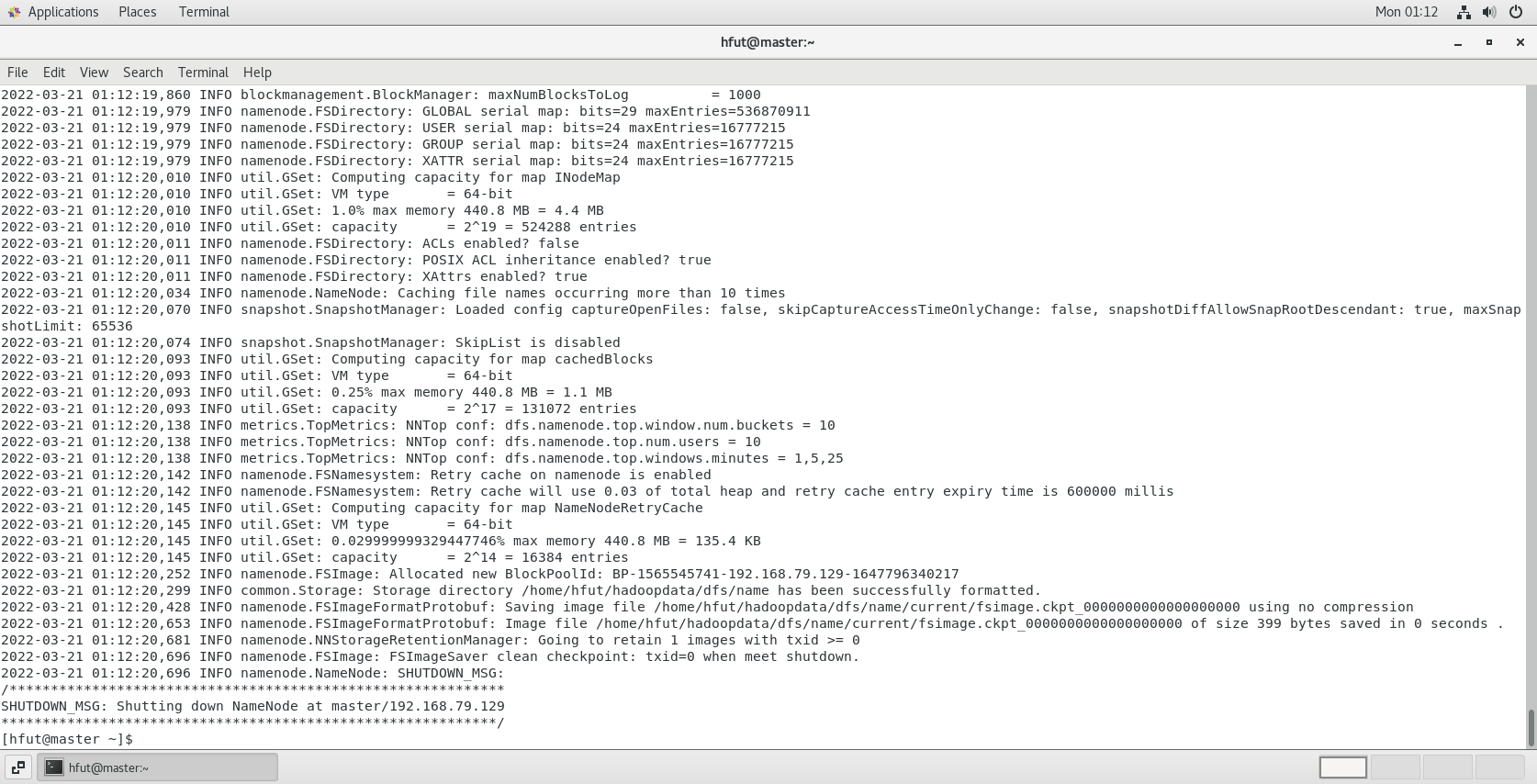
**（3）启动Hadoop集群**

① 格式化文件系统

格式化命令如下，该操作需要在HadoopMaster节点上执行：

[hfut@master ~]$ hdfs namenode -format

看到下图的打印信息表示格式化成功，如果出现Exception/Error，则表示出现问题。



② 启动Hadoop

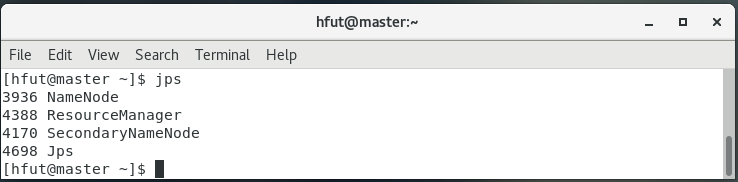
使用start-all.sh启动Hadoop集群，首先进入Hadoop安装主目录，然后执行启动命令：

[hfut@master ~]$ start-all.sh

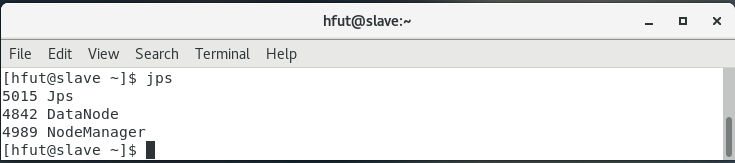
执行命令后，提示输入yes/no时，输入yes。

③ 查看进程是否启动

在HadoopMaster的终端执行jps命令，在打印结果中会看到4个进程，分别是ResourceManager、Jps、NameNode和SecondaryNameNode，如下图所示。如果出现这4个进程表示主节点进程启动成功。

