# 实验一 安装Hadoop

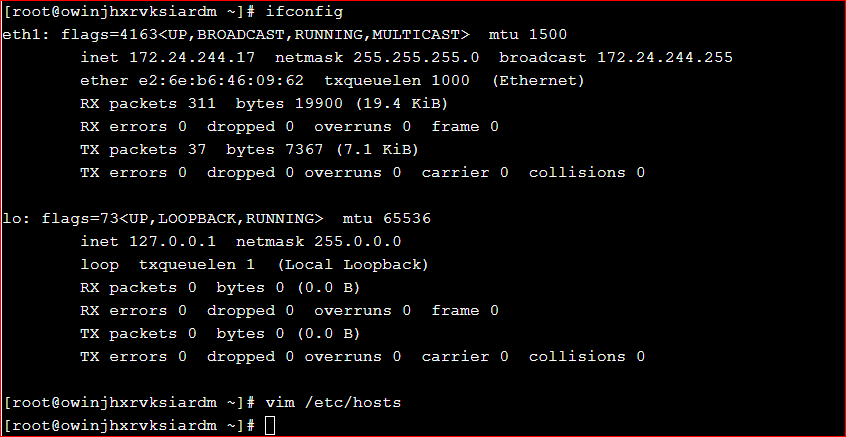
**·实验步骤**

1. 配置hosts文件

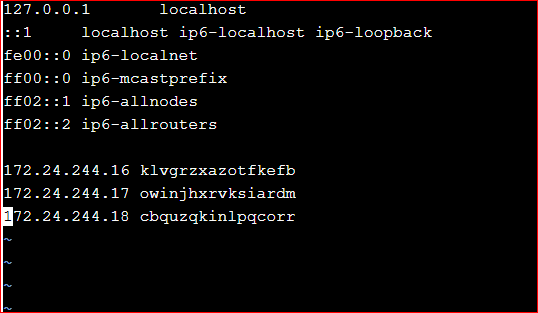
（三台服务器：server-1、server-2、server-3）

1.1、将各服务器的IP和名称配置到/etc/hosts文件中，可以让服务器之间通过服务器名称来访问。

|  |
| --- |
| 使用ifconfig命令查看本机IP  使用hostosname查看本机服务器名称 |



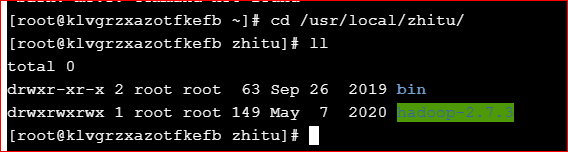
|  |
| --- |
| 使用vim命令进入/etc/hosts文件，将三台服务器名称的IP和名称写入文件  三台服务器均需要配置 |



1. 下载Hadoop

2.1、下载好后已解压到目录/usr/local/zhitu/下

|  |
| --- |
| cd进入目录/usr/local/zhitu/查看 |



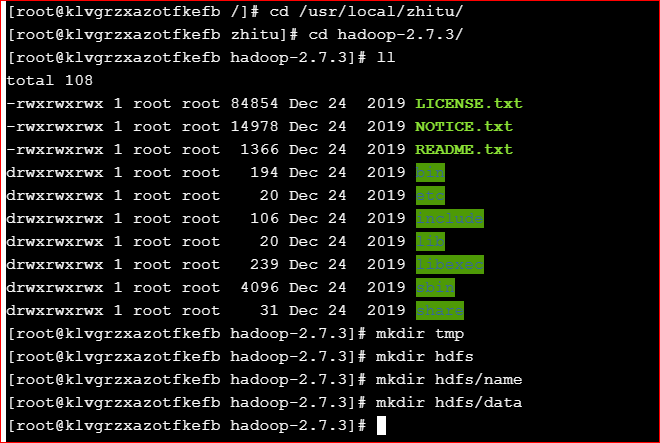
1. 配置Hadoop

3.1、在Hadoop-2.7.3目录下创建4个目录：tmp、hdfs、hdfs/name、hdfs/data

tmp目录用来存储hadoop的临时文件

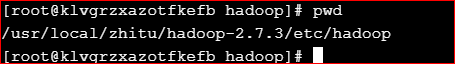
hdfs/data目录用来存储hdfs的文件块

hdfs/name目录用来存储元数据信息



3.2、配置core-site.xml文件

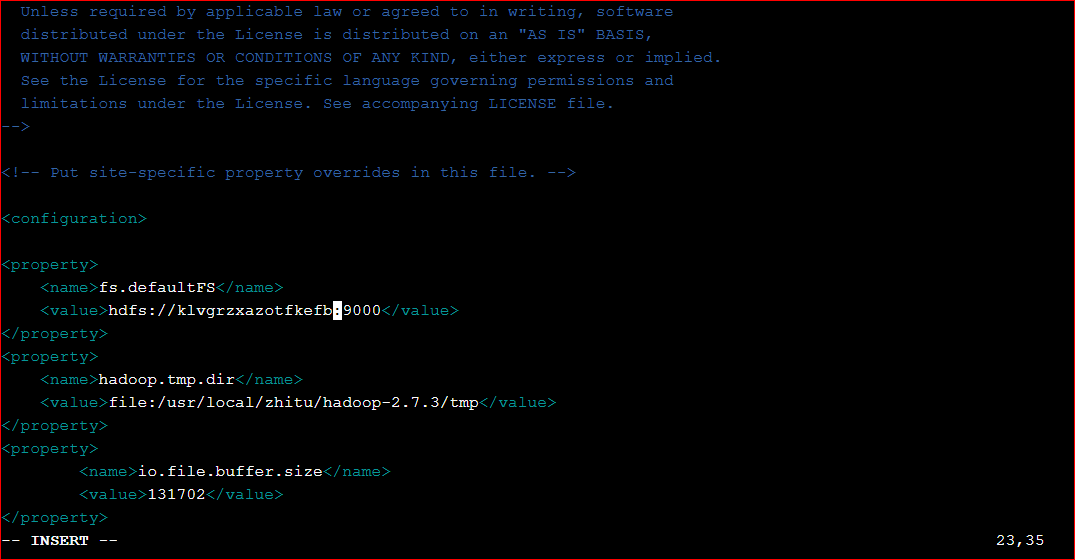
|  |
| --- |
| 使用pwd查看当前所在位置 |



进入hadoop-2.7.3目录下的etc/hadoop目录下，修改core-site.xml文件，在配置中增加如下配置：

|  |
| --- |
| <property>  <name>fs.defaultFS</name>  <value>hdfs://server-1:9000</value>  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>file:/usr/local/zhitu/hadoop-2.7.3/tmp</value>  </property>  <property>  <name>io.file.buffer.size</name>  <value>131702</value>  </property> |

