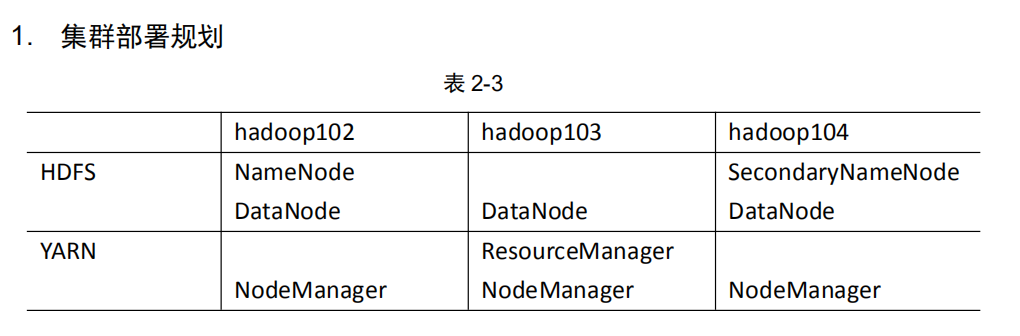
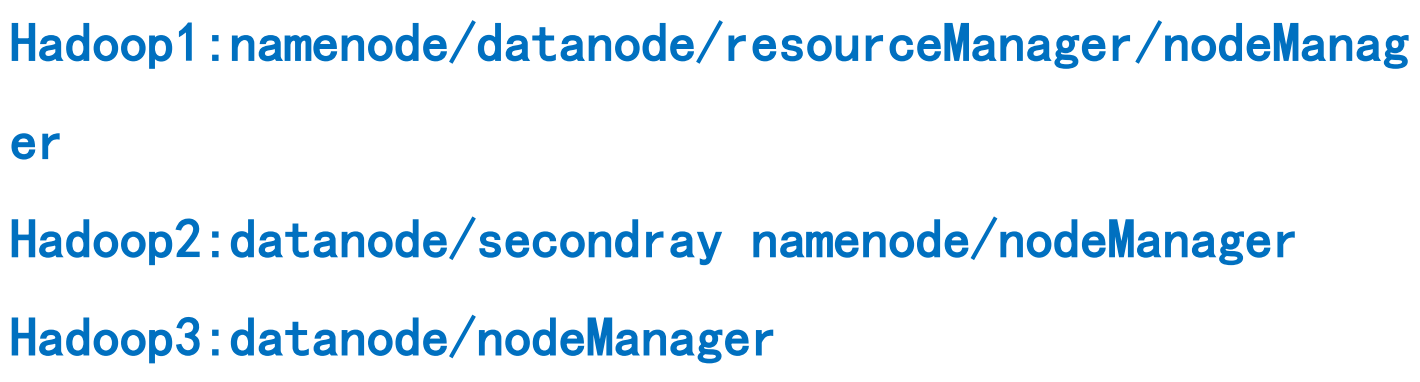
集群环境启动

1. **hadoop**





先格式化namenode，然后启动hadoop。（都在master进行操作，就是你namenode所在的服务器）

先格式化namenode

bin/hdfs namenode -format

然后启动hadoop。

sbin/start-all.sh

或者使用

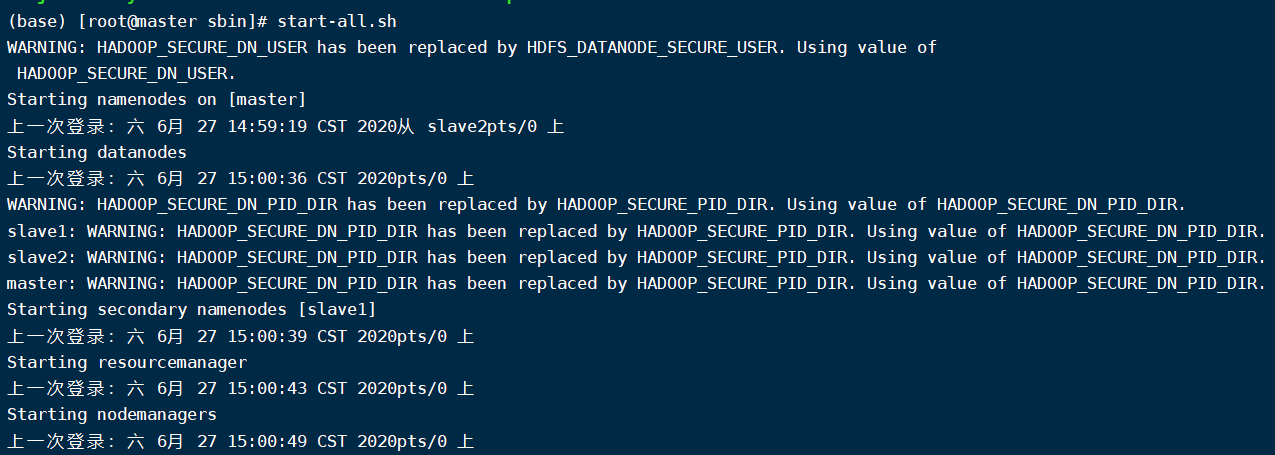
sbin/start-dfs.sh ;

