环境

2020年11月16日

13:20

Anaconda安装：

下载对应版本的Anaconda

命令：bash A\*\*\*\*\*\*\*\*.sh

Yes 到底

最后 [source ~/.bashrc]

输入python查看结果：

计算机生成了可选文字:
Last login ： Mon NOV 16 10 ： 37 ： 19 2020 from 192 ． 168 ． 80 ． 
1 
16 ： 14 ： 56 ） 
more information ． 
[ root@Pspark 、 ] # python 
Python 3 ． 6 ． 5 IAnaconda, Inc. 丨 (default, 
LGCC 7 ． 2 ． 0 ] on linux 
Ty pe "help" 
" COPY right " 
" C redits " 0 r 
Apr 29 2018 ， 
''license" for 

Scala

计算机生成了可选文字:
#J AVA HOME 
expo rt J AVA HOME—/home/app/jdkI.8 
expo rt PATH—$PATH ： $JAVA HOME/bin 
#HADOOP 
expo rt HAD()0P_HOME—/home/app/hadoop2 ． 6 
expo rt PATH—$PATH ： $HADOOP HOME/bin ： $HADOOP HOME/sbin 
#SCALA 
expo rt SCALA HOME—/home/app/scala-2 ． 11 ． 12 
expo rt PATH—$PATH ： $SCALA HOME 
/bin 
"/etc/profile" 100L ， 2095C 

maven安装：

计算机生成了可选文字:
#J AVA HOME 
—/home/app/jdkl ． 8 
expo rt J AVA HOME— 
expo rt PATH—$PATH ： $ ] AVA HOME/bin 
#HADOOP 
expo rt 
expo rt 
#SCALA 
expo rt 
expo rt 
#MVN 
expo rt 
expo rt 
—/home/ app/hadoop2 ． 6 
HADOOP HOME— 
$HADOOP HOME/bin ： $HADOOP HOME/sbin 
PATH—$PATH 
—/home/app/scala-2 ． 11 ． 12 
SCALA HOME— 
$SCALA HOME/bin 
PATH—$PATH 
/home/app/maven3 ． 3 ． 9 
MAV E N HOME— 
$MAVEN HOME/bin 
PATH—$PATH 
"/etc/profile" 101L ， 2171C 

修改一下缓存路径：

计算机生成了可选文字:
丨 The path t 0 the 10Ca1 repository maven will I-ISe t 0 store artifacts ． 
丨 Default ： $fuser.homeH .m2/repository 
«IocaIRepositorp/path/t0/IocaI/repo«/IocaIRepositorp 
«IocaIRepositorp/home/mvn repository«/IocaIRepositorp 

Hadoop：2.6.1

Spark：2.4.6-hadoop2.6

Python 3.6.5 ==anaconda3.5

Java： 1.8

Scala： 2.11.12

Maven 3.3.9

ElasticSearch 6.3.0

Kibana：6.3.0

Cen7info： root root

网络连接以及Hadoop安装：

**分布式**

**换源：**

mv /etc/yum.repos.d/CentOS-Base.repo /etc/yum.repos.d/CentOS-Base.repo.backupcache

wget -O /etc/yum.repos.d/CentOS-Base.repo <http://mirrors.aliyun.com/repo/Centos-7.repo>

wget -O /etc/yum.repos.d/epel.repo <http://mirrors.aliyun.com/repo/epel-7.repo>

yum clean all

yum makecache

yum -y install autoconf automake libtool cmake ncurses-devel openss-devel lzo-devel zlib-devel gcc gcc-c++

yum -y install zlib-devel bzip2-devel openssl-devel ncurses-devel sqlite-devel readline-devel tk-devel gdbm-devel db4-devel libpcap-devel xz-devel