5



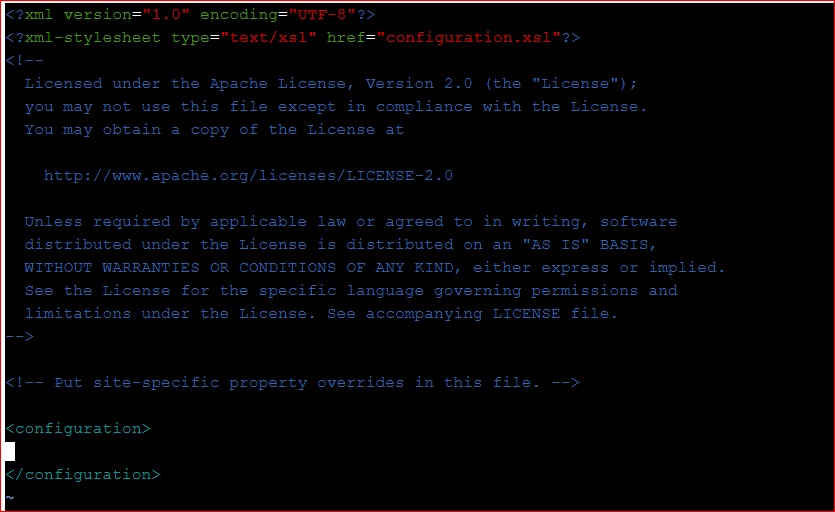
注意主服务器名称

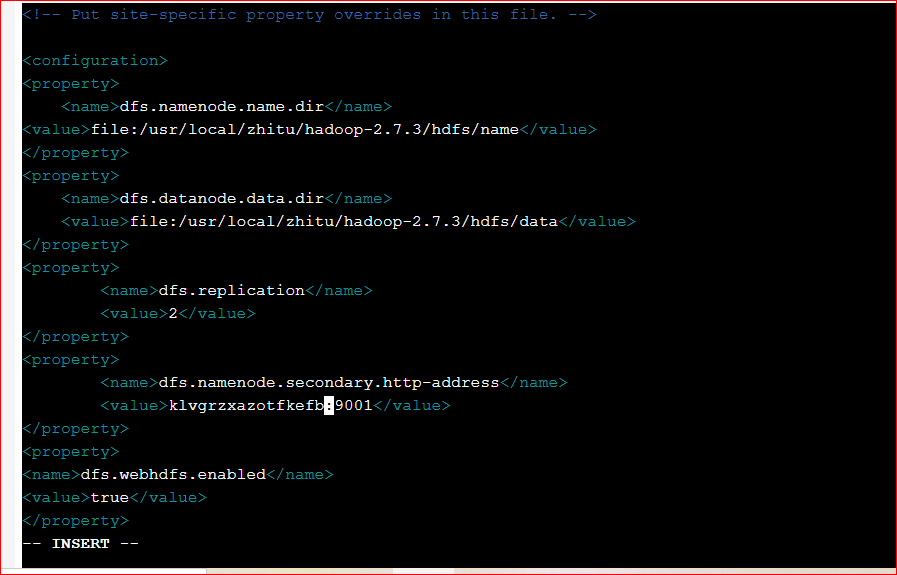
3.3、配置配置hdfs-site.xml文件

进入hadoop-2.7.3目录下的etc/hadoop目录下，修改hdfs-site.xml文件，该文件主要是hdfs系统的相关配置，在配置中增加如下配置:

|  |
| --- |
| ​<property>      <name>dfs.namenode.name.dir</name>  <value>file:/usr/local/zhitu/hadoop-2.7.3/hdfs/name</value>  </property>  <property>      <name>dfs.datanode.data.dir</name>      <value>file:/usr/local/zhitu/hadoop-2.7.3/hdfs/data</value>  </property>  <property>          <name>dfs.replication</name>          <value>2</value>  </property>  <property>          <name>dfs.namenode.secondary.http-address</name>          <value>server-1:9001</value>  </property>  <property>  <name>dfs.webhdfs.enabled</name>  <value>true</value>  </property> |

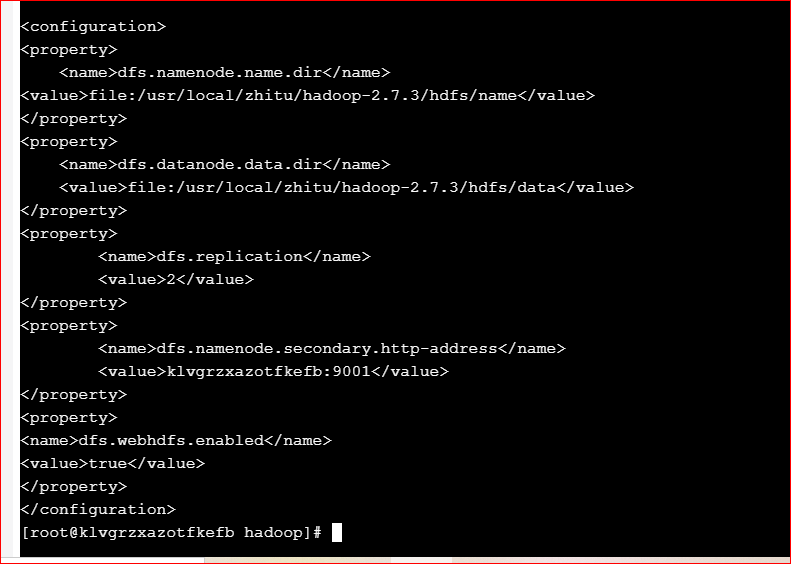
7





使用more命令查看文件内容：

10



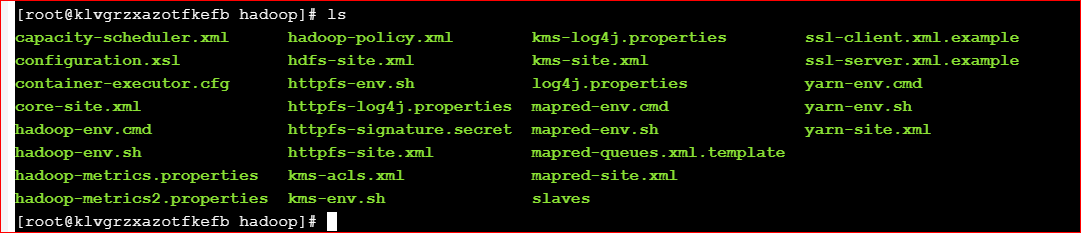
3.4、配置mapred-site.xml文件

进入hadoop-2.7.3目录下的etc/hadoop目录下，将文件mapred-site.xml.template使用mv命令重命名为mapred-site.xml，并修改该文件增加如下配置：

|  |
| --- |
| <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>mapreduce.jobhistory.address</name>  <value>server-1:10020</value>  </property>  <property>  <name>mapreduce.jobhistory.webapp.address</name>  <value>server-1:19888</value>  </property> |

