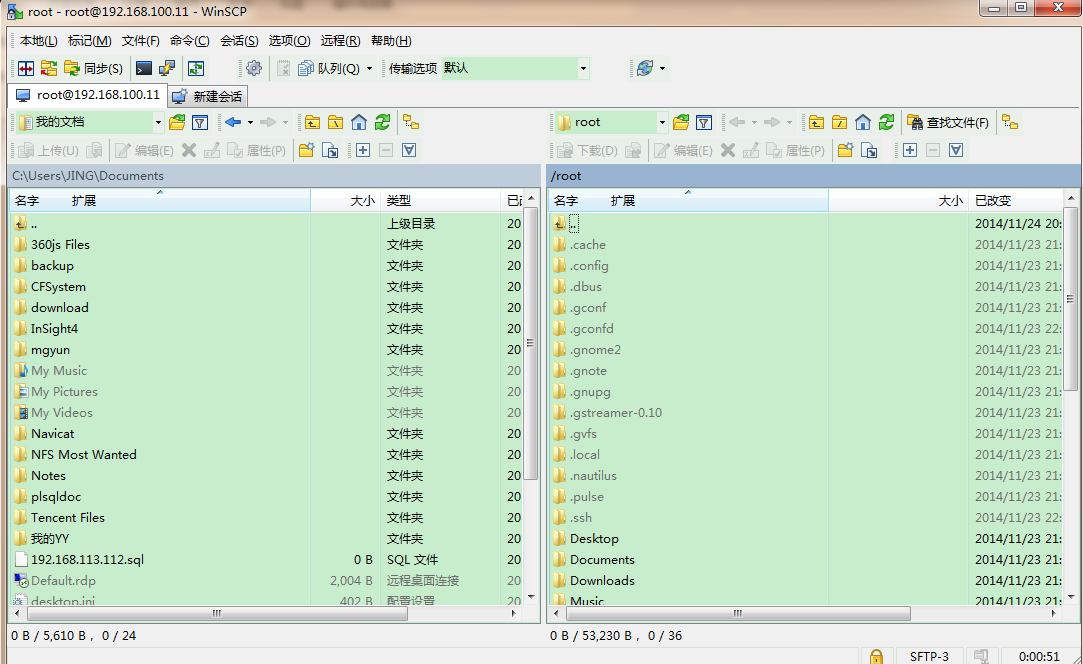
Hadoop学习第二天

继续伪分布安装

1. 安装JDK

使用winscp将hadoop和jdk复制到linux目录下

* 1. 打开Winscp输入虚拟机的ip，用户，密码，保存密码后点击登录



左边windows，右边linux。在左边找到要复制的文件(hadoop1.1.2.tar.gz,JDK-xx-xx.bin)，直接拖到linux系统下的/usr下的local中.