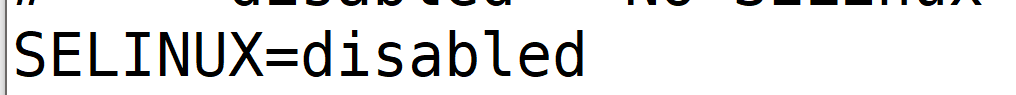
**1、关闭防火墙：**

$ systemctl stop firewalld.service

$ systemctl disable firewalld.service

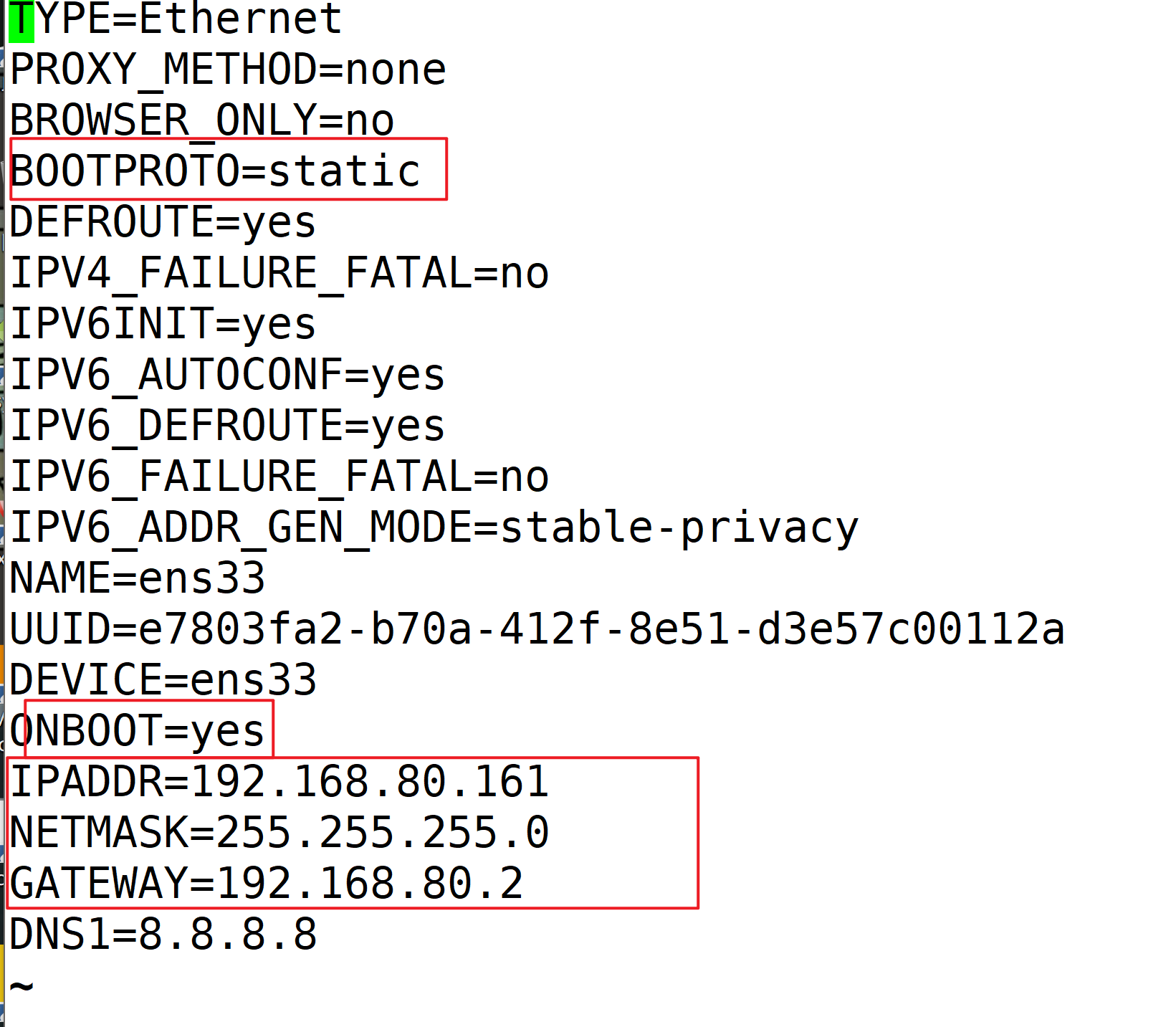
**2、禁用selinux**

$ vim /etc/sysconfig/selinux

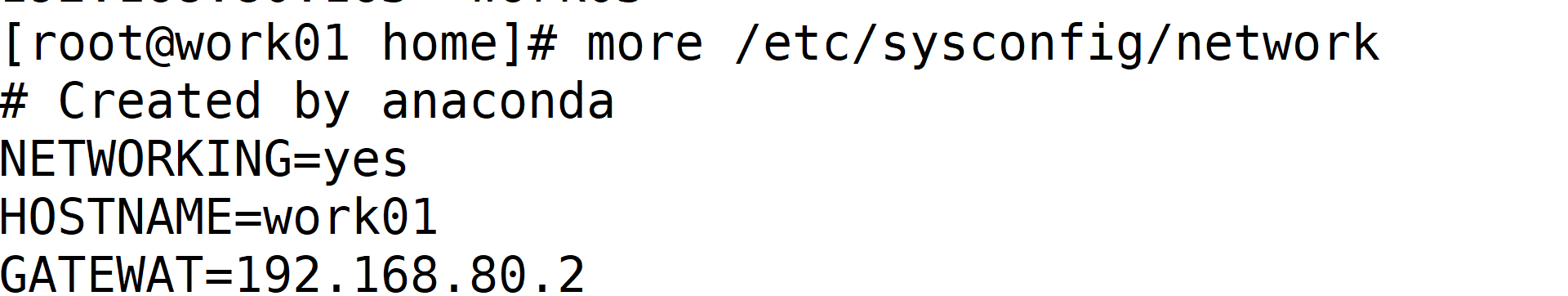


**3、网络相关配置，改成自己相对应的**

**$ vi /etc/sysconfig/network-scripts/ifcfg-ens33**

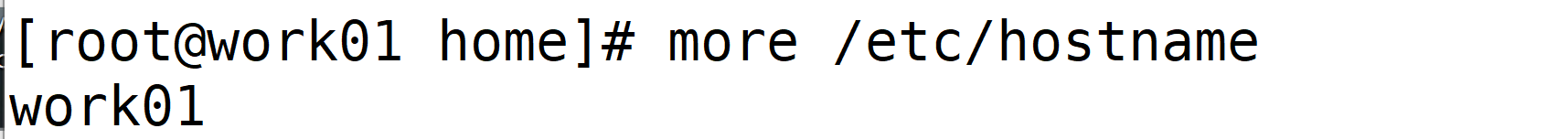


**$ vim /etc/sysconfig/network**

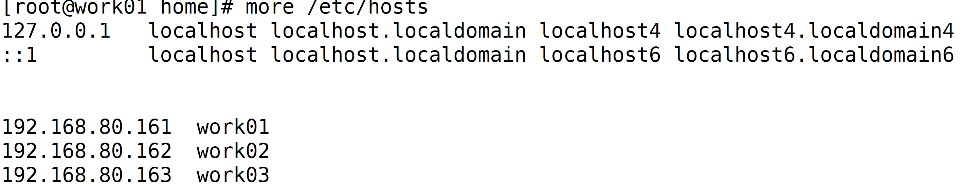


**$ vim /etc/hostname**

**重启生效**



**$ vim /etc/hosts**



**4、SSH免密登录**

$ ssh-keygen -t rsa //四下回车

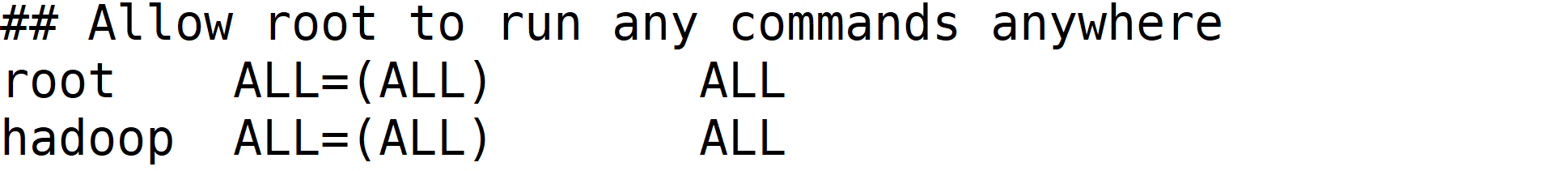
$ ssh-copy-id 主节点ip

$ spc -r /root/.ssh 其余节点ip:/root

**5、添加用户组**

$adduser hadoop

$vim /etc/sudoers



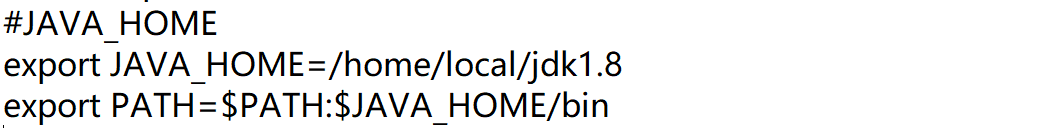
**修改用户所属用户组：**

$chown -R hadoop:hadoop /home/local/hadoop-2.9.2

**6、Jdk8安装**

$ tar -zxvf /home/tools/jdk-8u231-linux-x64.tar.gz -C /home/local

 $ vim /etc/profile



**7、 Hadoop需修改文件**

文件路径：【安装路径/etc/hadoop】

**core-site.xml**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://**主机名**:9000</value>

</property>

<property>

<name>io.file.buffer.size</name>

<value>131072</value>

</property>

<property>

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