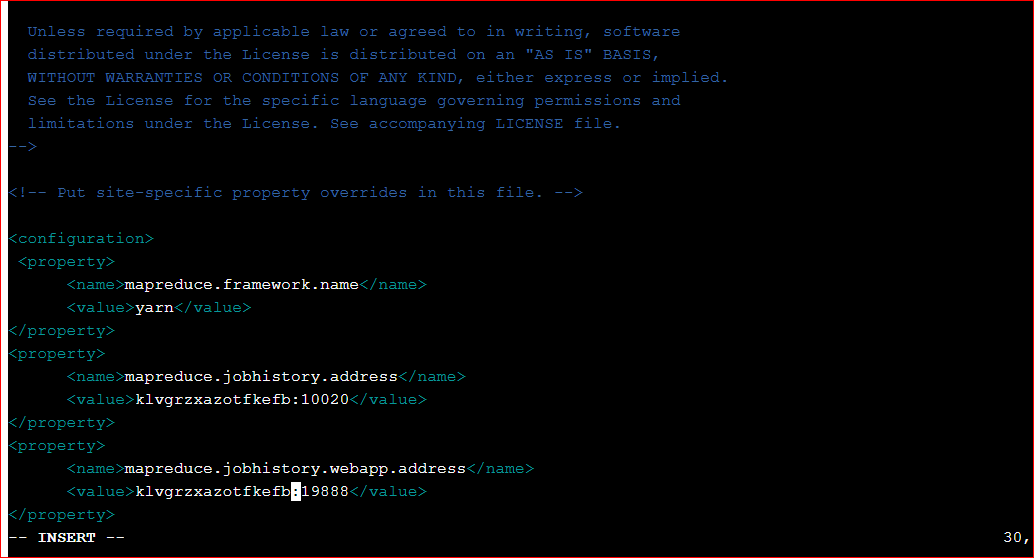
使用mv命令更改文件名，并查看：

12

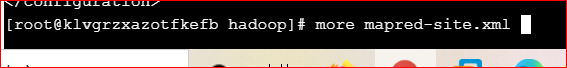


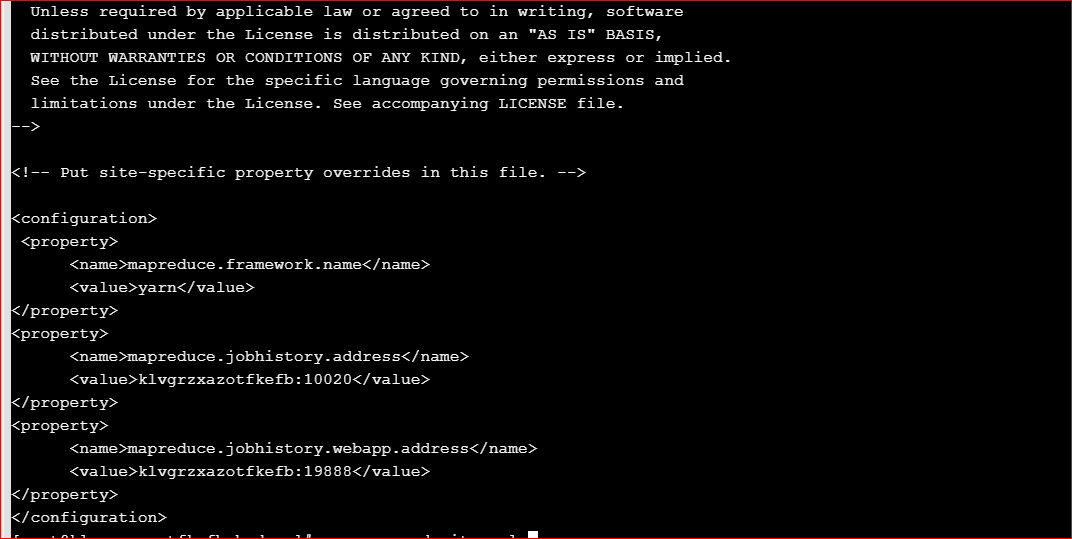
|  |
| --- |
| 使用vim进入mapred-site.xml文件增加配置： |

14



使用more命令查看该文件：





可以看到增加的配置已成功保存。

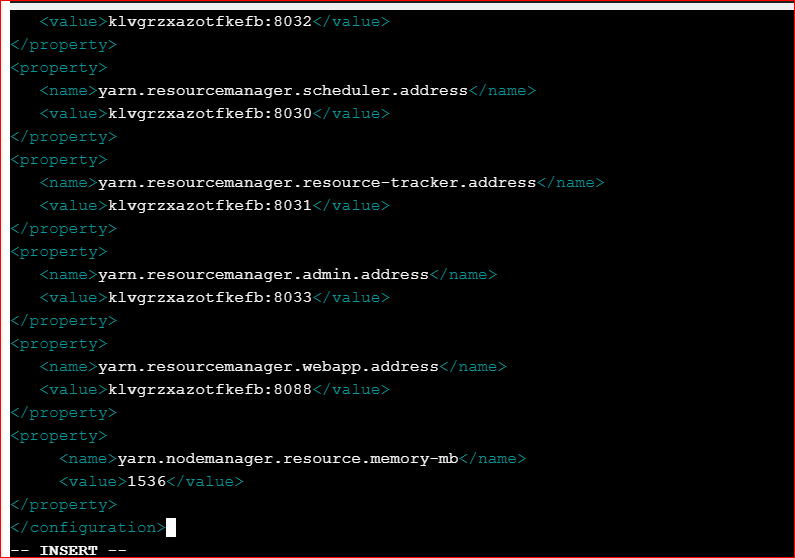
3.5、配置yarn-site.xml文件

进入hadoop-2.7.3目录下的etc/hadoop目录下，修改yarn-site.xml文件，增加如下配置：

|  |
| --- |
| <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>  <value>org.apache.hadoop.mapred.ShuffleHandler</value>  </property>  <property>  <name>yarn.resourcemanager.address</name>  <value>server-1:8032</value>  </property>  <property>  <name>yarn.resourcemanager.scheduler.address</name>  <value>server-1:8030</value>  </property>  <property>  <name>yarn.resourcemanager.resource-tracker.address</name>  <value>server-1:8031</value>  </property>  <property>  <name>yarn.resourcemanager.admin.address</name>  <value>server-1:8033</value>  </property>  <property>  <name>yarn.resourcemanager.webapp.address</name>  <value>server-1:8088</value>  </property>  <property>  <name>yarn.nodemanager.resource.memory-mb</name>  <value>1536</value>  </property> |

