**CentOS 6.5 配置hadoop 2.6.0伪分布式**

操作系统：CentOS 6.5    64位

Java环境：jdk 1.7.0\_71

hadoop下载地址： http://apache.fayea.com/hadoop/common/stable/hadoop-2.6.0.tar.gz

1.创建用户组和用户

使用su命令切换用户为root

groupadd hadoop

useradd hadoop hadoop

passwd  hadoop   #为用户添加密码   可以不设置密码

2.安装ssh

rpm -qa |grep ssh   #检查是否装了SSH包

yum install openssh-server   # 安装ssh

chkconfig --list sshd  #检查SSHD是否设置为开机启动

chkconfig --level 2345 sshd on  #如果没设置启动就设置下.

service sshd restart   #重新启动

3.配置ssh无密码登录

切换至hadoop用户

su hadoop

生成密钥

ssh-keygen -t rsa -P ""

执行后会在.ssh目录下生成id\_rsa和id\_rsa.pub两个文件

进入.ssh目录，并将id\_rsa.pub追加到authorized\_keys文件中

cd ./.ssh

cat id\_rsa.pub >> authorized\_keys

chmod 600 authorized\_keys # 修改用户权限

测试是否可以登录

ssh localhost  # 执行后会提示输入 yes or no. 输入yes后 如果提示为最后一次登录时间 则表明成功。

4.安装hadoop

将下载的hadoop解压并移动到期望的安装目录,修改其访问权限

tar -xvf hadoop-2.6.0.tar.gz

mv hadoop-2.6.0 /usr/opt/hadoop

chmod -R 775 /usr/opt/hadoop

chown -R hadoop:hadoop /usr/opt/hadoop

配置hadoop 的环境变量（所有的环境变量都是必须的）

su root

vim /etc/profile

然后加入以下内容

export HADOOP\_HOME=/usr/opt/hadoop

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

export HADOOP\_MAPRED\_HOME=${HADOOP\_HOME}

export HADOOP\_COMMON\_HOME=${HADOOP\_HOME}

export HADOOP\_HDFS\_HOME=${HADOOP\_HOME}

export YARN\_HOME=${HADOOP\_HOME}

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=${HADOOP\_HOME}/lib/natvie

export HADOOP\_OPTS="-Djava.library.path=${HADOOP\_HOME}/lib:${HADOOP\_HOME}/lib/native"

设置hadoop-env.sh中的java环境变量

cd /usr/opt/hadoop

vim ./etc/hadoop/hadoop-env.sh

加入以下内容

export JAVA\_HOME= {你的java环境变量}

5.配置伪分布式

hadoop的配置文件主要有core-site.xml  、 hdfs-site.xml 、 yarn-site.xml 三个文件。

cd /usr/opt/hadoop/etc/hadoop

core-site.xml

**<configuration>**

**<property>**

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

**<value>**/usr/opt/hadoop/tmp**</value>**  <!--一定要配置  系统默认的缓存文件夹在每次重启后会被删除-->

**</property>**

**<property>**

**<name>**fs.defaultFS**</name>**

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

**</property>**

**</configuration>**

hdfs-site.xml

**<configuration>**

**<property>**

**<name>**dfs.replication**</name>**

**<value>**1**</value>**

**</property>**

**<property>**

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

**<value>**file:/usr/opt/hadoop/dfs/name**</value>**

**</property>**

**<property>**

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

**<value>**file:/usr/opt/hadoop/dfs/data**</value>**

**</property>**

**<property>**  **<**!--这个属性是为了防止后面eclipse存在拒绝读写设置的 --**>**

**<name>**dfs.permissions**</name>**

**<value>**false**</value>**

**</property>**

**</configuration>**

yarn-site.xml

**<configuration>**

**<property>**

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

**<value>**yarn**</value>**

**</property>**

**<property>**

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

**<value>**mapreduce\_shuffle**</value>**

**</property>**

**</configuration>**

接下来创建yarn-site.xml中参数指定的文件夹

cd /usr/opt/hadoop

mkdir tmp dfs dfs/name dfs/data

到目前为止所有的配置都已经完成。

6.运行

首先格式化文件系统

./bin/hdfs namenode -format

启动

./sbin/start-dfs.sh

./sbin/start-yarn.sh

提示如下则表明成功了。

Starting namenodes on [localhost]

localhost: starting namenode, logging to /usr/opt/hadoop-2.6.0/logs/hadoop-hadoop-namenode-.out

localhost: starting datanode, logging to /usr/opt/hadoop-2.6.0/logs/hadoop-hadoop-datanode-.out

Starting secondary namenodes [0.0.0.0]

0.0.0.0: starting secondarynamenode, logging to /usr/opt/hadoop-2.6.0/logs/hadoop-hadoop-secondarynamenode-.out

输入网址127.0.0.1:50070就可以看见hadoop的网页了。

PS ：如果出现警告提示

Unable to load native-hadoop library for your platform

也可以看到那个网页，但hadoop并没有完全装好。

出现这个问题有两种情况：

一、没有设置HADOOP\_OPTS这个环境变量

二、是操作系统与hadoop.so文件不兼容导致的。

hadoop 2.5以后的release版本都是64位编译的，如果自己的电脑是32位的就出现不兼容的情况。这时需要重新编译hadoop源码（编译hadoop源码时，不要使用jdk1.8.0，使用1.6和1.7就可以了，不然会报这个错误[ERROR] Failed to execute goal org.apache.maven.plugins:maven-javadoc-plugin:2.8.1:jar (module-javadocs) on project hadoop-annotations: MavenReportException: Error while creating archive ）