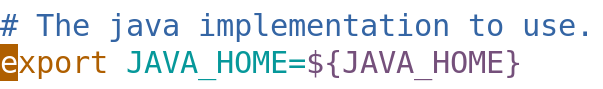
大数据第五周

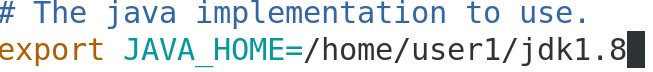
1配置hadoop集群

1.1配置环境

（1）配置hadoop-env.sh



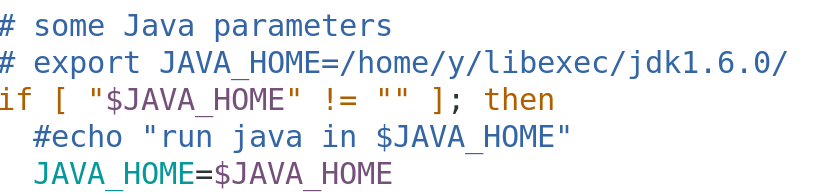
找到这句，把JAVA\_HOME写进来，如下图：



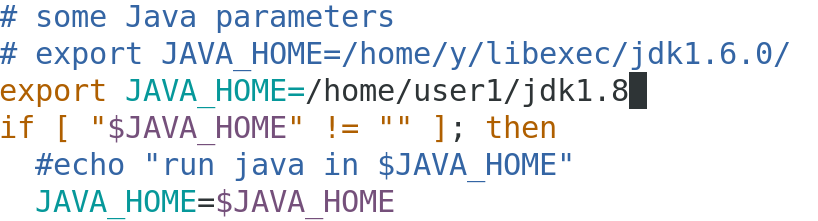
注意：是修改。

（2）配置yarn-env.sh

找到如下图位置：



将JAVA\_HOME变量写在if前面，例如：



1.2 配置核心配置文件core-site.xml

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

<property>

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

<value>/home/user1/hadoopdata</value>

</property>

1.3配置hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>master:9001</value>

</property>

说明：本文件如上的配置不可以用于实际使用。原因：（1）文件副本存储量是1，对文件而言毫无安全性可言；（2）second namenode是作为namenode失效后的后备使用的，我们当前的配置是把second namenode放到了和namenode相同的物理节点上，这样毫无意义，因为：当namenode失效时，second namenode也一并失效了。

我们这样配置是因为条件所限和仅仅是为了实验。

1.4配置yarn-site.xml

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce\_shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.resourcemanager.address</name>

<value>master:8032</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>master:8030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>master:8035</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>master:8033</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>master:8088</value>

</property>

1.5配置mapred-site.xml

先将模板文件存成正常文件：cp mapred-site.xml.template mapred-site.xml

<property>

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

<value>yarn</value>

</property>

1.6 配置slaves文件

把文件原内容删除，然后添加

slave0

slave1

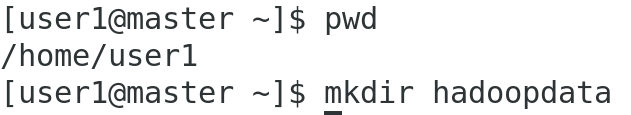
2.拷贝hadoop文件包到其他节点

命令是：scp –r hadoop-2.7 slave0:~，最后一个是波浪号。

命令是：scp –r hadoop-2.7 slave1:~，最后一个是波浪号。

3.格式化hdfs

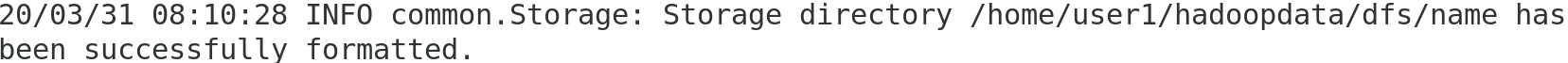
首先按照配置文件，在namenode上建立文件夹，我们配置的是/home/user1/hadoopdata。



然后，在namenode上执行：

hadoop namenode -format

如果看到如下内容，说明格式化成功了。

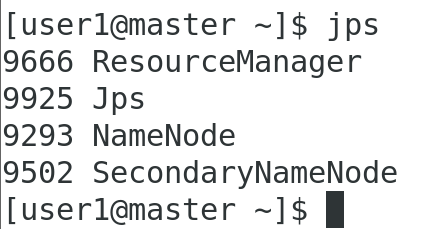


格式化不能多次执行，如果不成功，需要清空文件夹，才能再次执行格式化。

4.启动集群

启动命令：start-all.sh

启动后用jps查看java进程，namenode有4个，datanode有3个。如下图：





关闭命令：stop-all.sh

注意：关闭linux之前，务必先关闭集群，否则下次启动集群时容易出错。