使用vim进入yarn-site.xml文件：

18

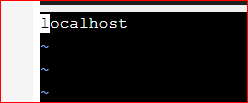


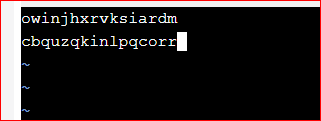
3.6、修改slaves，增加从服务器

进入hadoop-2.7.3目录下的etc/hadoop目录下，修改slaves文件，将其中的localhost去掉，并所有的从服务器名称添加进去，每台服务器名称占一行。

21

使用vim进入文件后出现如下显示，将localhost删掉更换成从服务器名称。





3.7、配置hadoop-env.sh、yarn-env.sh的JAVA\_HOME

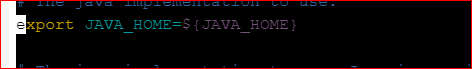
进入hadoop-2.7.3目录下的etc/hadoop目录下，修改hadoop-env.sh、yarn-env.sh 两个文件，增加JAVA\_HOME的环境变量配置。查看jdk目录可使用命令echo $JAVA\_HOME。

在hadoop-env.sh、yarn-env.sh两个文件中分别增加如下一段代码：

|  |
| --- |
| export JAVA\_HOME=/usr/lib/java/jdk1.8 |

使用vim进入hadoop-env.sh文件出现如图界面，找到​export JAVA\_HOME那一行并更改相关jd目录路径：

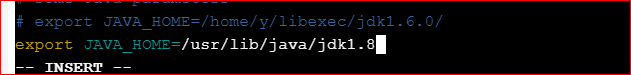
24



26

进入yarn-env.sh文件，找到​export JAVA\_HOME那一行并更改相关jd目录路径：

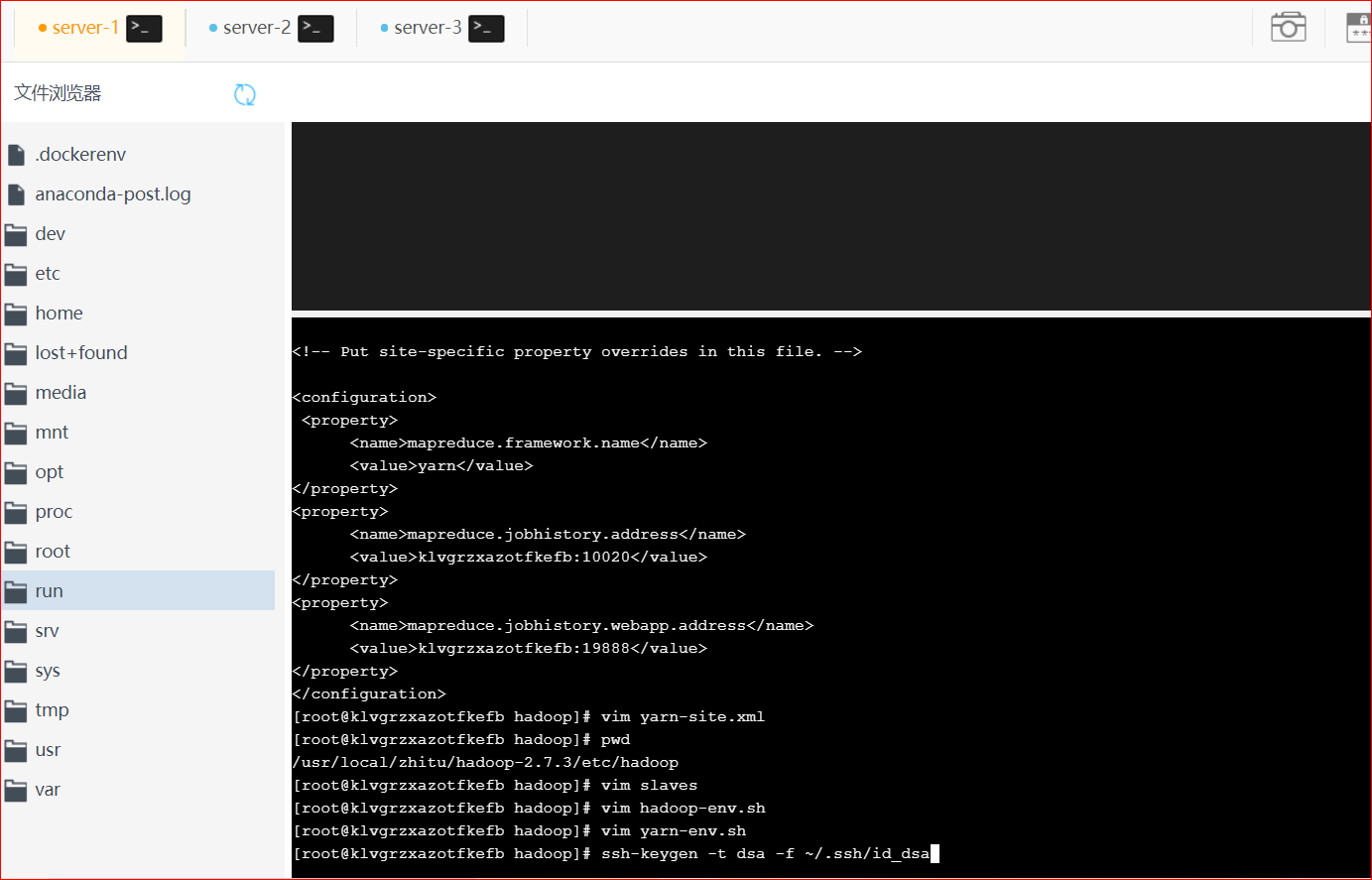
27

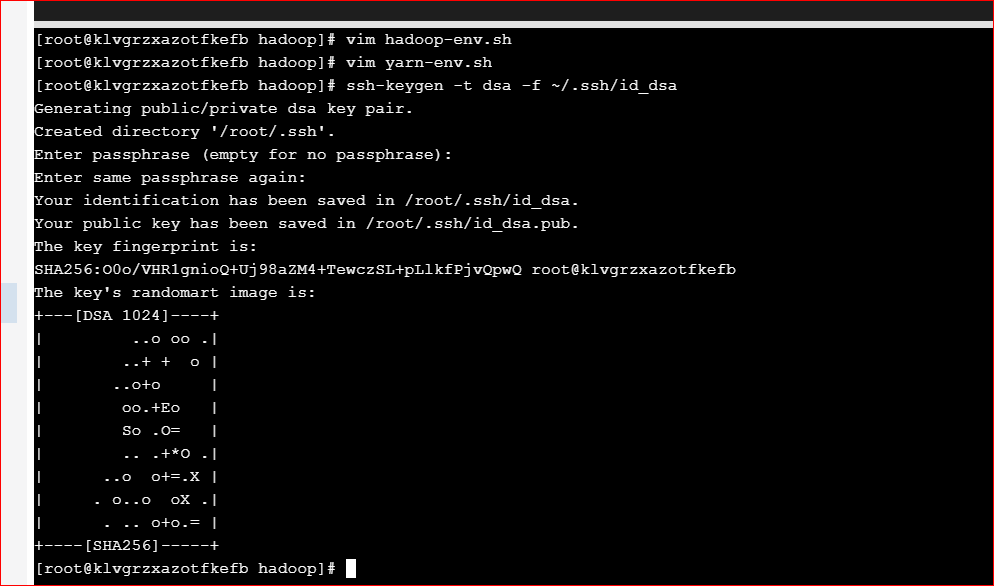


1. 配置主服务器免登陆访问从服务器

4.1、为每个节点分别产生公、私秘钥配置

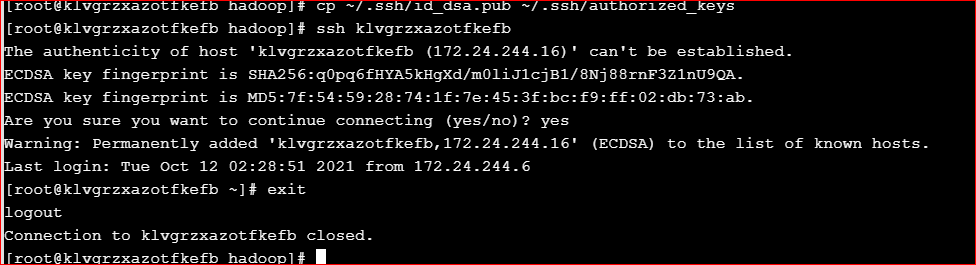
|  |
| --- |
| ssh-keygen -t dsa -f ~/.ssh/id\_dsa  将公钥复制成authorized\_keys文件。  cp ~/.ssh/id\_dsa.pub ~/.ssh/authorized\_keys |



