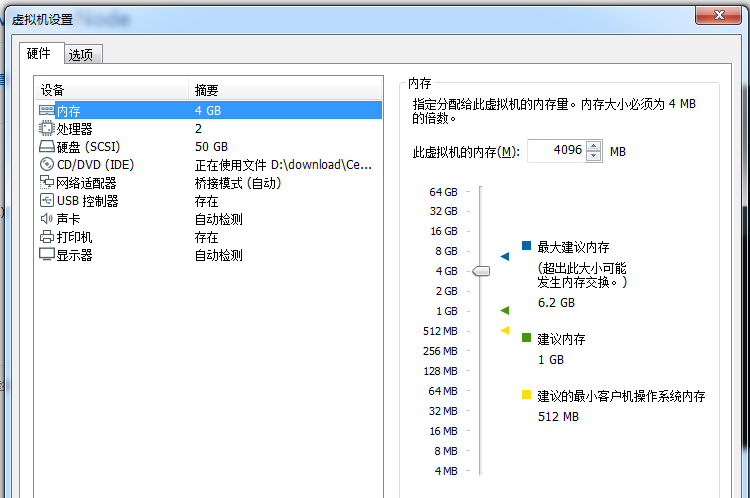
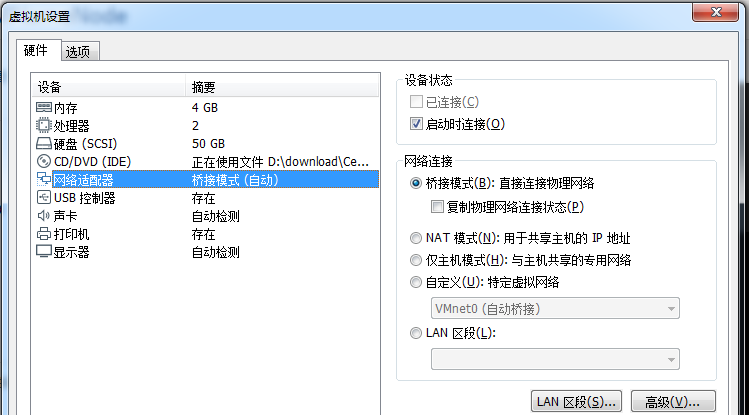
大数据第11周AM

1. 物理集群的搭建

开启本组所有电脑（4台），拷贝虚拟机映像，打开。在开启虚拟机之前，设置虚拟机内存是4G，网络连接是桥接模式。





把桥接网卡绑定到具体的物理网卡，不要使用自动绑定。



设置完成后，启动虚拟机。

1. 设置网络连接

编辑网络配置文件：/etc/sysconfig/network-scripts/ifcfg-ens33，IP地址设置如下：

IPADDR=172.21.12. 你自己的ip

PREFIX=24

GATEWAY=172.21.12.254

DNS1=172.16.3.8

1. 设置hostname和hosts

修改集群各个机器的名字和域名解析。

1. 设置ssh免密登陆

由于以前已经做过免密，所以会提示出错，看提示是否有修改错误的指令，如果没有，删除~/.ssh后重做：rm -r .ssh。

5.设置时钟同步

将时区设置为上海：sudo timedatectl set-timezone Asia/Shanghai

使用chrony。因为centos7已经内置了chrony，所以只要配置就可以了。

选择一个本地时钟服务器，例如master，对于时钟服务器修改/etc/chrony.conf，要允许其他节点进行连接：

allow 172.21.12.0/24。

从节点时钟服务器地址修改成master，其他的注释掉，用#。

所有节点，执行

sudo systemctl enable chronyd.service

sudo systemctl start chronyd.service

重启所有节点。

在主节点上运行：chronyc sources

可以看到各个时钟服务器的延迟时间。

在从节点上运行：chronyc sources

可以看到连接到master的延迟时间，如果不是0，就说明连接正常了。

1. 修改配置文件

修改slaves文件，加入新的节点。

Slaves文件位置：hadoop-2.7/etc/hadoop。

1. Namenode格式化

删除所有节点上的集群自动建立的文件夹：hadoopdata。

在master上操作：hadoop namenode -format。

要看到successfully提示，说明成功了。

1. 启动集群

命令：start-all.sh

用jps查看java进程：按我们的配置，master是4个java进程，slave是3个java进程。

上传一个文件，查看是否正常。

运行命令：hadoop jar hadoop-mapreduce-examples-2.7.7.jar pi 10 10

如果运行正常，说明集群工作一切正常。

8. 安装spark

Spark配置成：yarn模式运行。

Spark集群需要配置两个文件，spark-env.sh、slaves。

建立spark-env.sh ：cp spark-env.sh.template spark-env.sh

编辑：vi spark-env.sh，添加如下内容，红色部分替换成自己的目录。

export SPARK\_CONF\_DIR=/home/caiyishuai/spark-2.4.5/conf

export HADOOP\_CONF\_DIR=/home/caiyishuai/hadoop-2.7/etc/hadoop

export YARN\_CONF\_DIR=/home/caiyishuai/hadoop-2.7/etc/hadoop

export JAVA\_HOME=/home/caiyishuai/jdk1.8

建立slaves文件：cp slaves.template slaves

编辑：vi slaves

把文件原内容删除，添加slave节点名：例如

slave0

slave1

slave2

文件编辑后拷贝到集群所有节点。

9.启动spark集群

由于我们是用yarn模式运行，首先要启动yarn，启动yarn的方式是用hadoop下的start-all.sh。启动顺序是：hadoop的start-all.sh，spark的start-all.sh。

由于两个文件同名，所以必须区分是哪一个。区分的方式可以是，spark路径不配置环境变量，而是用全路径执行。Spark要在master节点上启动。

启动后，用jps查看，master节点：

NameNode

Jps

ResourceManager

Master

SecondaryNameNode

Slave节点：

DataNode

NodeManager

Worker

Jps

说明启动成功。