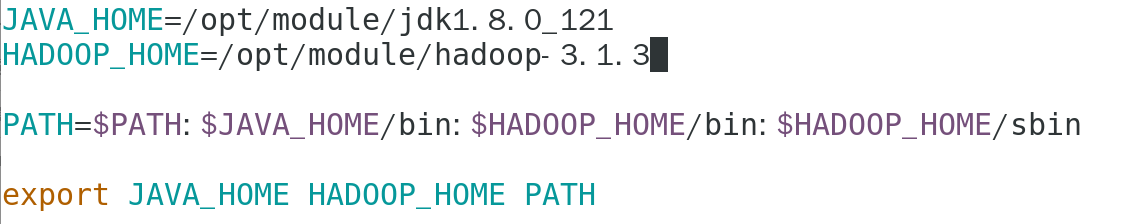
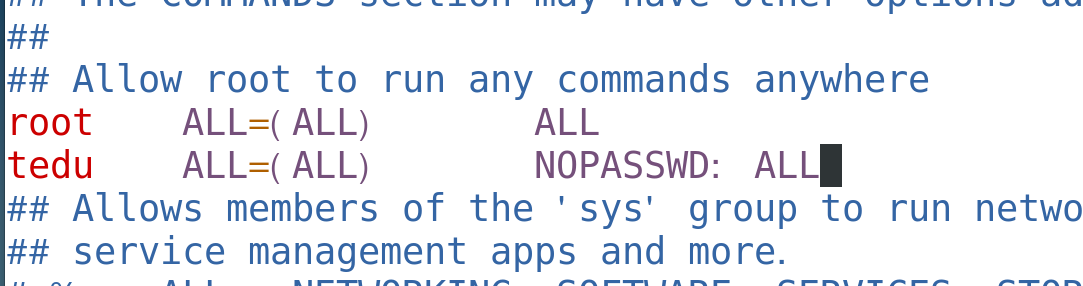
**集群搭建步骤**

1. 打开虚拟机后，使用root用户登录，在根目录下的opt目录中创建两个目录，一个soft，另一个module。Soft目录用于存放压缩文件，module目录用于存放解压后的文件。
   1. 命令：mkdir /opt/soft /opt/module
2. 上传jdk以及hadoop的压缩包到/opt/soft（注意：现目前用户是否是root用户）
3. 将jdk以及hadoop解压到module目录中。
   1. tar -zxvf jdk..... -C /opt/module
   2. tar -zxvf hadoop... -C /opt/module
4. 配置环境变量，在/etc/profile中配置以下内容：



配置集群前的准备工作：

1. 创建一个新用户。（如果已经存在新用户，该步骤可以省略）
2. 给该用户root权限



1. 关闭防火墙

systemctl stop firewalld 停止防火墙

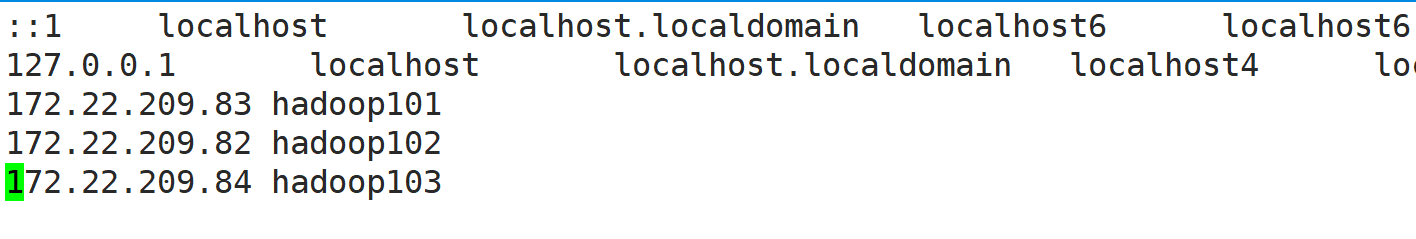
systemctl disable firewalld 永久禁止防火墙开机自启

1. 把opt目录下soft目录以及module目录交给普通用户

chown -R tedu:tedu /opt/soft /opt/module

克隆（保证有3台机器运行）

1. 保证3台机器能够相互通信，也就是需要配置/etc/hosts,在此之前，需要把主机名更改为不同名字，如hadoop101，hadoop102，hadoop103。
   1. 更改主机名的文件在/etc/hostname,将原有名字全部删除，并写入新的名字，保存并退出，再重启。
   2. 将每台机器的ip地址复制，粘贴到/etc/hosts文件中，并在后方写入对应的主机名，如下图：