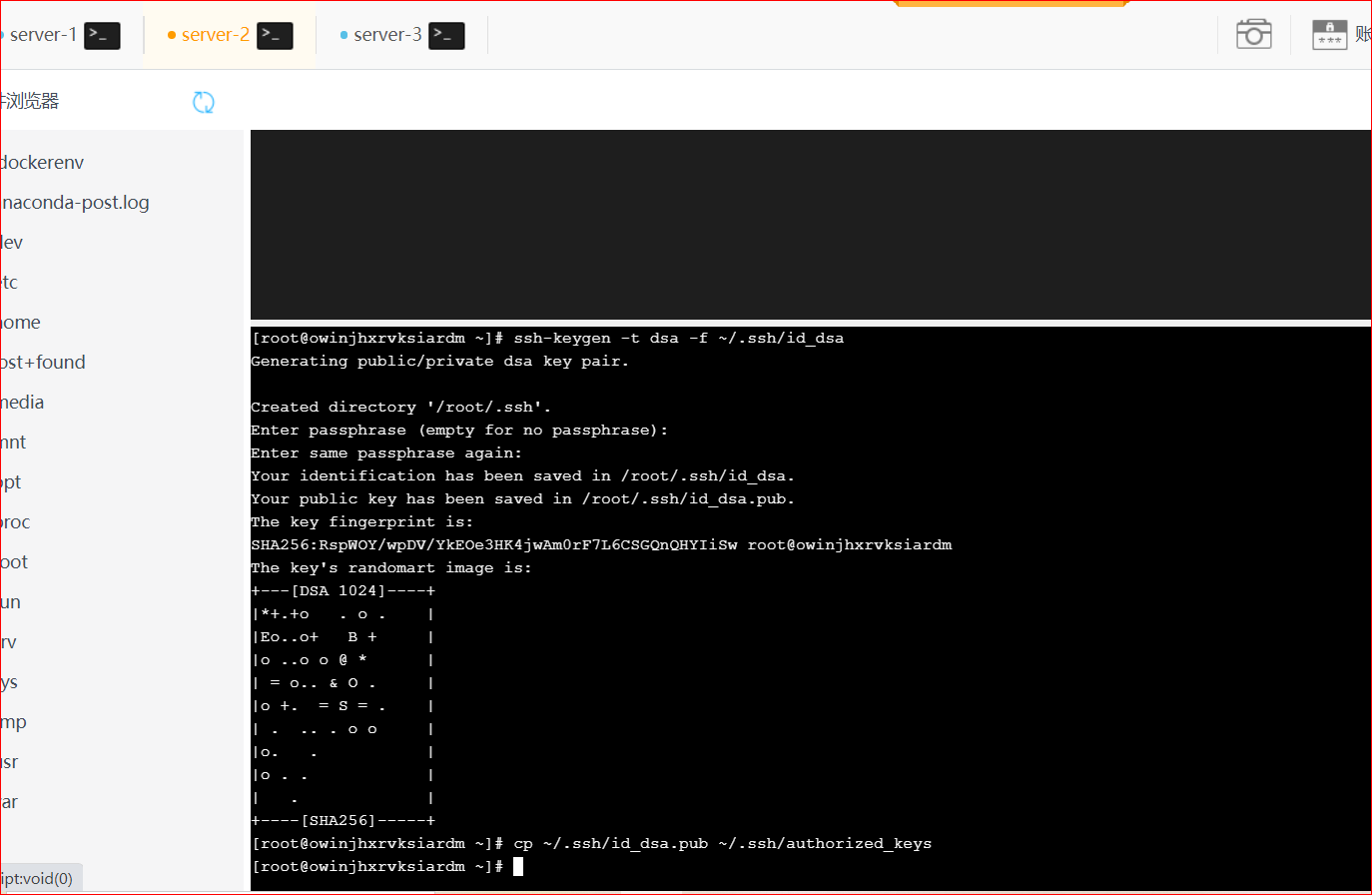


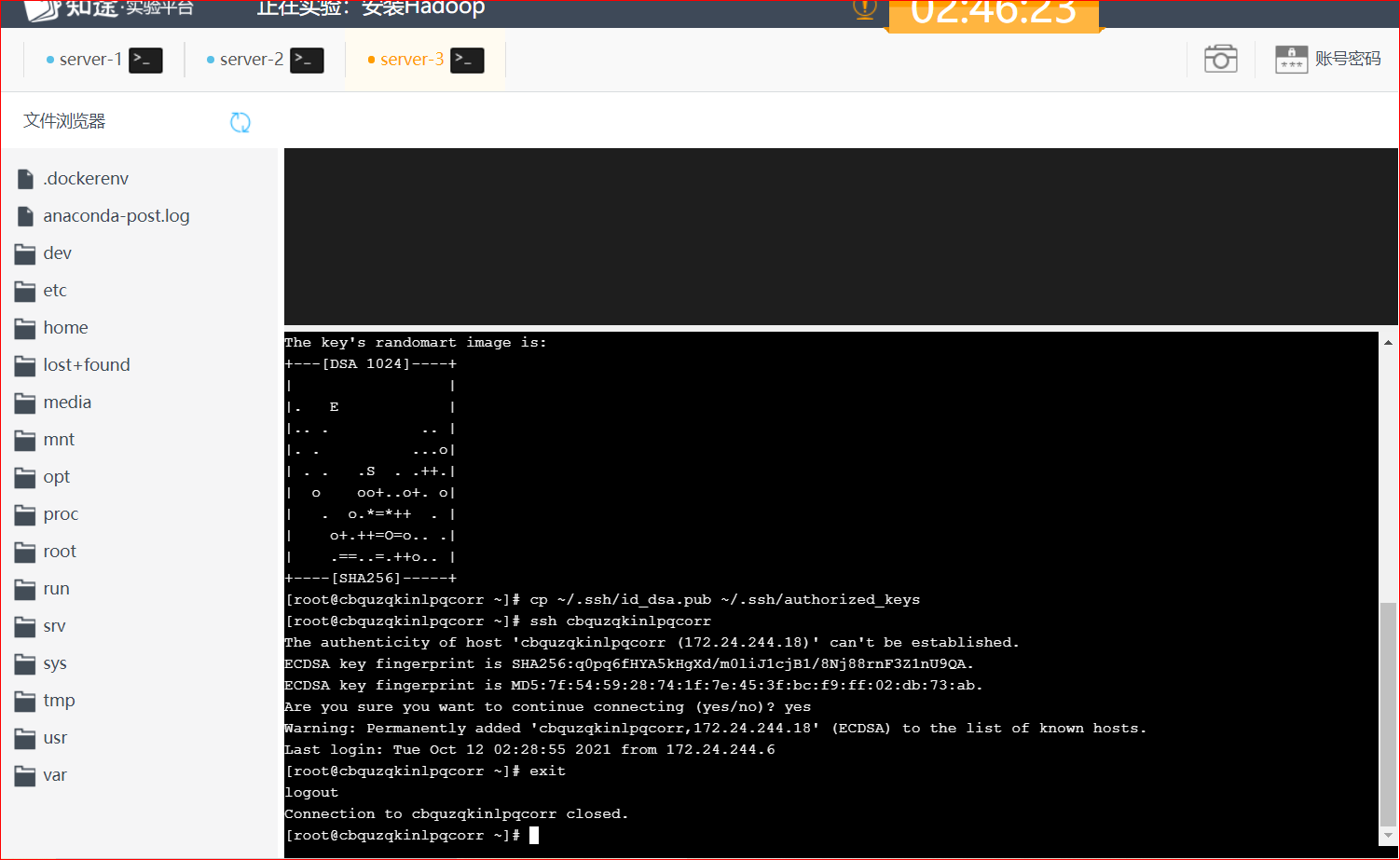
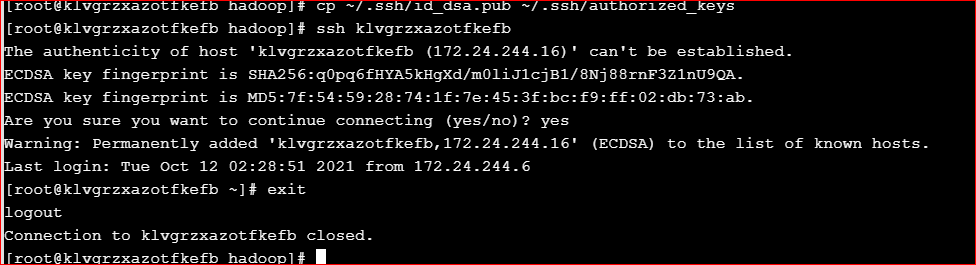
31

执行完上述操作之后，用ssh命令免密登陆本台服务器测试，如图（记得退出）：



接下来重复刚刚server-1的操作在server-2和server-3：

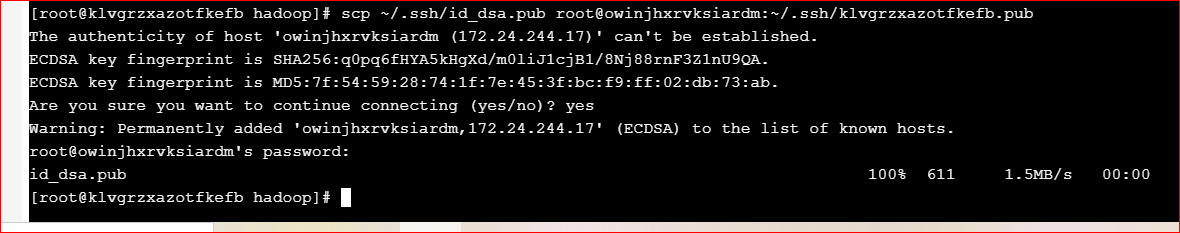
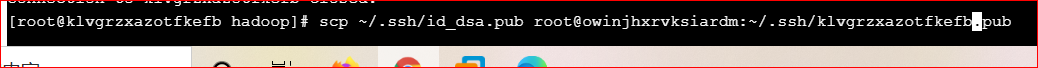




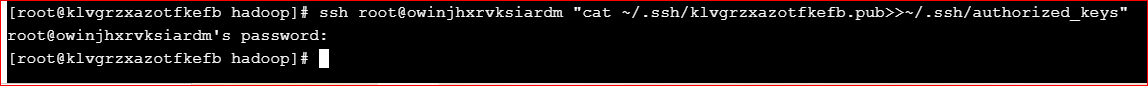
4.2、让主结点能通过SSH免密码登录两个子结点

将主服务器的公钥文件内容添加到从服务器的authorized\_keys文件里，就可以实现主服务器免密码登录从服务器了。

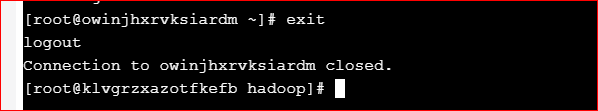
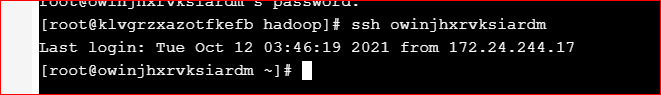
使用scp命令，将主服务器的公钥文件id\_dsa.pub复制到从服务器上，并命令为server-1（名称）.pub，随后根据提示输入从服务器登录密码。



