

hadoop配置

2014年12月4日 21:19

Author: Yao Yu
E-mail: csyuyao@sjtu.edu.cn

以下为hadoop完全分布式的配置，实验已成功。

环境：VMware下3个CentOS7, hadoop2.6.0, jdk1.8.0_25

1. jdk安装：(all hosts needed)

- 删除当前jdk：rpm -qa|grep jdk （列出所有被安装的rpm package）
- rpm -e (--nodeps) {rpm package}
- vim /etc/profile 配置环境变量，追加
export JAVA_HOME={/usr/java/jdk1.7.0} #jdk所在路径
export CLASSPATH=.:\$JAVA_HOME/jre/lib/rt.jar:\$JAVA_HOME/lib/dt.jar:
\$JAVA_HOME/lib/tools.jar
export PATH=\$PATH:\$JAVA_HOME/bin
- source /etc/profile 使生效

2. 创建hadoop用户：(all hosts needed)

- ① useradd hadoop 创建hadoop用户 ② passwd hadoop 为hadoop用户设置密码
之后的操作没有权限（root）要求，一律在hadoop用户下。
- vim /etc/hostname 设置主机名字（如master，slave1等，便于辨识）
- vim /etc/hosts 设置ip与主机名的映射（每台主机都要统一），如：

192.168.182.135	master
192.168.182.134	slave1
192.168.182.131	slave2

3. ssh无密码配置：（在master上配置，scp到slaves上）

- 在master上，ssh-keygen -t rsa 生成密钥对 id_rsa 和 id_rsa.pub，默认存储在 /home/hadoop/.ssh
- cd .ssh, cat id_rsa.pub >> authorized_keys
- 修改文件权限 chmod 600 authorized_keys
- 把公钥发布到各节点上 scp authorized_keys hadoop@slave1:~/.ssh
- 更改权限 chmod 700 .ssh, chmod 600 .ssh/authorized_keys
远程ssh登陆貌似要所有机器拥有统一一个账户名，如hadoop，所以全部用hadoop账户
- vim /etc/ssh/sshd_config 设置无密码登陆有效（这一步基本不需要）

4. hadoop配置：（在master上配置，scp到slaves上）

- ① tar -zxf hadoop-2.X.X.tar.gz -C /usr/local/ 解压 ② cd /usr/local/ ③ ln -s hadoop2.X.X
hadoop软连接方便操作
- vim etc/profile 配置环境变量，追加：
export HADOOP_PREFIX="/usr/local/hadoop"
export PATH=\$PATH:\$HADOOP_PREFIX/bin:\$HADOOP_PREFIX/sbin
export HADOOP_COMMON_HOME=\${HADOOP_PREFIX}
export HADOOP_HDFS_HOME=\${HADOOP_PREFIX}

```

export HADOOP_MAPRED_HOME=${HADOOP_PREFIX}
export HADOOP_YARN_HOME=${HADOOP_PREFIX}
c. ① vim etc/hadoop/yarn-env.sh 导入JAVA_HOME , export JAVA_HOME={/usr/java/jdk1.7.0}
   ② vim etc/hadoop/hadoop-env.sh 导入JAVA_HOME , 同上 ;
d. 修改xml文件 :
   ① core-site.xml
   <configuration>
       <property>
           <name>fs.defaultFS</name>
           <value>hdfs://master:9000</value>
           <description>The name of the default file system.</description>
       </property>
   </configuration>
   ---
   ② hdfs-site.xml
   <configuration>
       <property>
           <name>dfs.replication</name>
           <value>3</value>
       </property>
       <property>
           <name>dfs.namenode.name.dir</name>
           <value>file:///usr/local/hadoop/dfs/name</value>
           <final>true</final>
       </property>
       <property>
           <name>dfs.datanode.data.dir</name>
           <value>file:///usr/local/hadoop/dfs/data</value>
       </property>
   </configuration>
   ---
   ③ yarn-site.xml
   <configuration>
   <!-- Site specific YARN configuration properties -->
       <property>
           <name>yarn.nodemanager.aux-services</name>
           <value>mapreduced_shuffle</value>
       </property>
       <property>
           <name>yarn.resourcemanager.hostname</name>
           <value>master</value>
       </property>
   </configuration>
   ---
   ④ mapred-site.xml
   默认没有mapred-site.xml , cp mapred-site.xml.template mapred-site.xml
   <configuration>
       <property>
           <name>mapreduce.framework.name</name>
           <value>yarn</value>
           <final>true</final>
       </property>
   </configuration>

```

- e. vim etc/hadoop/slaves ,如 :
slave1
slave2
- f. 至此 , master上的配置完成 , scp到slaves上 :
scp -r /usr/local/hadoop hadoop@slave1:/usr/local/hadoop
scp -r /usr/local/hadoop hadoop@slave2:/usr/local/hadoop
- g. 在每台主机上都执行 : chown hadoop /usr/local/hadoop 避免权限问题 , 见下
- 5. hadoop启动 :
 - a. 在/usr/local/hadoop/sbin下 , 初始化 : hdfs namenode -format
 - b. start-all.sh 全部启动 (或分为两步 : ①start-yarn.sh ②start-dfs.sh)
 - c. 在浏览器下 , master:50070 查看Namenode Information
 - d. stop-all.sh 全部关闭
- 6. hadoop添加datanode :
 - a. 新增节点设置ip和hostname , 如192.168.182.132 slave3
 - b. 安装jdk , 配置ssh无密码
 - c. 在所有的NameNode上和DataNode上修改/etc/hosts 追加 : 192.168.182.132 slave3
 - d. 修改master上的hadoop/etc/hadoop/slaves 追加 : slave3 , 并且原先的DataNode需要同步
 - e. scp hadoop到slave3
 - f. 启动DataNode : 在hadoop用户下 , 在usr/local/hadoop下
 - ① sbin/hadoop-daemon.sh start datanode
 - ② sbin/hadoop-daemon.sh start tasktracker
 - g. 数据同步 : sbin/start-balancer.sh -threshold 3 (本语句还未实验)
- 7. 设置静态ip (有问题)
 - a. vim /etc/sysconfig/network-scripts/ifcfg-eno16777736 修改
BOOTPROTO="static"
ONBOOT="yes"
IPADDR="192.168.1.7"
PREFIX="255.255.255.0"
GATEWAY="192.168.1.1"
在虚拟机下 , NAT网络时 , ip、gateway要与vmnet8在同一网段
 - b. sbin/ifdown eno16777736
 - c. sbin/ifup eno16777736
 - d. ifconfig 检查是否成功
- 8. 问题 :
 - a. mkdir: cannot create directory `/usr/local/hadoop/bin/./logs': Permission denied
解决方案 : chown -hR hadoop /usr/local/hadoop
chown -hR 对软连接可行 , 但对其子目录不可行 ,
执行 chown -hR hadoop /usr/local/hadoop2.X.X
 - b. 打开 master:50070 livenode数量有误 (为0) , 一般是由于namenode和datanode中的
clusterID号不同 , 重复格式化造成。
解决方案 : /hadoop/dfs/name/current/VERSION下 , 把namenode的clusterID复制到每个
datanode的/hadoop/dfs/data/current/VERSION中 , 使保持一致。