**Hadoop3.3.4伪分布式环境搭建**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

前言：想学习大数据开发，搭建大数据Hadoop环境是免不了的工作，这一步至关重要。而搭建Hadoop环境不是一件容易的事情，步骤很多，配置繁琐，这也是大数据门槛高的一个体现，很多新手都卡在这一关。经过实践和参考相关资料，个人总结整理成文档，为自己为他人，希望都有帮助。

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

现在有一台虚拟机：**hadoop01。**

操作系统：centos7.4（已经搭建好）

虚拟机：1台（hadoop01）

内存：2G

硬盘：20G

安装文件准备：

hadoop-3.3.4.tar.gz，

jdk-8u121-linux-x64.rpm

**一、前期环境配置**

切换到 root 用户

su root

**1、网络配置**

1.1修改主机名为：

主机名称需要改为：hadoop301

其中3是班级，01是你学号最后两位数。

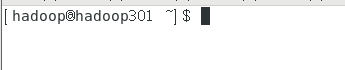
比如你的学号后面两位是23，主机名改为hadoop323

输入命令：

hostnamectl set-hostname hadoop301

然后输入命令重启

reboot

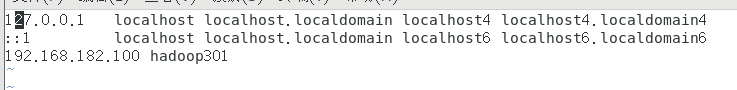


1.2 修改ip与主机名的对应关系

vi /etc/hosts

编辑输入：

192.168.182.100 hadoop301



**2、SSH免密码登录**

Hadoop301执行：

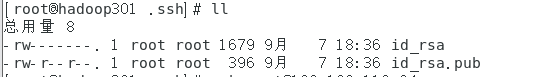
ssh-keygen -t rsa

连续按三个回车

cd /root/.ssh/

ll

查看生成的秘钥



然后将公钥上传到hadoop301服务器的

ssh-copy-id [root@192.168.182.100](mailto:root@192.168.116.94)

这个过程中提示需要输入密码时，输入root用户的密码，然后回车

免密登录验证

ssh root@192.168.182.100

如果没有提示要输入密码，能登录成功，则表示配置成功。

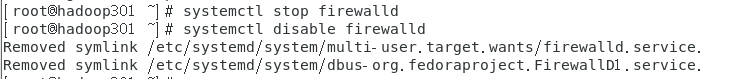
**3**、**关闭防火墙**

停止 firewalld 防火墙服务

禁用 firewalld 防火墙服务开机自启动

systemctl stop firewalld

systemctl disable firewalld



4、**关闭SELINUX**

永久关闭：

sed -i 's/enforcing/disabled/' /etc/selinux/config

临时关闭：

setenforce 0



**5、安装JDK**

###### 5.1通过XFTP工具上传文件jdk-8u121-linux-x64.rpm到下载目录



执行按照JDK命令：

rpm -ivh jdk-8u121-linux-x64.rpm

###### 5.2配置环境变量，修改/etc/profile：

vi /etc/profile

输入组合键shitf+g跳到文件末尾，

然后编辑输入：

export JAVA\_HOME=/usr/java/jdk1.8.0\_121

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

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

保存后退出

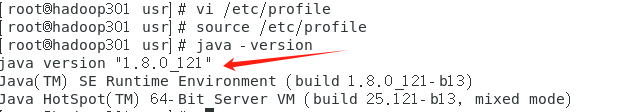
:wq

###### 5.3配置信息生效：

source /etc/profile

###### 5.4查看java版本：

java -version



/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**虚拟机做好快照**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**6、上传和解压Hadoop安装包**