sbin/start-yarn.sh:

注意：NameNode 和 ResourceManger 如果不是同一台机器，不能在 NameNode 上启动 YARN，应该在 ResouceManager 所在的机器上启动 YARN。

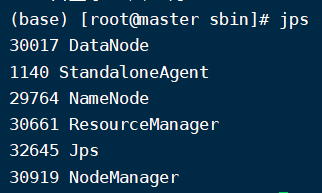
yarn监控界面：http://172.16.29.109:8088/

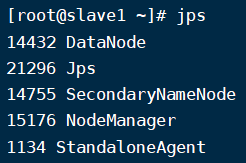
hadoop监控界面：http://172.16.29.109:9870/

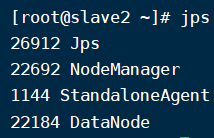


WARINING解决方法（不解决也没事）：

<https://blog.csdn.net/weixin_38763887/article/details/79157652>





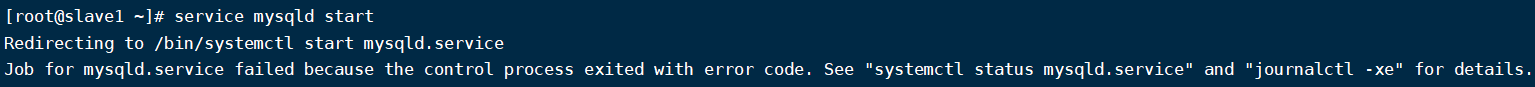


1. **hive**

**2.1. slave2（赵磊）mysql出问题了**



ERROR 2002 (HY000): Can't connect to local MySQL server through socket '/var/lib/mysql/mysql.sock' (111)