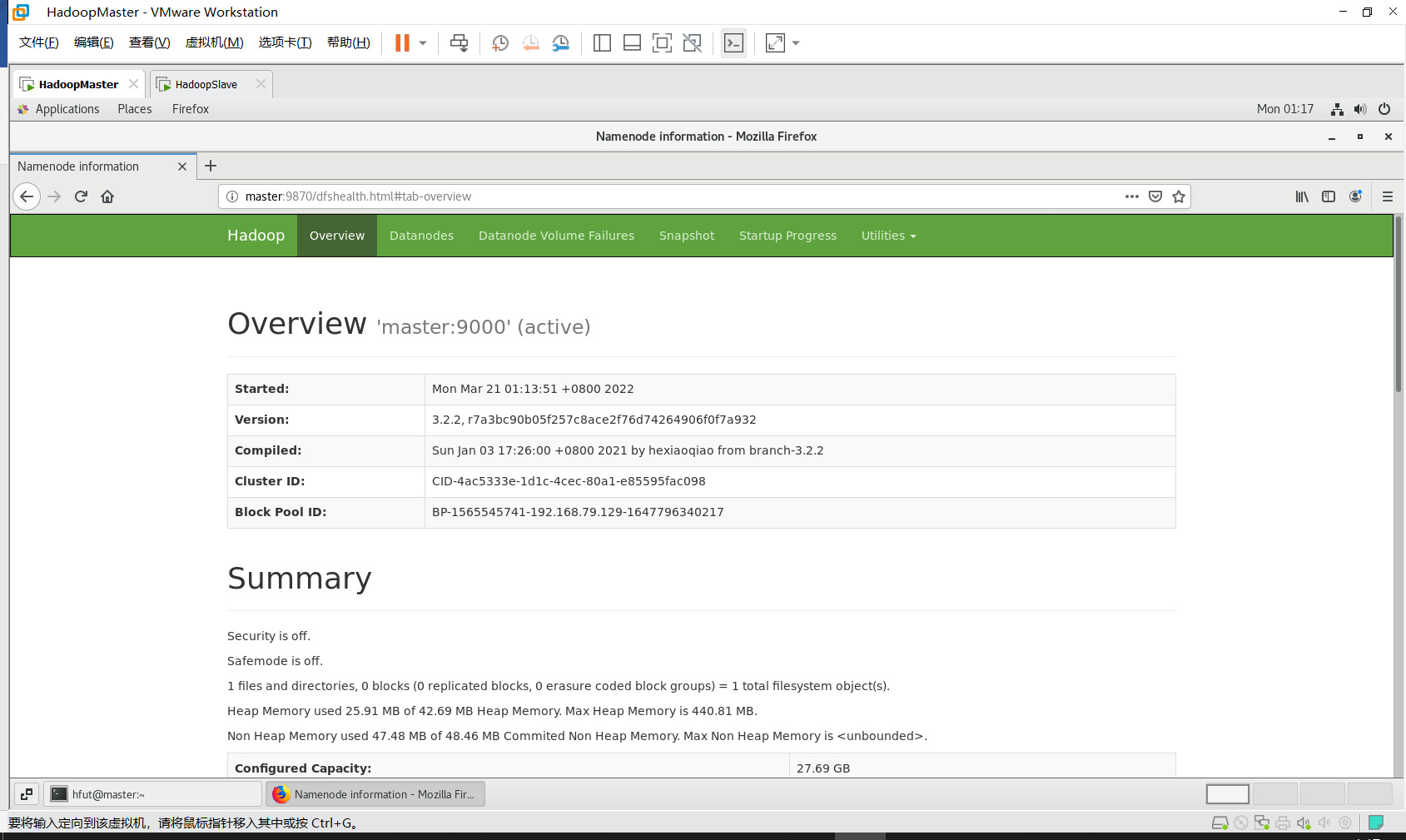


在HadoopSlave的终端执行jps命令，在打印结果中会看到3个进程，分别是Nodemanager、Datanode和Jps，如下图所示。如果出现这3个进程表示主节点进程启动成功。