###### 6.1 创建目录/opt/software

mkdir -p /opt/software

###### 6.1上传hadoop-3.3.4.tar.gz文件到下载目录



###### 6.2 创建目录：/opt/module

mkdir -p /opt/module

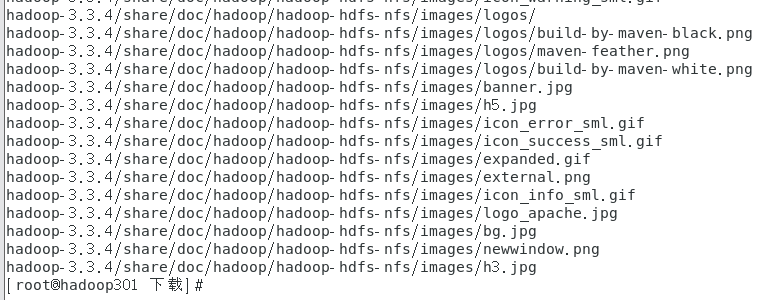
###### 6.3 进入到Hadoop安装包路径下：

cd /home/hadoop/下载

###### 6.4 解压安装文件到/opt/module下面：

tar -zxvf hadoop-3.3.4.tar.gz -C /opt/module/





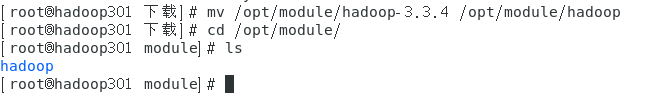
###### 6.5 查看是否解压成功：

ls /opt/module/hadoop-3.3.4



###### 6.6 重命名：

mv /opt/module/hadoop-3.3.4 /opt/module/hadoop



**7、配置环境变量**

###### 7.1 将Hadoop添加到环境变量

进入 hadoop目录

cd /opt/module/hadoop

获取Hadoop安装路径

pwd

路径：

/opt/module/hadoop

打开/etc/profile.d/my\_env.sh文件，/etc/profile.d这个目录用于存放系统全局的环境变量配置脚本。

vi /etc/profile.d/my\_env.sh

在profile文件末尾添加JDK路径：（shitf+g）

#HADOOP\_HOME

export HADOOP\_HOME=/opt/module/hadoop

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

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



保存后退出

:wq

###### 7.2 添加hadoop的环境变量和启动用户为root

执行命令：vi /etc/profile

export HADOOP\_HOME=/opt/module/hadoop

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

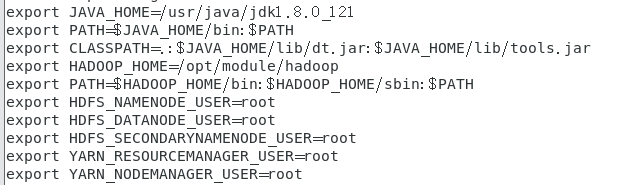
export HDFS\_NAMENODE\_USER=root

export HDFS\_DATANODE\_USER=root

export HDFS\_SECONDARYNAMENODE\_USER=root

export YARN\_RESOURCEMANAGER\_USER=root

export YARN\_NODEMANAGER\_USER=root



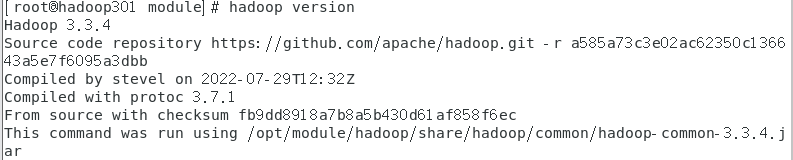
让环境变量配置生效

执行命令：source /etc/profile

###### 7.3 查看hadoop版本

执行命令：

hadoop version

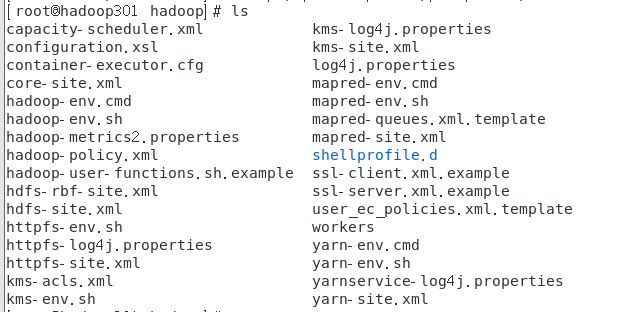


###### 7.4 查编辑Hadoop环境配置文件 - hadoop-env.sh

执行命令：

cd /opt/module/hadoop/etc/hadoop

进入[hadoop配置](https://so.csdn.net/so/search?q=hadoop%E9%85%8D%E7%BD%AE&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/lxyandqing/article/details/_blank)目录