* 1. 解压hadoop-1.1.2.tar.gz

#tar –xzvf hadoop-1.1.2.tar.gz

* 1. 在linux中二进制文件一般是以bin结尾，但仍然需要执行权限才能执行

#chmod 755 jdk-6u24-linux-i586.bin

解压jdk

#./ jdk-6u24-linux-i586.bin

* 1. 配置环境变量才能使得jdk生效，

#vim /etc/profile

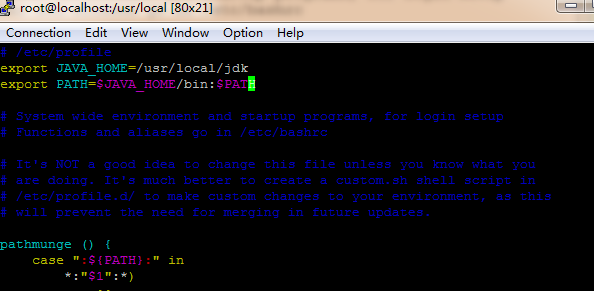
修改前先备份一份

#cp /etc/profile /etc/profile.bak

在开头的空白行添加

export JAVA\_HOME=/usr/local/jdk

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



就类似于windows下的jdk配置环境变 量

* 1. 重命名hadoop和jdk的文件夹，使得容易使用

#mv hadoop-1.1.2.tar.gz hadoop

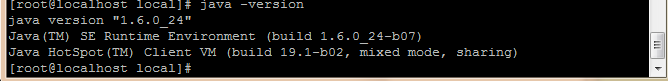
#mv jdk-6u24-linux-i586.bin jdk

* 1. 使得环境变量生效

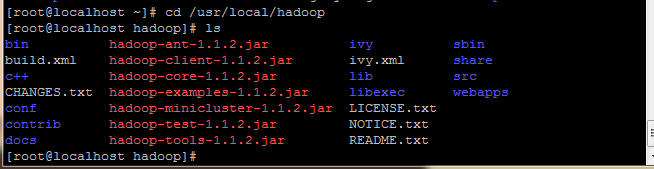
#source /etc/profile

* 1. 验证java版本

#java -version



1. 安装hadoop
2. 因为在第七步已经解压并且重命名了hadoop，直接切换到hadoop目录



bin下存放的是hadoop的可执行文件

conf下放的是配置文件

lib目录下放的是hadoop依赖的jar包

src目录是存放hadoop源代码的位置

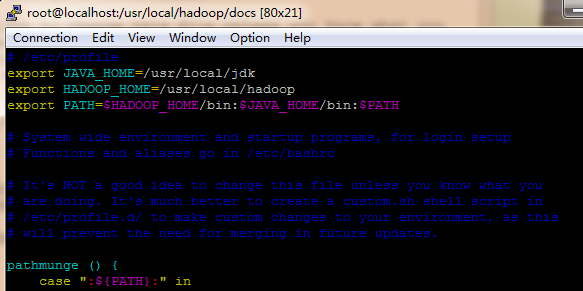
docs存放的是hadoop的文档

1. 为了执行hadoop命令方便，将hadoop的bin目录添加到环境变量文件中

#vim /etc/profile

export HADOOP\_HOME=/usr/local/hadoop

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



1. 使环境变量生效

#source /etc/profile

1. 修改hadoop的配置文件

进入conf目录

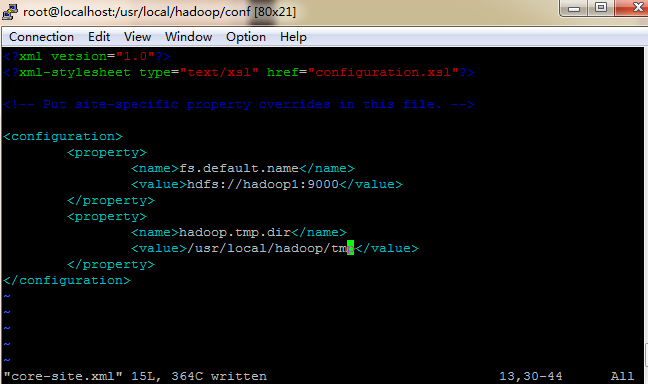
hadoop#cd conf

主要修改的配置文件有四个

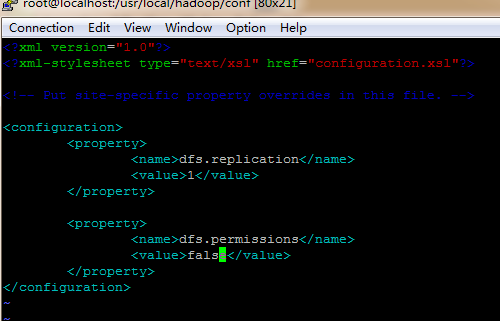
hadoop-env.sh \ core-site.xml\mapred-site.xml\hdfs-site.xml

hadoop-env.sh中将第九行的#去掉，改为 export JAVA\_HOME=/usr/loal/jdk/

修改core-site.xml配置文件内容如下，hadoop1是我设置的主机名

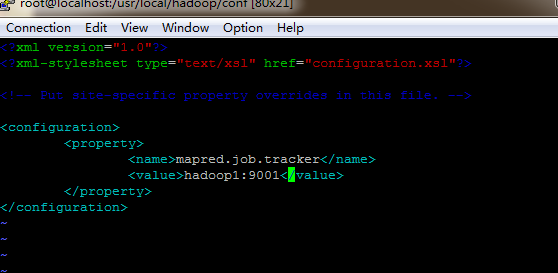
 修改hdfs-site.xml配置文件内容如下：

Replication,设置副本的数目，默认为三个



修改mapred-site.xml配置文件内容如下：

hadoop1修改为自己的主机名



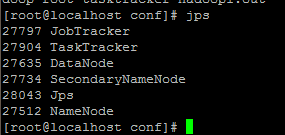
1. 启动hadoop前的文件系统格式化

# hadoop namenode -format

1. 启动hadoop

#start-all.sh 启动与hadoop有关的所有进程

1. 查看hadoop是否正常启动，使用jps命令可以查看java进程



如果看到这五个java进程代表hadoop启动成功

JobTracker、TaskTracker、DataNode、SecondaryNameNode、NameNode

启动时最有可能造成NameNode启动不成功的可能原因

(1). 没有格式化

(2). 环境变量设置错误

(3).ip和hostname绑定失败

h) 验证hadoop是否正在运行

浏览器地址栏中：<http://hadoop1:50070> (hadoop1需要在windows的host文件中绑定)显示正常网页表示NameNode正在运行

<http://hadoop1:50030> 的链接，显示正常表示mapreduce正在运行