## 安装ftp

## 安装ssh免密：

安装ssh：

sudo apt-get install openssh-server

配置免密：

exit # 退出刚才的 ssh localhost

cd ~/.ssh/ # 若没有该目录，请先执行一次ssh localhost

ssh-keygen -t rsa # 会有提示，都按回车就可以

cat ./id\_rsa.pub >> ./authorized\_keys # 加入授权

然后ssh localhost就不需要输入密码了

## 安装java：

1. 卸载版本java

Java -version检查java版本

1. 把java和hadoop传到服务器上
2. 新建app目录用于存放java和hadoop ：mkdir app
3. 将ftp传过来的java及hadoop安装包移动到app文件夹中

命令：

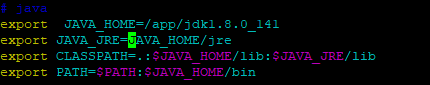
1. 解压java：
2. tar -zxvf jdk-8u141-linux-x64.gz -C app
3. 建立软连接： ln –s jdk1.8.0\_141 jdk
4. 修改java的环境变量
5. 再输入sudo vim /etc/profile
6. 在底端输入

export JAVA\_HOME=~/app/jdk1.8.0\_141

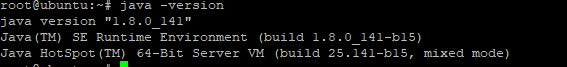
export JAVA\_JRE=JAVA\_HOME/jre

export CLASSPATH=.:$JAVA\_HOME/lib:$JAVA\_JRE/lib

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

1. 
2. 更新profile源
3. 保存退出，并使/etc/profile文件生效

source /etc/profile

1. 输入java -version可以查看java的版本，如果出现说明安装成功，否则是路径的问题。
2. 

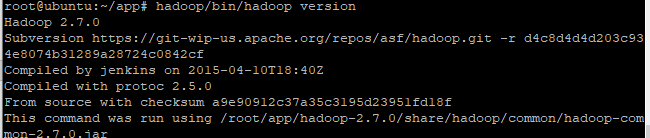
## 安装hadoop

打开app目录，解压hadoop：



建立软连接：ln -s hadoop-2.7.0 hadoop

验证单机下hadoop是否安装成功：hadoop/bin/hadoop version

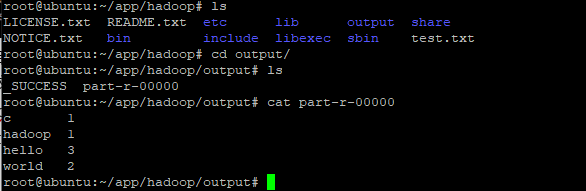


单机模式下运行自带的程序查看自己文件的wordcount

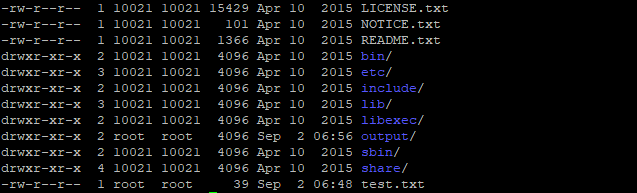
新建一个test.txt文件。在里面随便输入单词。

运行：bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.0.jar wordcount test.txt output

可以在output文件目录下查看



ll查看hadoop的目录结构



（1）bin目录：存放对Hadoop相关服务（HDFS,YARN）进行操作的脚本

（2）etc目录：Hadoop的配置文件目录，存放Hadoop的配置文件

（3）lib目录：存放Hadoop的本地库（对数据进行压缩解压缩功能）

（4）sbin目录：存放启动或停止Hadoop相关服务的脚本

（5）share目录：存放Hadoop的依赖jar包、文档、和官方案例

## 配置伪分布式

进入hadoop/etc/hadoop目录，修改相关配置文件

cd etc/hadoop

vim core-site.xml

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

<property>

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

<value>file:/fata/tmp</value>

</property>

<property>

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

<value>\*</value>

</property>

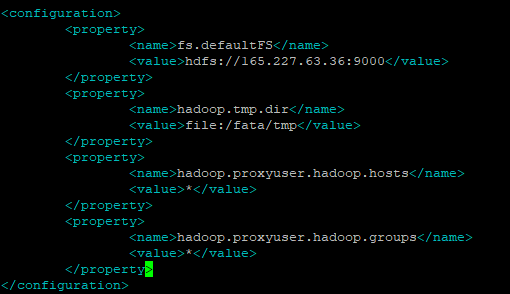
<property>

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

<value>\*</value>

</property>

</configuration>



Vim hdfs-site.xml配置文件

<configuration>

<property>

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

<value>/data/dfs/name</value>

<final>true</final>

</property>

<property>

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

<value>/data/dfs/data</value>

<final>true</final>

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

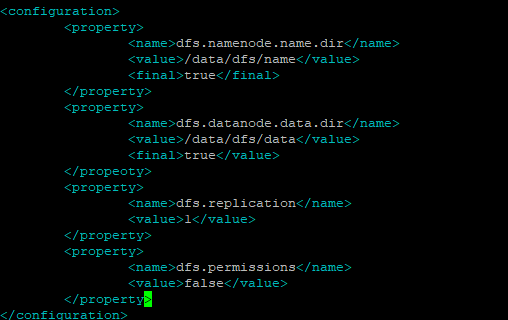
<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>



Vim mapred-site.xml.template配置文件

<configuration>

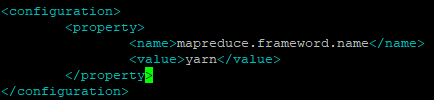
<property>

<name>mapreduce.frameword.name</name>

<value>yarn</value>

</property>

</configuration>



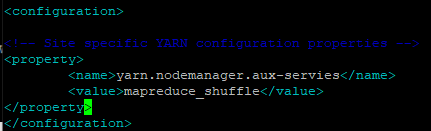
修改yarn-site.xml配置文件

<property>

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

<value>mapreduce\_shuffle</value>

</property>



Vim ~/.bashrc

在bashrc目录下加入环境变量

JAVA\_HOME=~/app/jdk

HADOOP\_HOME=~/app/hadoop

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

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

Source ~/.bashrc

创建hadoop相关数据目录

在hadoop相关配置文件中配置了多个数据目录，提前建立这些文件夹

mkdir -p data/tmp

mkdir -p data/dfs/name

mkdir -p data/dfs/data

给予所以用户data目录的权限：sudo chmod 777 -R /data

格式化namenode

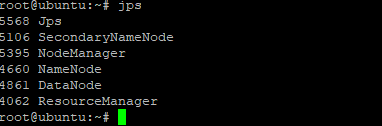
bin/hdfs namenode -format

启动hdfs：

启动hadoop伪分布式集群

sbin/start-all.sh

输入jps查看hadoop启动与否：



### 遇到的问题