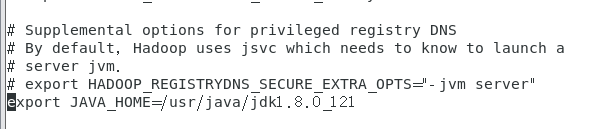


执行命令：

vi hadoop-env.sh

最后位置添加java的环境变量配置

export JAVA\_HOME=/usr/java/jdk1.8.0\_121



保存退出后，执行命令

source hadoop-env.sh

让配置生效

**8、编辑Hadoop核心配置文件**

###### 8.1编辑Hadoop核心配置文件 - core-site.xml

创建存放hadoop临时数据的目录：/opt/module/hadooptmp

mkdir -p /opt/module/hadooptmp

执行命令：vi core-site.xml

进入文件之后，滑到文件最下边，在<configuration></configuration>中添加如下内容

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop301:9000</value>

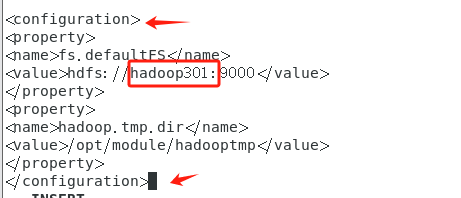
</property>

<property>

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

<value>/opt/module/hadooptmp</value>

</property>



添加好之后保存退出

注意：这里的hadoop301是我的主机名，对应改成自己的主机名

这里的/opt/module/hadooptmp是前一步创建hadooptmp的路径

###### 8.2编辑HDFS配置文件 - hdfs-site.xml

vi hdfs-site.xml

进入文件之后，滑到文件最下边，在<configuration></configuration>中添加如下内容

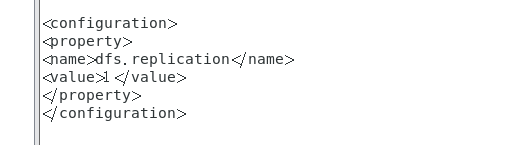
<property>

<name>dfs.replication</name>

<value>1</value>

</property>

添加好之后保存退出



###### 8.3配置MapReduce运行框架的yarn容器

vi mapred-site.xml

进入文件之后，滑到文件最下边，在<configuration></configuration>中添加如下内容

<property>

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

<value>yarn</value>

</property>

添加好之后保存退出



###### 8.4配置ResourceManager所在的主机及MapReduce的shuffer方式

vi yarn-site.xml

进入文件之后，滑到文件最下边，在<configuration></configuration>中添加如下内容

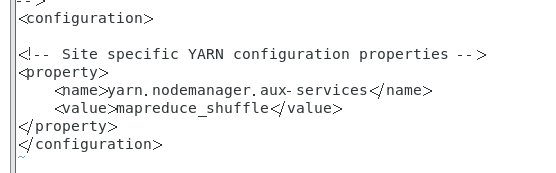
<property>

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

<value>mapreduce\_shuffle</value>

</property>

添加好之后保存退出



/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

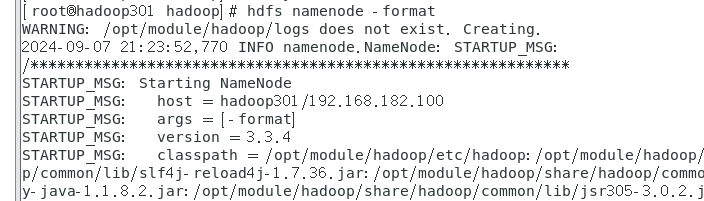
**虚拟机做好快照，后面如果有问题，恢复到这里，继续执行**

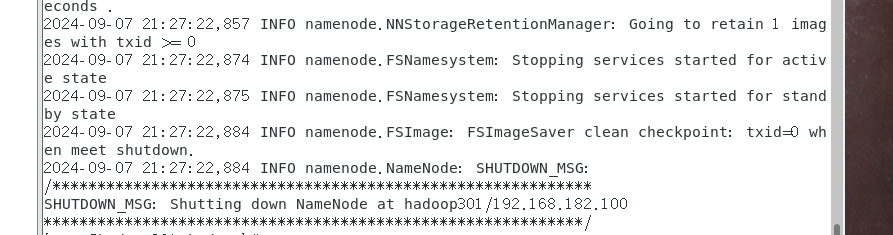
/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**9、初始化namenode主节点**

