## hadoop伪分布环境配置安装

##### 第一步：

首先我们先要有一个linux系统（root用户登入），或是虚拟机或是真正的系统。

##### 第二步、关闭linux防火墙并设置防火墙不自动检测开启：

linux防火墙命令：

*service iptables [start][stop][status]*：分别对应开启防火墙，关闭防火墙和查看防火墙状态。

*chkconfig iptables off*：关闭linux防火墙的自动检查开启功能。

*chkconfig --list*：显示所有的linux系统自动检查项

*chkconfig --list | grep iptables*：只是过滤防火墙的检查项

##### 第三步、设置主机名并设置与IP地址的对应：

查看主机名：*hostname*

临时会话修改主机名：*hostname crown*，将主机名修改为crown，但是会话结束主机名会变回原来的。

永久修改主机名：

用vi编辑命令：*vi /etc/sysconfig/network*，进入编辑界面，将里面的*HOSTNAME=...*改成自己的主机名保存退出即可。

设置主机名和IP地址对应：

*vi /etc/hosts*：编辑hosts，增加一项映射：*192.168.75.10 crown*。保存退出即可。

注意：linux的IP必须设置为静态IP地址。

##### 第四步、设置ssh自动登入：

设置ssh免密码登入：

先执行：*ssh-keygen -t rsa*，之后一直按回车，直到生成密码，密码文件生成后会默认保存在：

*/root/.ssh/id\_rsa.pub*文件中。

进入这个目录后复制文件为：authorized\_keys，既执行命令：*cp id\_rsa.pub authorized\_keys*

验证：执行*ssh localhost*命令，如果正常则说明设置成功。

**第五步，安装linux下的jdk：**

先将jdk文件放在*/usr/local*目录下，之后进入这个目录。

解压这个bin文件。执行命令：*./jdk-6u24-linux-i586.bin*，等待解压完成。（如不不行，用chmod命令看看有没有rwx的权限。）

修改环境变量：执行命令，*vi /etc/profile*，然后在空白处添加下面两行内容：

*export JAVA\_HOME=/usr/local/jdk(这里记得把这个路径下的相应文件重命名为jdk)*

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

最后执行，*source /etc/profile*，这样环境变量就配置好了，可以通过*java -version*验证。

注意：等号的两边不能有空格。

##### 第六步、安装hadoop：

将*hadoop-1.1.2.tar.gz*放在*/usr/local*目录下，之后执行命令解压它，命令如下：

*tar -zxvf hadoop-1.1.2.tar.gz*，得到*tar -zxvf hadoop-1.1.2*文件夹。

将其用mv命令重命名为*hadoop*。

在/etc/profile文件中再增加一项HADOOP\_HOME环境变量，如下：

*export JAVA\_HOME=/usr/local/jdk*

*export HADOOP\_HOME=/usr/local/hadoop*

*export PATH=.:$HADOOP\_HOME/bin:$JAVA\_HOME/bin:$PATH(这里记得修改)*

接着用命令*source /etc/profile*重新编译一下profile文件。

##### 第七步、修改hadoop的配置文件：

配置文件的目录为*/usr/local/hadoop/config/*

**1.hadoop-env.sh**

export JAVA\_HOME=/usr/local/jdk/

**2.core-site.xml**

<configuration>

<!-- 用来指定HDFS的NameNode的地址 -->

<property>

<name>fs.default.name</name>

<value>hdfs://hadoop0:9000</value>

</property>

<!-- 用来指定Hadoop运行时产生文件的存放目录-->

<property>

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

<value>/usr/local/hadoop/tmp</value>

</property>

</configuration>

这里将hdfs://hadoop0:9000改成hdfs://crown:9000，存放目录应该是具体目录。

**3.hdfs-site.xml**

<configuration>

<property>

<!-- 用来指定Hadoop文件副本的数量 -->

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>

**4.mapred-site.xml**

这里没有mapred-site.xml这个文件，但是只有它的模板mapred-site.xml.template，需要先创建。

<configuration>

<!-- 用来指定mp以后运行在yarn上-->

<property>

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

<value>yarn</value>

</property>

<!-- 这个现在不用配置，A版本用-->

<property>

<name>mapred.job.tracker</name>

<value>hadoop0:9001</value>

</property>

<!-- 这个现在不用配置，A版本用-->

</configuration>

这里将hdfs://hadoop0:9000改成hdfs://crown:9000

5、yarn-site.xml

<configuration>

<!-- NodeManager获取数据的方式为shuffle方式-->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<!-- 指定yarn的老大（resourcemanager）的地址-->

<property>

<name>yarn.resourcemanager.hostname</name>

<value>localhost或127.0.0.1或crown</value>

</property>

</configuration>

##### 第八步、对hadoop存储空间进行格式化：

执行命令：*hadoop namenode -format(过时了，A版本用)*

*hdfs namenode -format(现在用)*

**启动hadoop**

命令：*start-all.sh*

**验证hadoop**

用jps来验证hadoop进程

如果出现：NameNode，DataNode，JobTracker，TaskTracker，SecondaryNameNode这五个进程说明配置成功。（过时了*，A版本用*）

如果出现：NameNode，DataNode，ResourceManager，NodeManager，SecondaryNameNode这五个进程说明配置成功。

可以在浏览器中输入：

192.168.57.88:50070（hdfs的管理界面）

或是

192.168.57.88:50030

或是

192.168.57.88:8088（yarn的管理界面）

进行访问，访问成功，说明配置成功。

##### Hadoop版本信息：

A版本：hadoop-1.1.2.tar.gz

B版本：hadoop-2.6.0.tar.gz