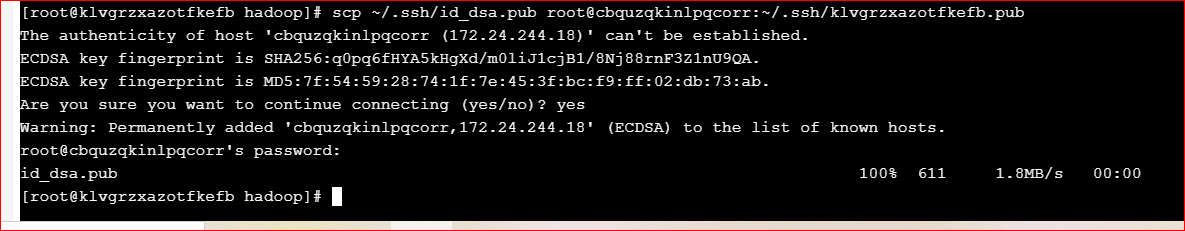
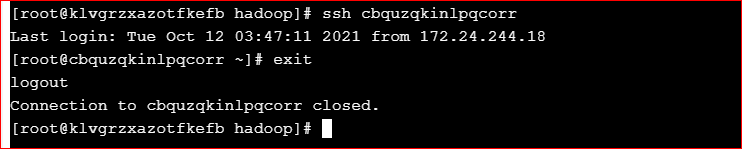
再将上一步生成的server-1.pub文件内容追加到server-2的authorized\_keys文件中，可以登录到server-2上去操作，也可以在server-1上使用ssh命令远程操作。这里我们使用ssh命令，在server-1上远程操作；并根据提示再次输入从服务器登录密码：



在主服务器上测试，是否可以免密登陆server-2的服务器（记得退出）：



接下来将上步操作从服务器server-2换成从服务器server-3的名称，并测试是否可以在主服务器上免密登陆从服务器server-3：

43

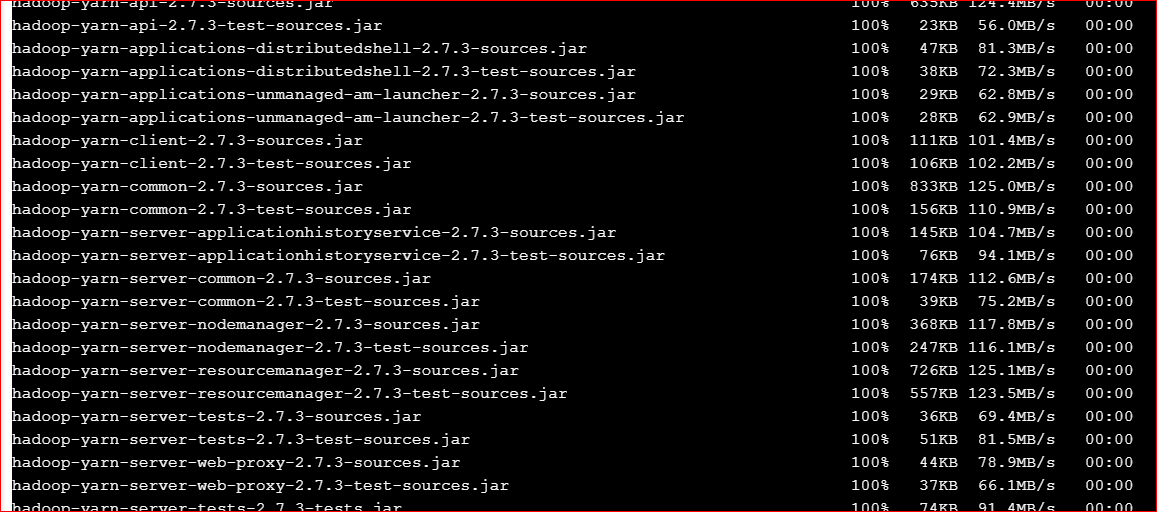
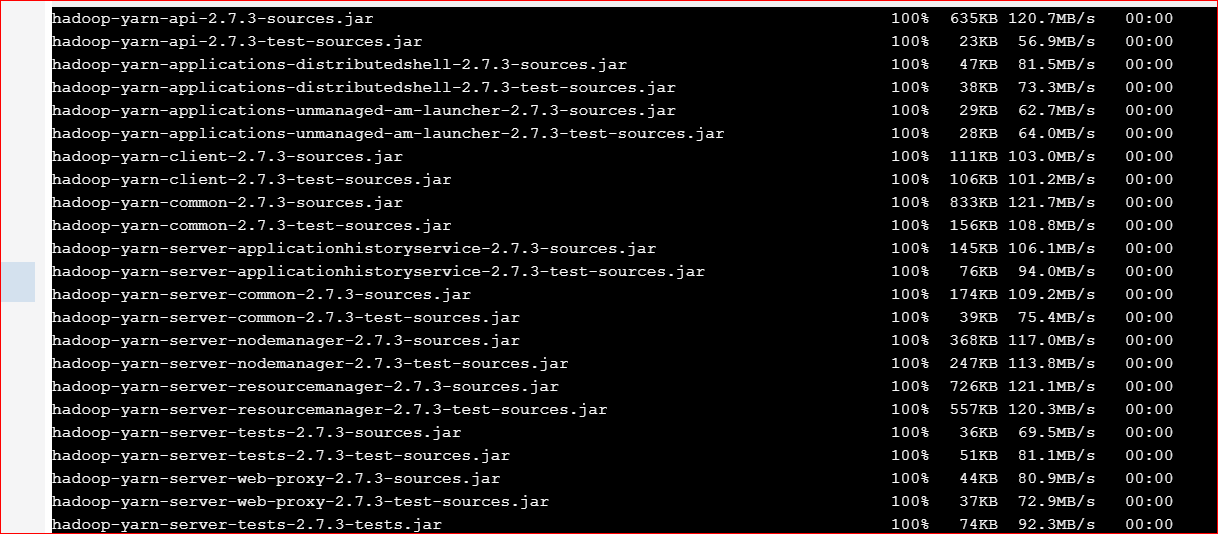
1. 将Hadoop复制到各从服务器