执行命令：

hdfs namenode -format

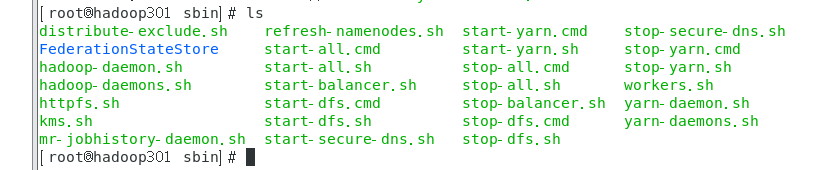




**10、启动Hadoop服务**

10.1进入hadoop的sbin目录：

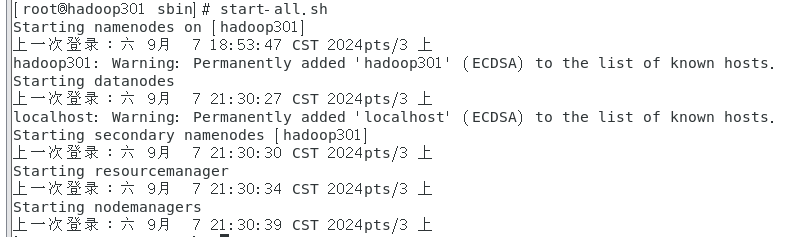
cd /opt/module/hadoop/sbin



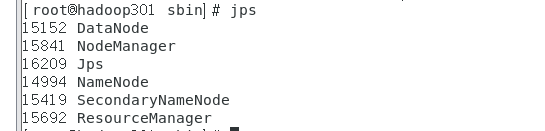
10.2执行启动启动所有服务的命令：

start-all.sh

或者直接执行：/opt/module/hadoop/sbin/start-all.sh

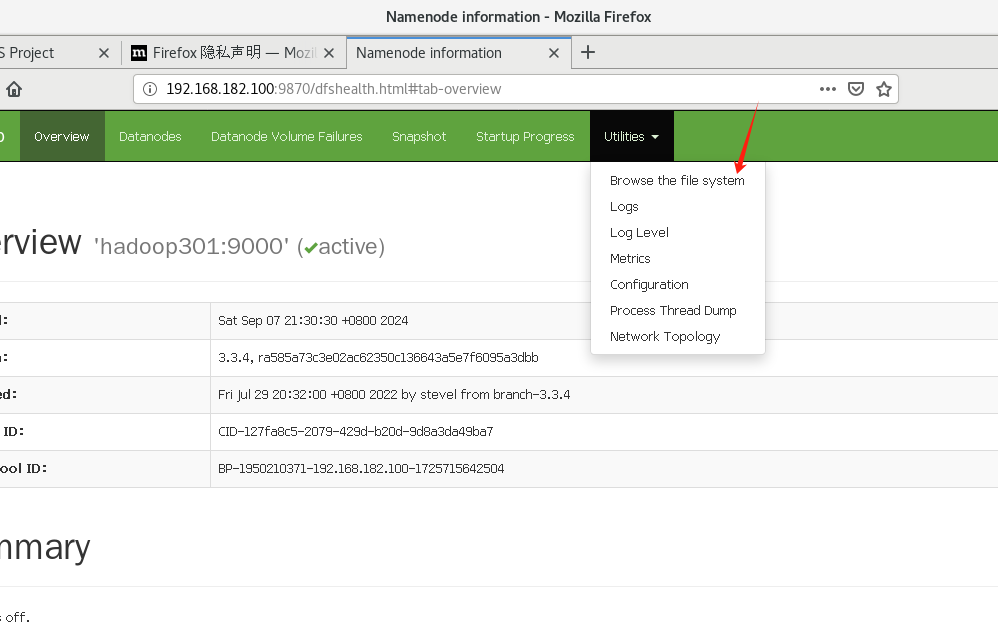


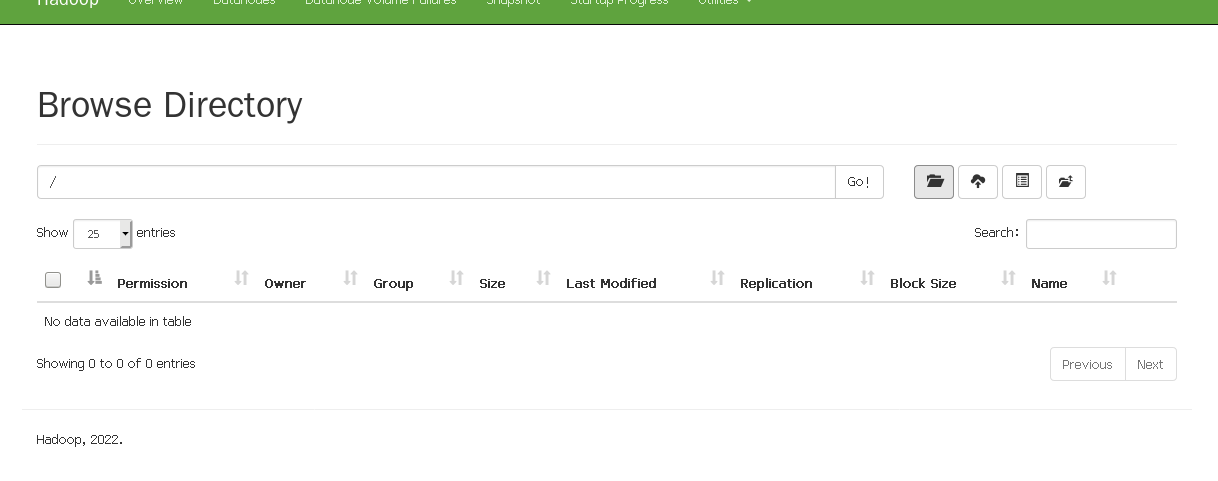
10.3 jps命令查看进程



10.4查看Hadoop WebUI

在浏览器里访问：<http://192.168.182.100:9870/>





**至此，Hadoop伪分布式环境已经搭建成功**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**虚拟机做好快照**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**个人心得：集群搭建好后，一定要做好快照备份。各个重要环节要做好快照，出现问题，就恢复到前面的快照，再操作。**