先按照这个卸载了

<https://www.cnblogs.com/nicknailo/articles/8563456.html>

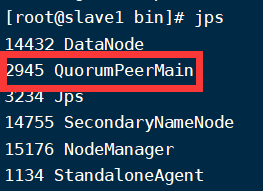
再按这个重安的

方法一：<https://www.cnblogs.com/Ishtarin/p/13202666.html>

方法二：<https://blog.csdn.net/qq_37598011/article/details/93489404>

1. **zookpeper**

bin/zkServer.sh start （每台服务器都得去启动）



报错一：集群上的三个服务器时间不一致

<https://www.cnblogs.com/lenmom/p/10348150.html?tdsourcetag=s_pctim_aiomsg>



主要看这四步

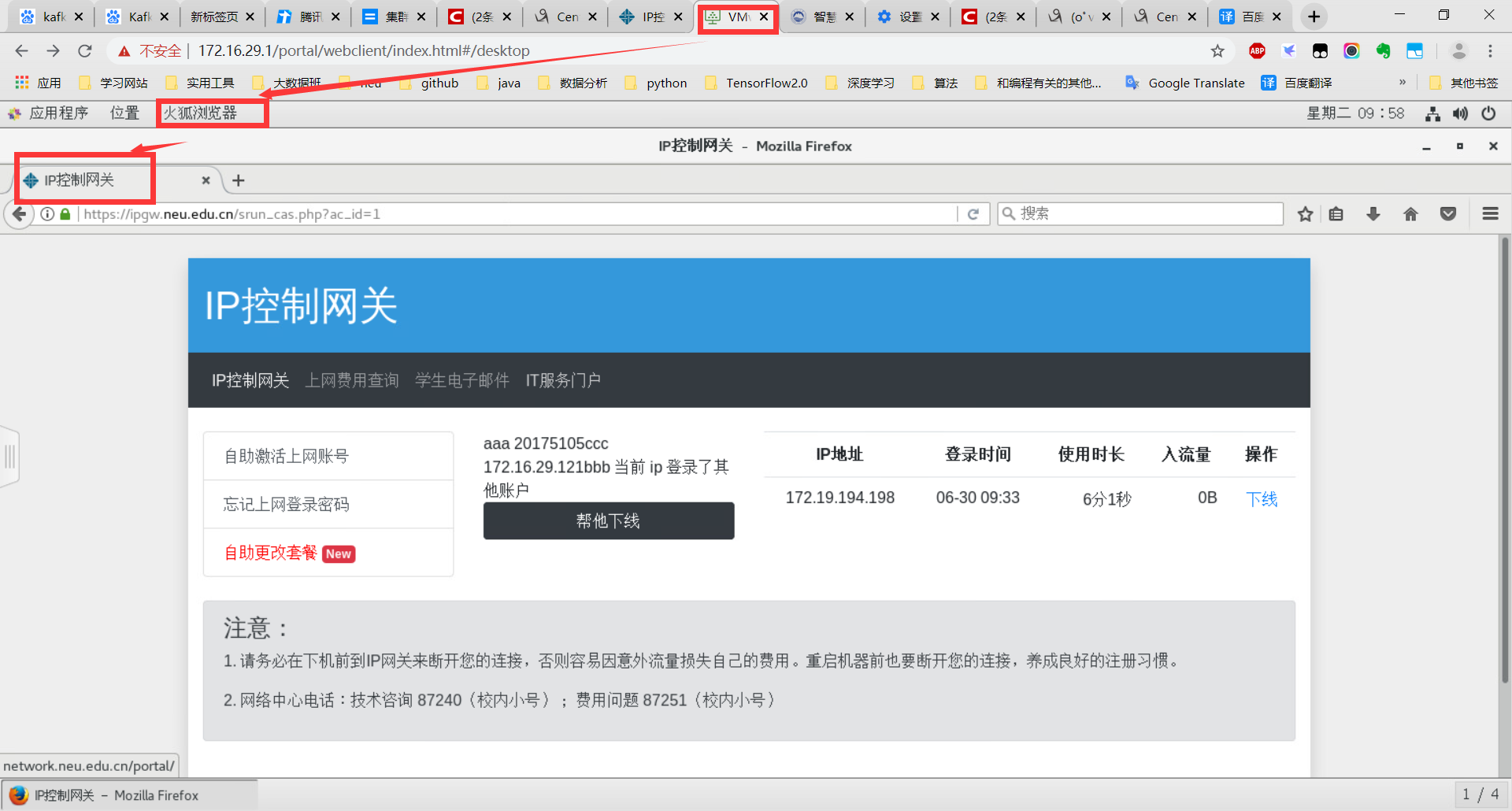


第二步需要联网

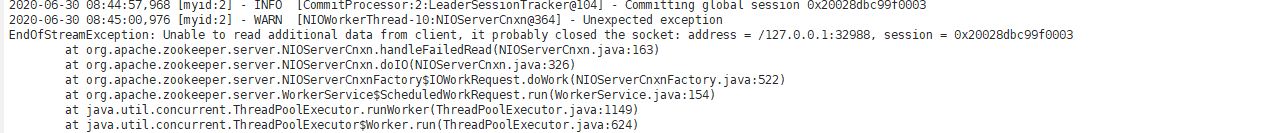
东北大学联网方式ipgw：

方法一（不要图形界面）：<https://github.com/unbyte/ipgw>

方法二（需要图形界面）：去网页版的Vmware Horizon，进入自己服务器，点击火狐浏览器，之后像Windows那样去联网就行



报错二：



原因：打开 zkCli.sh，然后 ctrl + c 关闭

解决方案，输入 quit退出，就没这个问题了

1. **hbase**

**以下所有命令到master执行就行**

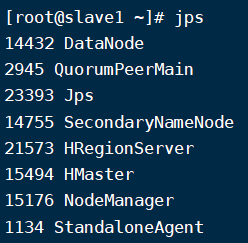
bin/hbase-daemon.sh start master

bin/hbase-daemon.sh start regionserver

或者

bin/start-hbase.sh

bin/stop-hbase.sh



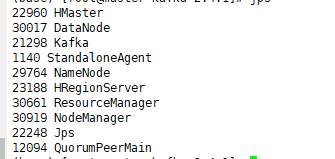
http://172.16.29.109:16010/master-status

