1. 虚拟机安装

安装vmware

1. 镜像安装

安装ContOS6.7-x86\_64-bin-DVD1镜像

minmal是最小安装版，类似ubuntu的server版，DVD版附带桌面，更容易操作

1. 系统初步配置
2. 用户权限赋予
3. 网络静态IP与各主机之间IP映射配置

网卡目录 /etc/sysconfig/network-scripts/ifcfg-eth0

网卡设备名称与mac地址的对应关系目录 /etc/udev/rules. 70-persistent-net.rules

主机名 /etc/sysconfig/network

IP映射文件 /etc/hosts

四、安装vmtools

五、完成虚拟机与xshell的连接

六、安装jdk

a.

update-alternatives --install /usr/bin/java java /opt/jdk1.8.0\_181/bin/java 300

update-alternatives --install /usr/bin/javac javac /opt/jdk1.8.0\_181/bin/javac 300

update-alternatives --install /usr/bin/jar jar /opt/jdk1.8.0\_181/bin/jar 300

update-alternatives --config java

b. source /etc/profile

1. # 配置 JDK 系统环境变量

export JAVA\_HOME=/export/servers/jdk

export PATH=$PATH:$JAVA\_HOME/bin

export CLASSPATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

七、SSH免登录配置

ssh-keygen -t rsa

ssh-copy-id Hadoop00X

一定要给自己也ssh一下！

八、Hadoop安装

# 配置Hadoop系统环境变量

export HADOOP\_HOME=/export/servers/hadoop-2.7.4

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

source /etc/profile

九、Hadoop集群配置

a.配置Hadoop集群主节点

/etc/hadoop文件夹下

1. 修改hadoop-env.sh文件

JAVA\_HOME

export JAVA\_HOME=/export/servers/jdk

2.修改core-site.xml文件

<configuration>

<!--开始修改-->

<!--用于设置Hadoop的文件系统,由URI指定-->

<property>

<name>fs.defaultFS</name>

<!--用于指定namenode地址在Hadoop001虚拟机上-->

<value>hdfs://Hadoop001:9000</value>

</property>

<!--配置Hadoop的临时目录,默认/tmp/hadoop-${user.name}-->

<property>

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

<value>/export/servers/hadoop-2.7.4/tmp</value>

</property>

</configuration>

3.修改hdfs-site.xml文件

<configuration>

<!--指定HDFS副本的数量-->

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<!--secondary namenode 所在主机的IP和端口-->

