hadoop2.4.1集群搭建

1.准备Linux环境

1.0先将虚拟机的网络模式选为NAT

1.1修改主机名

~~vi /etc/sysconfig/network~~

~~NETWORKING=yes~~

~~HOSTNAME=itcast ###~~

vi /etc/hostname （CentOS7）

1.2修改IP

两种方式：

第一种：通过Linux图形界面进行修改（强烈推荐）

进入Linux图形界面 -> 右键点击右上方的两个小电脑 -> 点击Edit connections -> 选中当前网络System eth0 -> 点击edit按钮 -> 选择IPv4 -> method选择为manual -> 点击add按钮 -> 添加IP：192.168.1.101 子网掩码：255.255.255.0 网关：192.168.1.1 -> apply

第二种：修改配置文件方式（屌丝程序猿专用）

vim /etc/sysconfig/network-scripts/ifcfg-eth0

DEVICE="eth0"

BOOTPROTO="static" ###

HWADDR="00:0C:29:3C:BF:E7"

IPV6INIT="yes"

NM\_CONTROLLED="yes"

ONBOOT="yes"

TYPE="Ethernet"

UUID="ce22eeca-ecde-4536-8cc2-ef0dc36d4a8c"

IPADDR="192.168.1.101" ###

NETMASK="255.255.255.0" ###

GATEWAY="192.168.1.1" ###

1.3修改主机名和IP的映射关系

vim /etc/hosts

192.168.1.101 itcast

1.4关闭防火墙

~~#查看防火墙状态~~

~~service iptables status~~

~~#关闭防火墙~~

~~service iptables stop~~

~~#查看防火墙开机启动状态~~

~~chkconfig iptables --list~~

~~#关闭防火墙开机启动~~

~~chkconfig iptables off~~

CentOS7关闭防火墙

关闭防火墙

sudo systemctl stop firewalld.service

关闭开机启动

sudo systemctl disable firewalld.service

1.5修改sudo

vim /etc/sudoers

给hadoop用户添加执行的权限

1.6关闭linux服务器的图形界面：

#查看当前运行的级别

runlevel //仍然可用

**CentOS6：**

vi /etc/inittab

# Default runlevel. The runlevels used are:

# 0 - halt (Do NOT set initdefault to this)

# 1 - Single user mode

# 2 - Multiuser, without NFS (The same as 3, if you do not have networking)

# 3 - Full multiuser mode

# 4 - unused

# 5 - X11 图形化

# 6 - reboot (Do NOT set initdefault to this)

#

id:3:initdefault:

**CentOS7：**

ls -ltr /lib/systemd/system/runlevel\*.target

lrwxrwxrwx. 1 root root 13 1月 19 00:55 /lib/systemd/system/runlevel1.target -> rescue.target

lrwxrwxrwx. 1 root root 15 1月 19 00:55 /lib/systemd/system/runlevel0.target -> poweroff.target

lrwxrwxrwx. 1 root root 17 1月 19 00:55 /lib/systemd/system/runlevel4.target -> multi-user.target

lrwxrwxrwx. 1 root root 17 1月 19 00:55 /lib/systemd/system/runlevel3.target -> multi-user.target

lrwxrwxrwx. 1 root root 17 1月 19 00:55 /lib/systemd/system/runlevel2.target -> multi-user.target

lrwxrwxrwx. 1 root root 13 1月 19 00:55 /lib/systemd/system/runlevel6.target -> reboot.target

lrwxrwxrwx. 1 root root 16 1月 19 00:55 /lib/systemd/system/runlevel5.target -> graphical.target

如设置命令行级别(init 3)方法:

ln -svf /lib/systemd/system/runlevel3.target /etc/systemd/system/default.target

或

ln -svf /lib/systemd/system/multi-user.target /etc/systemd/system/default.target

或

systemctl set-default multi-user.target

1.7重启Linux

reboot

2.安装JDK

2.1上传

alt+p 后出现sftp窗口，然后put d:\xxx\yy\ll\jdk-7u\_65-i585.tar.gz

2.2解压jdk

#创建文件夹

mkdir /home/hadoop/app

#解压

tar -zxvf jdk-7u55-linux-i586.tar.gz -C /home/hadoop/app

2.3将java添加到环境变量中

vim /etc/profile

#在文件最后添加

export JAVA\_HOME=/home/hadoop/app/jdk-7u\_65-i585

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

#刷新配置

source /etc/profile

3.安装hadoop2.4.1

先上传hadoop的安装包到服务器上去/home/hadoop/

注意：hadoop2.x的配置文件$HADOOP\_HOME/etc/hadoop

伪分布式需要修改5个配置文件

3.1配置hadoop

第一个：hadoop-env.sh

vim hadoop-env.sh

#第54行

export JAVA\_HOME=/usr/java/jdk1.7.0\_65

第二个：core-site.xml

<configuration>

<!-- 指定HADOOP所使用的文件系统schema（URI），HDFS的老大（NameNode）的地址 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://Hadoop01:9000</value>

</property>

<!-- 指定hadoop运行时产生文件的存储目录 -->

<property>

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

<value>/home/hadoop/app/hadoop-3.0.0-alpha3/tmp</value>

</property>

</configuration>

第三个：hdfs-site.xml

<configuration>

<!-- 指定HDFS副本的数量 -->

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<!-- 解决找不到dataNode节点的问题 -->

<property>

<name>dfs.namenode.datanode.registration.ip-hostname-check</name>

<value>false</value>

</property>

<!-- 配置Secondary NameNode -->

<property>

<name>dfs.secondary.http.address</name>

<value>Hadoop02:50090</value>

</property>

</configuration>

第四个：mapred-site.xml

mv mapred-site.xml.template mapred-site.xml

vim mapred-site.xml

<configuration>

<!-- 指定mr运行在yarn上 -->

<property>

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

<value>yarn</value>

</property>

<!-- 解决使用mapreduce时出错 -->

<property>

<name>mapreduce.admin.user.env</name>

<value>HADOOP\_MAPRED\_HOME=$HADOOP\_COMMON\_HOME</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP\_MAPRED\_HOME=$HADOOP\_COMMON\_HOME</value>

</property>

</configuration>

第五个：yarn-site.xml

<configuration>

<!-- 指定YARN的老大（ResourceManager）的地址 -->

<property>

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

<value>Hadoop01</value>

</property>

<!-- reducer获取数据的方式 -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>

3.2将hadoop添加到环境变量

vim /etc/profile

export JAVA\_HOME=/usr/java/jdk1.7.0\_65

export HADOOP\_HOME=/itcast/hadoop-2.4.1

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

source /etc/profile

3.3格式化namenode（是对namenode进行初始化）

hdfs namenode -format (~~hadoop namenode -format 以前方式~~)

3.4.0 配置启动hdaoop

在nameNode中的etc/hadoop/的works文件添加内容

vim workers

Hadoop02

Hadoop03

Hadoop04

3.4启动hadoop

先启动HDFS

sbin/start-dfs.sh

再启动YARN

sbin/start-yarn.sh

3.5验证是否启动成功

使用jps命令验证

[hadoop@Hadoop01 hadoop]$ jps

2738 DataNode

3122 ResourceManager

2632 NameNode

3513 Jps

3231 NodeManager

~~http://Hadoop01:50070 （HDFS管理界面） 3.0以前版本~~

http://Hadoop01:9870 （HDFS管理界面） 3.0以后版本

http://Hadoop01:8088 （MR管理界面）

https://issues.apache.org/jira/browse/HDFS-9427

修补程序将HDFS默认HTTP / RPC端口更新为非临时端口。更改如下：  
Namenode端口：50470 - > 9871，50070 - > 9870,8020 - > 9820  
辅助NN端口：50091 - > 9869,50090 - > 9868   
Datanode端口：50020 - > 9867， 50010 - > 9866，50475 - > 9865，50075 - > 9864

4.配置ssh免登陆

#生成ssh免登陆密钥

#进入到我的home目录

cd ~/.ssh

ssh-keygen -t rsa （四个回车）

执行完这个命令后，会生成两个文件id\_rsa（私钥）、id\_rsa.pub（公钥）

将公钥拷贝到要免密登陆的目标机器上

ssh-copy-id localhost

ssh免登陆：

生成key:

ssh-keygen

复制从A复制到B上:

ssh-copy-id B

验证：

ssh localhost/exit，ps -e|grep ssh

ssh A #在B中执行