1. **kafka**

每台节点在 /BigData/kafka/ 下运行 bin/kafka-server-start.sh config/server.properties &

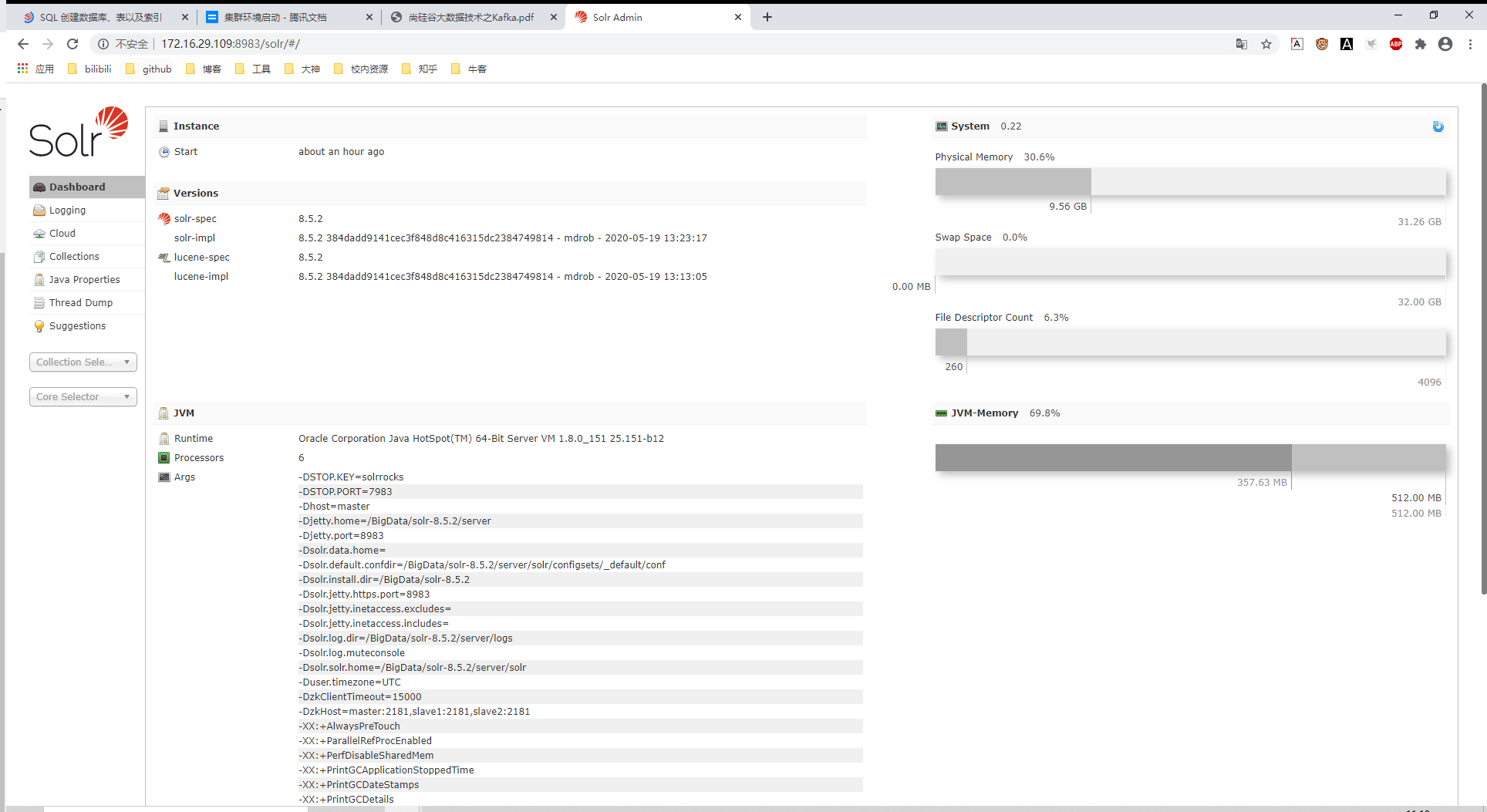
遇到的问题：因为之前集成了 atals，所以现在kafka会在控制台上隔断时间输出 atlas 的消息