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>Hadoop002:50090</value>

</property>

</configuration>

4.修改mapred-site.xml 文件

<configuration>

<!--指定MapReduce运行的框架，这里指定在YARM上,默认是local-->

<property>

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

<value>yarn</value>

</property>

</configuration>

5.修改yarn-site.xml文件

<configuration>

<!-- Site specific YARN configuration properties -->

<!--指定YARN集群的管理者(ResourceManager)的地址-->

<property>

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

<value>Hadoop001</value>

</property>

<property>

<!--定义Nodemanager上要提供给正在运行的容器的全部可用大小-->

<name>yarn.nodemanager.resource.memory-mb</name>

<value>2048</value>

</property>

<property>

<!--资源管理器中分配给每个容器请求的最小内存限制-->

<name>yarn.scheduler.minimum-allocation-mb</name>

<value>2048</value>

</property>

<property>

<!--NodeManager可以分配的CPU核数-->

<name>yarn.nodemanager.resource.cpu-vcores</name>

<value>1</value>

</property>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>

6.修改slaves文件

Hadoop001

Hadoop002

Hadoop003

b.将集群主节点的配置文件分发到其它子节点

scp /etc/profile Hadoop003:/etc/profile

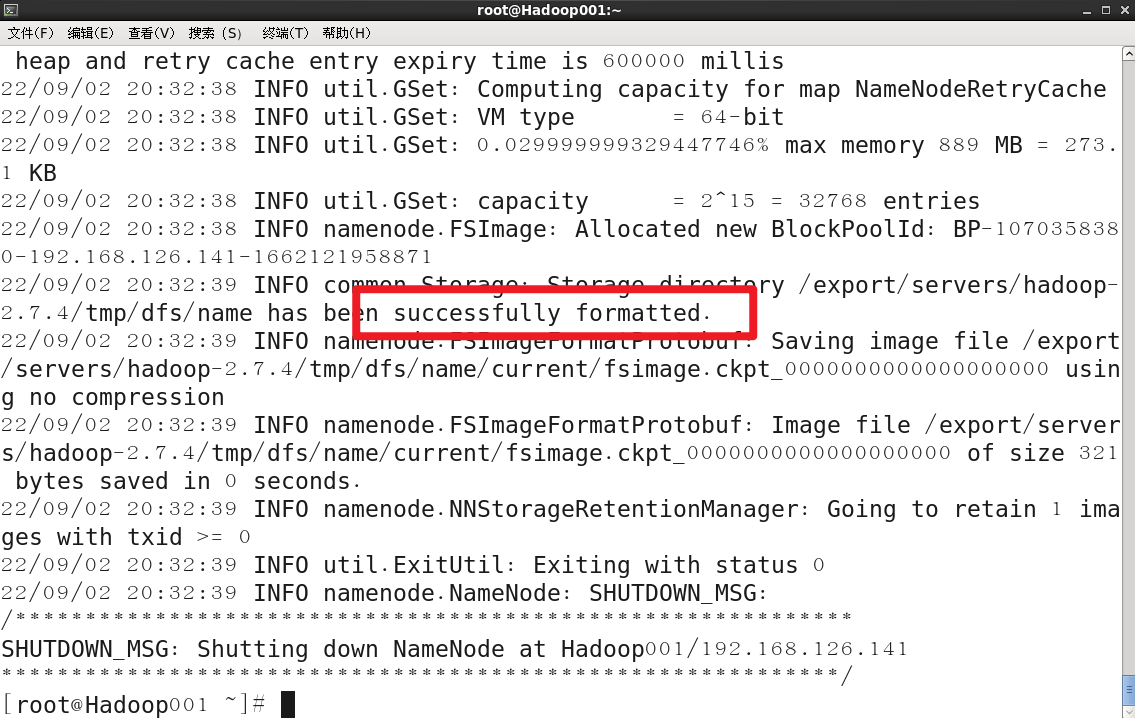
scp -r /export/ Hadoop003:/

十、Hadoop集群测试

a.格式化文件系统

对主节点进行格式化处理 hdfs namenode -format

或者 hadoop namenode -format



b.启动和关闭Hadoop集群

1.单节点逐个启动和关闭

hadoop-daemon.sh start namenode

hadoop-daemon.sh start datanode

yarn-daemon.sh start resourcemanager

yarn-daemon.sh start nodemanager

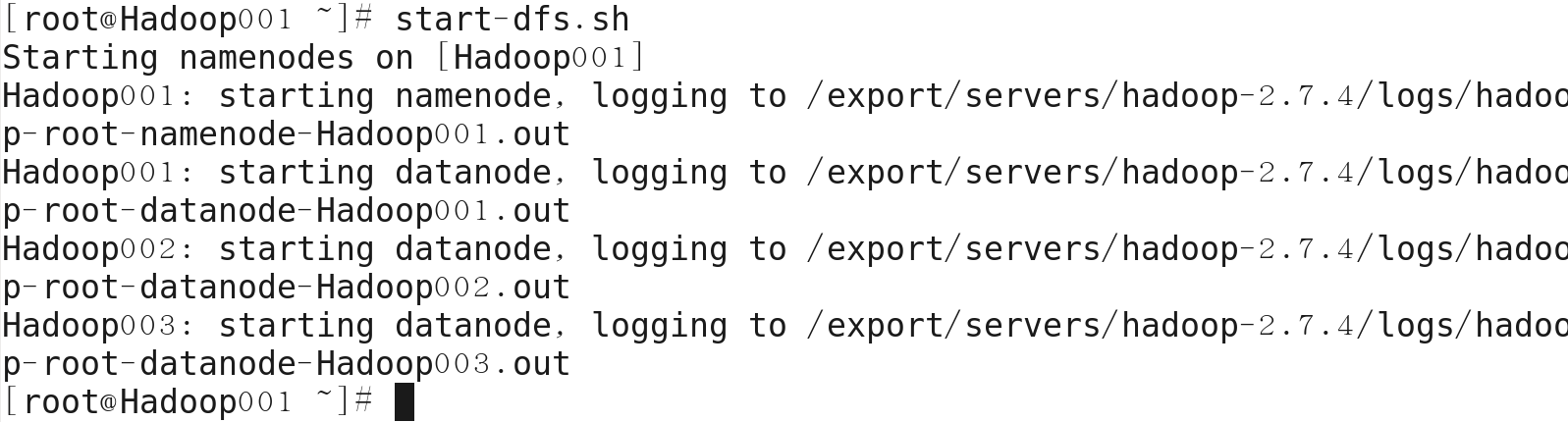
hadoop-daemon.sh start secondarynamenode

