五月项目进展:完成hadoop的配置

Hadoop: 完成 hadoop的配置:

参照相关的配置hadoop的书籍和以下方法配置完hadoop:

**创建用户 hadoop**

$ useradd *-d /home/hadoop -s /bin/bash* hadoop

$ passwd hadoop

hadoop

可以把hadoop用户加入到 sudoer 中（使用 visudo命令），方便管理。

更改 /usr/hadoop 属主为 hadoop

**chown -R hadoop:hadoop hadoop**

cd /usr/hadoop

mkdir tmp

**配置hadoop**

hadoop配置目录 /usr/hadoop/conf

vim ./hadoop-env.sh

export JAVA\_HOME=/usr/java/jdk1.7.0\_21

[debian] 的 openjdk 主目录 /usr/lib/jvm/java-7-openjdk-armhf/

Hadoop配置文件在conf目录下，之前的版本的配置文件主要是Hadoop-default.xml和Hadoop-site.xml。由于Hadoop发展迅速，代码量急剧增加，代码开发分为了core，hdfs和map/reduce三部分，配置文件也被分成了三个**core-site.xml、hdfs-site.xml、mapred-site.xml**。core-site.xml和hdfs-site.xml是站在HDFS角度上配置文件；core-site.xml和mapred-site.xml是站在MapReduce角度上配置文件。

**core-site.xml**，请先备份。 【ip地址改成localhost也可】

vim core-site.xml

*<configuration>*

*<property>*

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

*<value>/usr/hadoop/tmp</value>*

*<description>.</description>*

*</property>*

*<property>*

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

*<value>hdfs://192.168.1.249:9000</value>*

*</property>*

*</configuration>*

**hdfs-site.xml**

*replication 是数据副本数量，默认为3，****salve****少于3台就会报错*

*<configuration>*

*<property>*

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

*<value>1</value>*

*</property>*

*<configuration>*

**mapred-site.xml**

修改Hadoop中MapReduce的配置文件，配置的是JobTracker的地址和端口。

*<configuration>*

*<property>*

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

*<value>http://192.168.1.249:9001</value>*

*</property>*

*</configuration>*

**masters**

127.0.0.1 #master的ip，如果是全分布模式，SecondaryNameNode在某台机器上，则该配置设定为SecondaryNameNode

**slaves** (主机上的文件),加入slave的hostname或者ip 如果是全分布模式，去掉 master （即namenode）

hostname

127.0.0.1

**在 slave 机器数配置 slave**

**scp复制 mster 的 /usr/hadoop 到各个 slave 即可**

**环境变量**

vim /etc/profile

export HADOOP\_HOME=/usr/hadoop

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

vim ./hadoop-env.sh

export JAVA\_HOME=/usr/java/jdk1.7.0\_10/

[note] debian 的 openjdk 主目录 /usr/lib/jvm/java-7-openjdk-armhf/

**slaves 文件无需配置；**

**配置 ssh 无密码通信**

yum instal openssh-server openssh openssh-client

各个服务器的 /home/hadoop/.ssh/authorized\_keys

本机的公钥(id\_rsa.pub)也要添加到本机的authorized\_keys中

注意：

1 设置 authorized\_keys 权限为 600 ，属主有rw权限，其他的必须没有w权限；

2 设置 vim /etc/ssh/sshd\_config

*RSAAuthentication yes # 启用 RSA 认证*

*PubkeyAuthentication yes # 启用公钥私钥配对认证方式*

*AuthorizedKeysFile .ssh/authorized\_keys # 公钥文件路径（和上面生成的文件同）*

重启ssh:service ssh restart // service ssh restart

测试本机登录：

ssh localhost

ssh 192.168.1.xxx

$ ssh-keygen

$ ls ~/.ssh

id\_rsa  id\_rsa.pub

对应公钥/私钥

复制公钥到 authorized\_keys

**启动 Hadoop**

格式化HDFS文件系统(master ~~& slaves~~)

h**adoop namenode -format**

**启动**

在启动前关闭集群中所有机器的防火墙，不然会出现datanode开后又自动关闭

service iptables stop

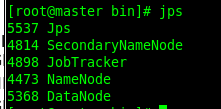
**启动：./bin/start-all.sh**

首先启动namenode 接着启动datanode1，datanode2，…，然后启动secondarynamenode。再启动jobtracker，然后启动tasktracker1，tasktracker2，…。

启动 hadoop成功后，在 Master 中的 tmp 文件夹中生成了 dfs 文件夹，在Slave 中的 tmp 文件夹中均生成了 dfs 文件夹和 mapred 文件夹。

**验证hadoop**

jps



hadoop dfsadmin -report

四个节点的信息