Hadoop的节点分配

Map

Reduce

resourceManager

nodeManager

datanode

namenode

secondaryNameNode

YARN

HDFS

MapReduce

如何避免单点故障：只需要咱们合理的分配节点在集群的机器中（均衡搭配）。

Hadoop101: namenode datanode nodemanager

Hadoop102: resourcemanager datanode nodemanager

Hadoop103: secondarynamenode datanode nodemanager

1. 要启动HDFS，首先需要配置core-site.xml文件

（该文件位于/opt/module/hadoop-3.1.3/etc/hadoop/）

* 1. 在文件中的<configuretion>标签中添加以下配置

<!-- 决定hdfs运行模式 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop101:8020</value>

</property>

<!-- 配置hadoop临时目录 -->

<property>

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

<value>/opt/module/hadoop-3.1.3/data</value>

</property>

<!-- 配置web端页面的静态用户 -->

<property>

<name>hadoop.http.staticuser.user</name>

<value>tedu</value>

</property>

<!--定义HDFS所开放的代理服务 -->

<property>

<name>hadoop.proxyuser.tedu.hosts</name>

<value>\*</value>

</property>

<property>

<name>hadoop.proxyuser.tedu.groups</name>

<value>\*</value>

</property>

(2). 格式化namenode(自动创建hadoop临时目录)

命令：hadoop namenode -format

(3). 启动namenode以及datanode节点(启动HDFS)

命令:hadoop-daemon.sh start namenode

hadoop-daemon.sh start datanode

1. 配置hdfs-site.xml（与core-site.xml在同级目录）

<property>

<name>dfs.namenode.http-address</name>

<value>hadoop101:9870</value>

</property>

<!-- 2nn web端访问地址-->

<property>

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

<value>hadoop103:9868</value>

</property>

1. 配置mapred-site.xml(与core-site.xml在同级目录)

<property>

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

<value>yarn</value>

</property>

<!-- 历史服务器运行机器以及端口 -->

<property>

<name>mapreduce.jobhistory.address</name>

<value>hadoop102:10020</value>

</property>

<!-- 历史服务器web端地址 -->

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>hadoop102:19888</value>

</property>

1. 配置yarn-site.xml（与core-site.xml在同级目录）

<!-- 指定MR走shuffle -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<!-- 指定ResourceManager的地址-->

<property>

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

<value>hadoop102</value>

</property>

<!-- 环境变量的继承 -->

<property>

<name>yarn.nodemanager.env-whitelist</name>

<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>

</property>

<!-- yarn容器允许分配的最大最小内存 -->

<property>

<name>yarn.scheduler.minimum-allocation-mb</name>

<value>1024</value>

</property>

<property>

<name>yarn.scheduler.maximum-allocation-mb</name>

<value>4096</value>

</property>

<property>

<name>yarn.nodemanager.resource.memory-mb</name>

<value>4096</value>

</property>

<property>

<name>yarn.nodemanager.pmem-check-enabled</name>

<value>false</value>

</property>

<property>

<name>yarn.nodemanager.vmem-check-enabled</name>

<value>false</value>

</property>

<property>

<name>yarn.log-aggregation-enable</name>

<value>true</value>

</property>

<!-- 设置日志聚集服务器地址 -->

<property>

<name>yarn.log.server.url</name>

<value>http://hadoop102:19888/jobhistory/logs</value>

</property>

<!-- 设置日志保留时间为7天 -->

<property>

<name>yarn.log-aggregation.retain-seconds</name>

<value>604800</value>

</property>

1. 集群启动与停止
   1. 启动：
      1. Hadoop101命令:
         1. hadoop-daemon.sh start namenode
         2. hadoop-daemon.sh start datanode
         3. yarn-daemon.sh start nodemanager
      2. Hadoop102命令：
         1. hadoop-daemon.sh start datanode
         2. yarn-daemon.sh start nodemanager
         3. yarn-daemon.sh start resourcemanager
      3. Hadoop103命令:
         1. hadoop-daemon.sh start secondarynamenode
         2. yarn-daemon.sh start nodemanager
         3. hadoop-daemon.sh start datanode
   2. 停止：

就是把启动命令中的start替换成stop