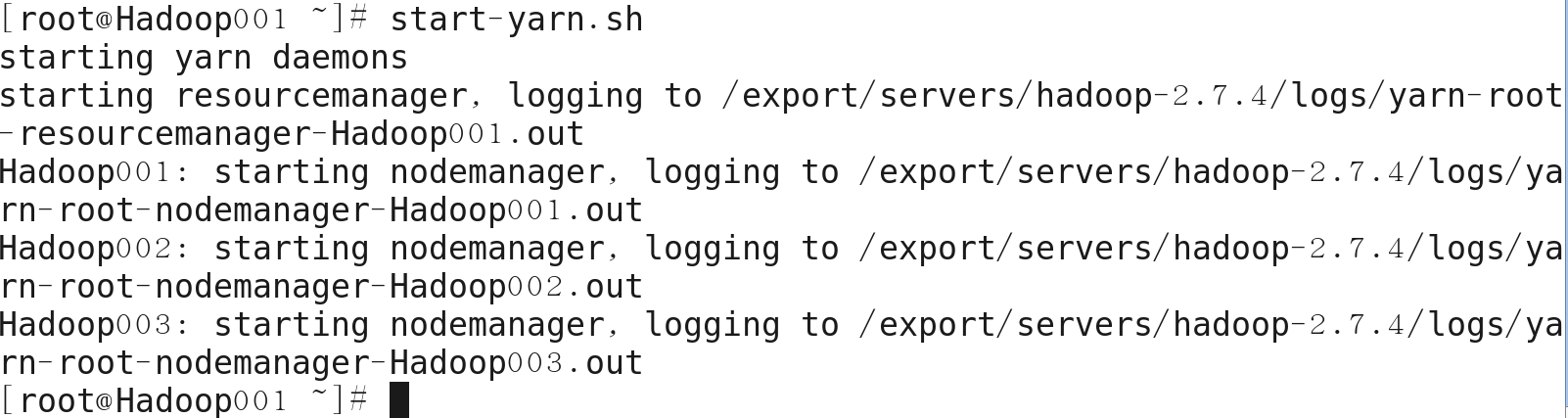
2.脚本一键启动和关闭

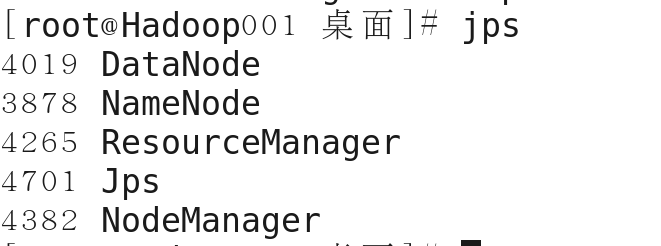
start-dfs.sh

start-yarn.sh

查看 – jps











问题：1.NodeManager启动后又停止

2.secondarynamenode 无法启TMMD端口号多加了个0

c.通过UI查看Hadoop运行状态

1.\Windows\System32\drivers\etc\hosts 文件添加IP映射

(重启主机)