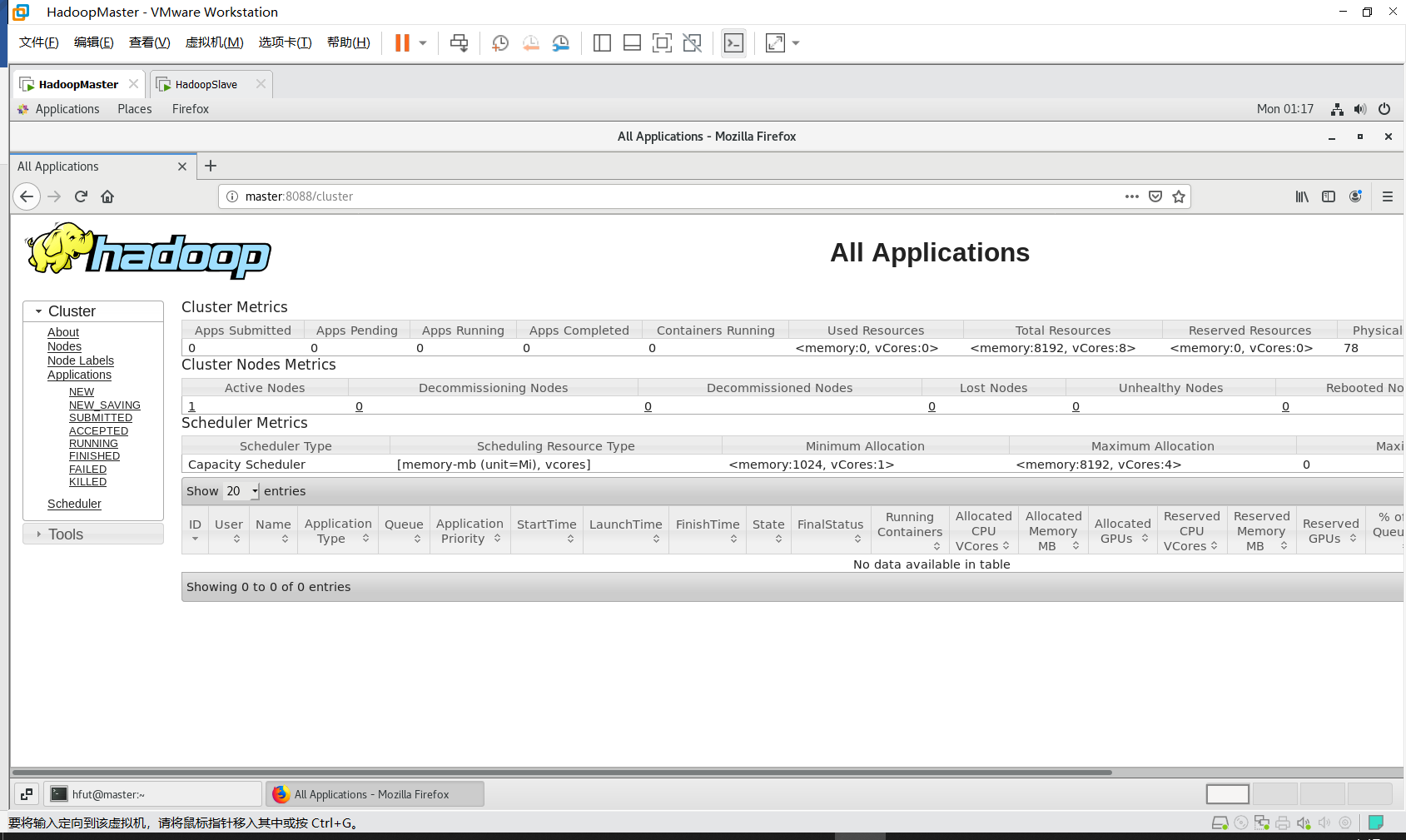


④ Web UI查看集群是否成功启动

在HadoopMaster上启动Firefox浏览器，在浏览器地址栏中输入<http://master:9870/>，检查namenode和datanode是否正常。UI页面如下图所示。



在HadoopMaster上启动Firefox浏览器，在浏览器地址栏中输入<http://master:8088/>，检查Yarn是否正常。页面如下图所示。



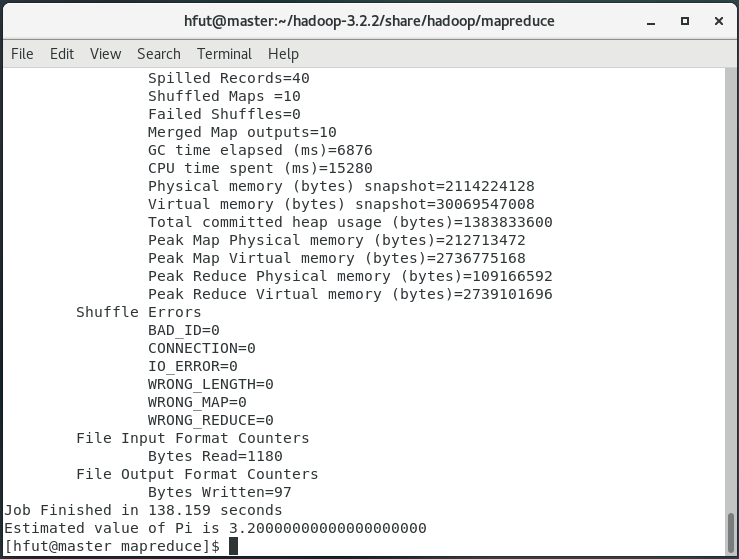
⑤ 运行PI实例检查Hadoop集群是否搭建成功

进入Hadoop安装主目录，执行下面的命令：

[hfut@master ~]$ cd ~/hadoop-3.2.2/share/hadoop/mapreduce/

[hfut@master mapreduce]$ hadoop jar ~/hadoop-3.2.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar pi 10 10

会看到如下的执行结果：



最后输出：

Estimated value of Pi is 3.20000000000000000000

如果以上的3个验证步骤都没有问题，说明集群正常启动。