5.1、将Hadoop复制到各从服务器

通过以上步骤我们基本已经配置好了hadoop的集群配置，现在我们需要将我们配置好的文件复制到其他的服务器上，如图：

|  |
| --- |
| scp -r /usr/local/zhitu/hadoop-2.7.3 root@server-2:/usr/local/zhitu/ |
| scp -r /usr/local/zhitu/hadoop-2.7.3 root@server-3:/usr/local/zhitu/ |

45

47

1. 格式化namenode

6.1、格式化namenode

在主服务器上，进入hadoop-2.7.3目录下，格式化NameNode：

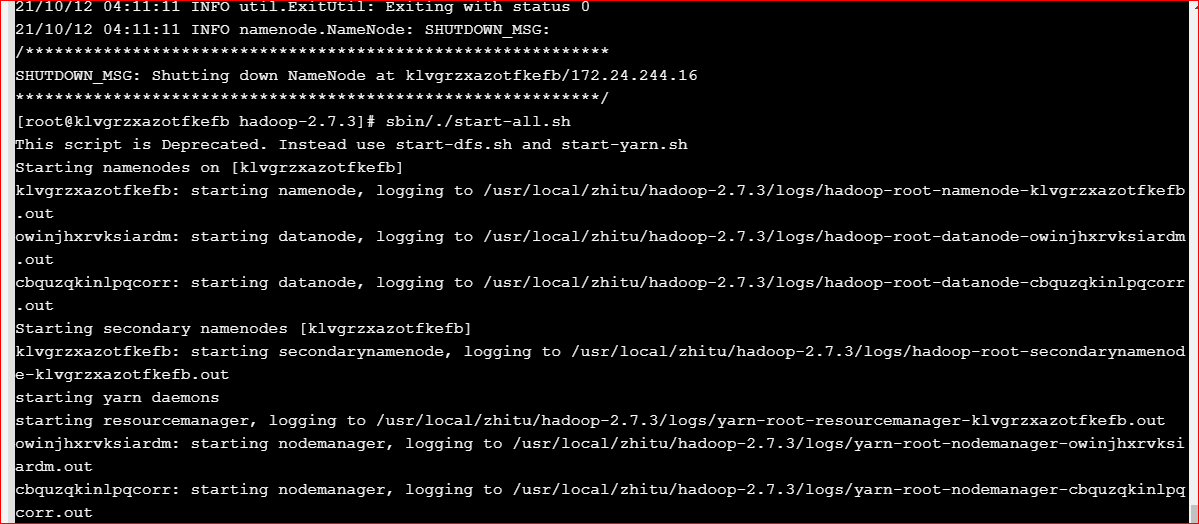
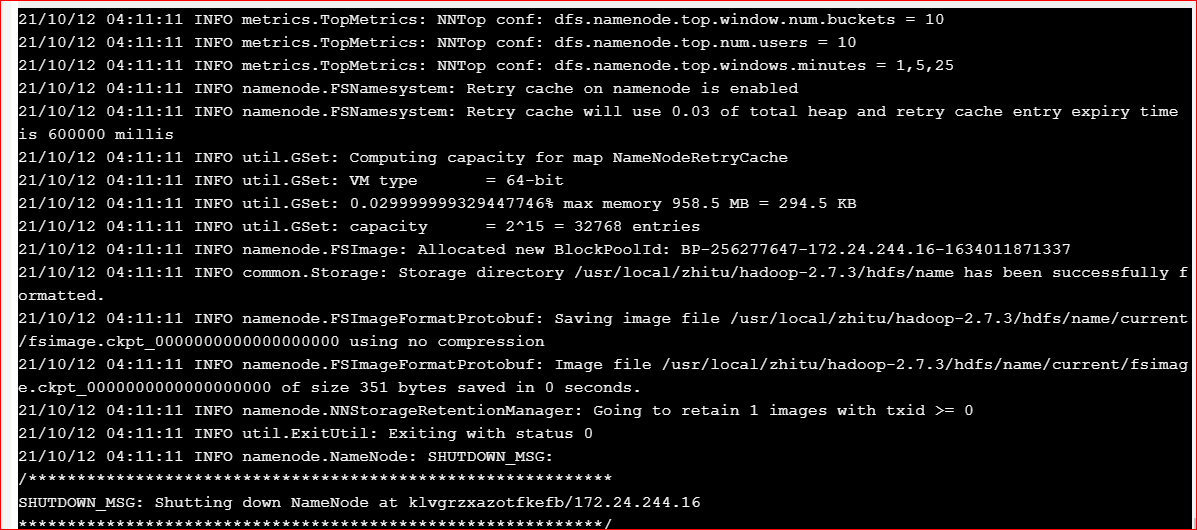
4950

1. 启动Hadoop

7.1、启动Hadoop

在主服务器上，进入hadoop-2.7.3目录下，启动Hadoop集群，结果显示如下：

52

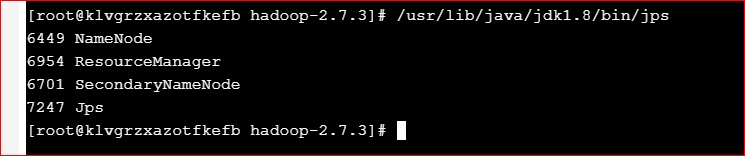


1. 验证服务是否安装成功

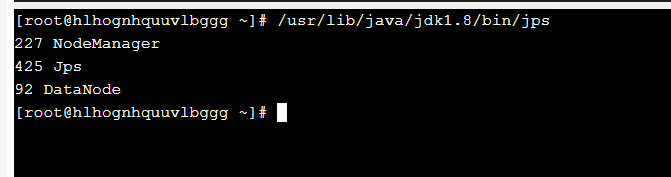
8.1、使用jps查看服务

jps命令为jdk提供的查看java进程的命令，hadoop服务如果启动成功了，可以在主服务器上查看到NameNode ,SecondaryNameNode,ResourceManager 三个服务，从服务器上看到到NodeManager, DataNode 两个服务。

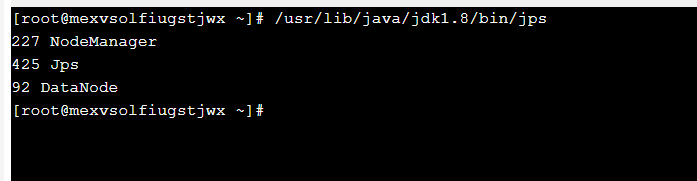
主服务器server-1显示结果如下：



从服务器server-2显示结果如下：



从服务器server-3显示结果如下：



**·实验心得**

通过这次试验学会了如何进行免密登陆的操作，如何测试免密登陆，如何更改文件名称，如何查看当前若在位置、查看更改的文件是否已经修改成功的，如何启动hadoop，还有是如何使用jps查看服务的。