<value>file:/home/local/hadoop-2.9.2/tmp</value>

</property>

<property>

<name>hadoop.proxyuser.root.hosts</name>

<value>\*</value>

</property>

<property>

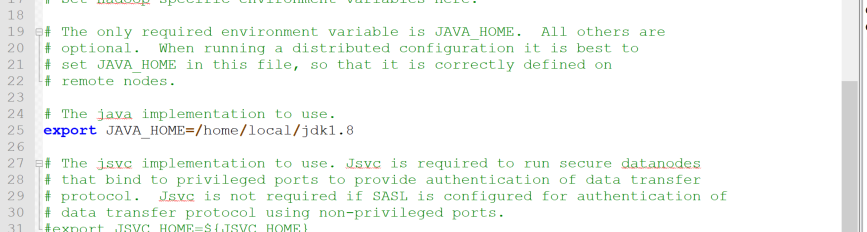
<name>hadoop.proxyuser.root.groups</name>

<value>\*</value>

</property>

</configuration>

**Hadoop-env.sh**【env文件都改为自己的路径】



**hdfs-site.xml** 主机名、文件存储位置

/dfs/name, /dfs/data 需自行创建

<configuration>

<property>

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

<value>**work01**:9001</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>**file:/home/app/hadoop2.6/dfs/name**</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>**file:/home/app/hadoop2.6/dfs/data**</value>

</property>

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<property>

<name>dfs.webhdfs.enabled</name>

<value>true</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

<property>

<name>dfs.web.ugi</name>

<value>supergroup</value>

</property>

</configuration>

**mapred-site.xml**

<configuration>

<property>

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

<value>yarn</value>

</property>

<property>

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

<value>**work01**:10020</value>

</property>

<property>

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

<value>**work01**:19888</value>

</property>

</configuration>

**Slaves 自己主机名**



**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.hostname</name>

<value>**work01**</value>

</property>

<property>

<name>yarn.resourcemanager.address</name>

<value>**work01**:8032</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>**work01**:8030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>**work01**:8031</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>**work01**:8033</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>**work01**:8088</value>

</property>

<property>

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

<value>2048</value>

</property>

<property>

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

<value>2048</value>

</property>

<property>

<name>yarn.nodemanager.vmem-pmem-ratio</name>

<value>2.1</value>

</property>

<property>

<name>mapred.child.java.opts</name>

<value>-Xmx2048m</value>

</property>

</configuration>

**格式化集群**

$ bin/hdfs namenode -format

**启动与关闭：**

安装目录下：

$ sbin/start-all.sh

$ sbin/stop-all.sh

**WEB管理界面**

Namenodeip\_:50070 //HDFS管理界面

resourmanger\_ip:8088 //yarn 管理界面

分布式：

一个namenode

2-3个datanode

伪分布式：