2.关闭所有防火墙的集群映射

service iptables stop -关闭防火墙

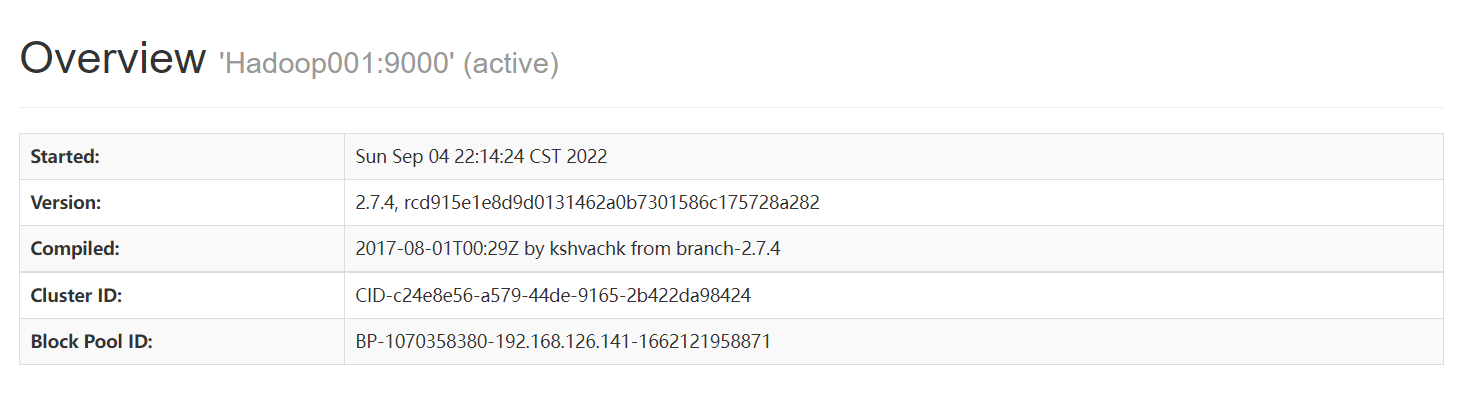
chkconfig iptables off -关闭开机自启

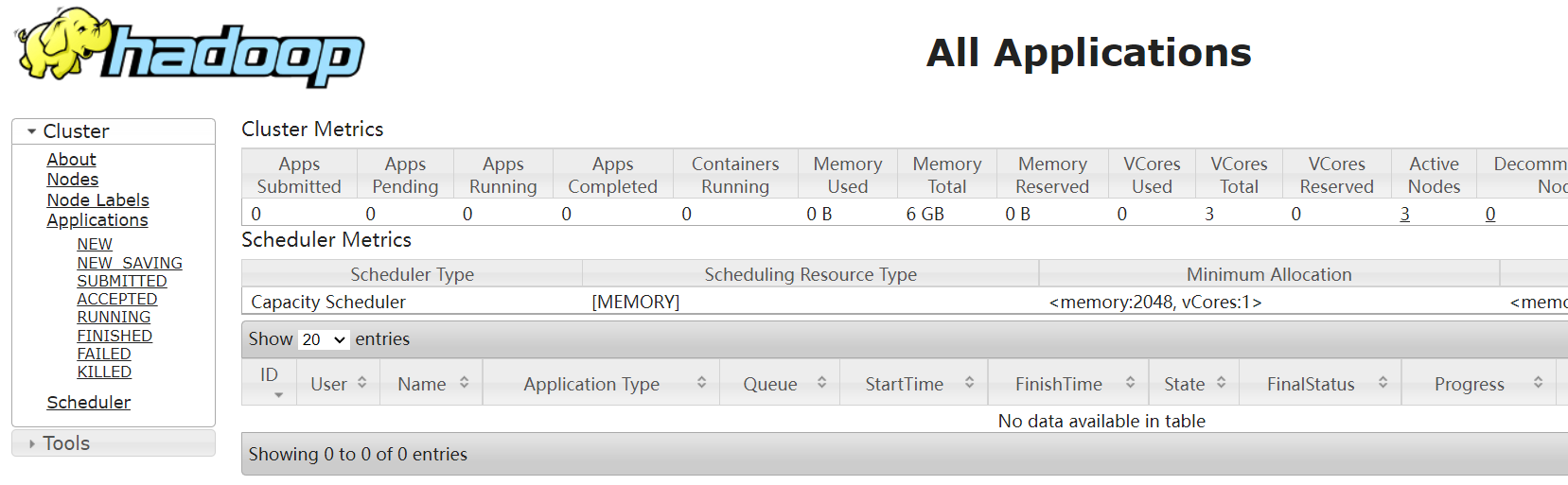
3.主机通过

<http://Hadoop001:50070> 查看HDFS

<http://Hadoop001:8088> 查看YARN

集群状态





十一、Hadoop集群初体验

hadoop fs -mkdir -p /wordcount/input

hadoop fs -put /export/data/word.txt /wordcount/input

Hadoop解压包中的/share/Hadoop/mapreduce/目录下ll

hadoop-mapreduce-examples-2.7.4.jar 单词计数例子

hadoop jar hadoop-mapreduce-examples-2.7.4.jar wordcount /wordcount/input /wordcount/output