解决方法：改变启动方式，在启动命令后加上 &>/dev/null 把日志输出到黑洞，再加上 & 后台运行



1. **solr**

每台节点在/BigData/solr-8.5.2/bin下运行 ./solr start -force，然后到 http:/ip:8983 下查看web页面



1. **atlas**

atlas 只需要启动一台节点即可，在任意节点的/BigData/atlas-2.0.0/bin 下运行 atlas-start.py 即可，到启动服务的节点ip:21000下看web 页面

1. **spark**

运行 /BigData/spark-3.0.0/sbin 下的 start-all.sh即可

master 下增加了 Master 进程

slave 下增加了 Worker 进程

**启动 spark shell 查看能否运行：**

spark on yarn: spark-shell --master yarn

yarn 的端口试 8088 172.16.29.109:8088

spark 的是 8888 172.16.29.109:8888

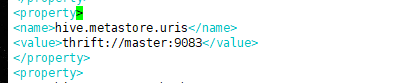
遇到的问题

1. **hive metastroe 的问题**

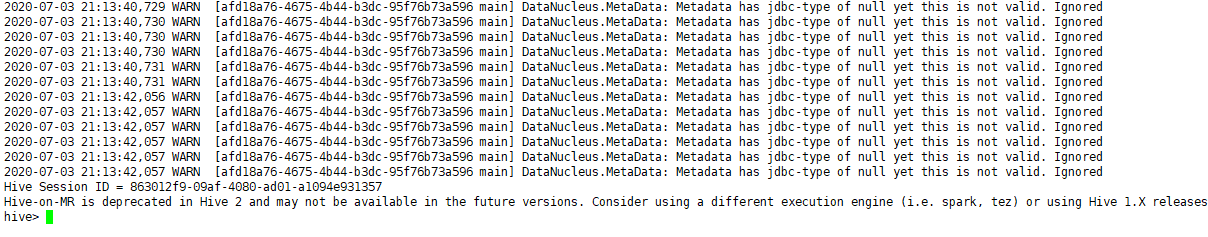
这个问题最早的时候是在启动 hive 的时候遇到的，当时报错是：



我把 hive-site.xml 中的这部分配置注释掉之后就不再报错了，但是会在启动 hive 的时候有警告，当时把这个问题直接忽略过去了，因为已经可以 hive 已经可以正常运行了

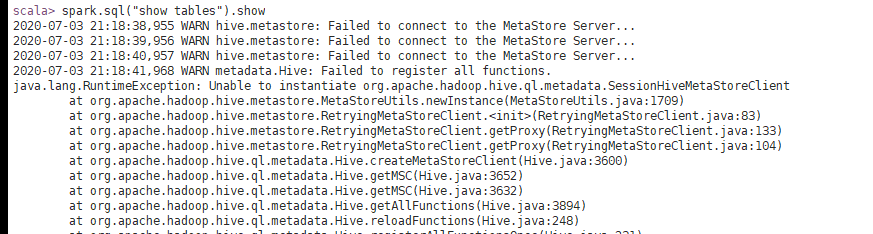


（被注释掉的代码）



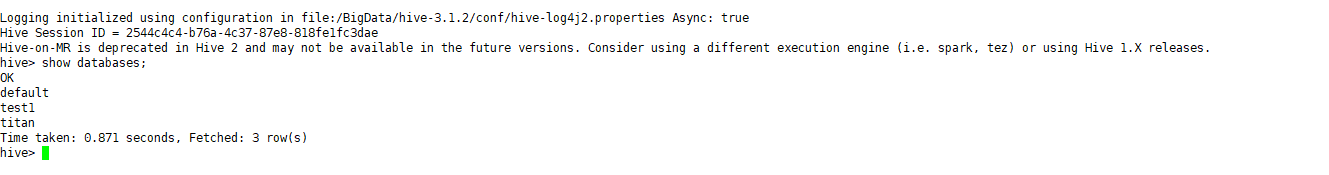
（启动 hive 时的警告）

后来在尝试 spark on hive 的时候，报了一个这个错误：

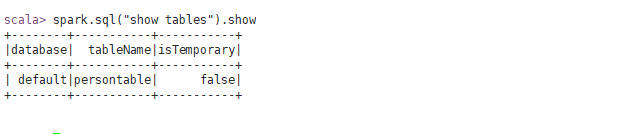


通过查资料，发现是需要启动 service metastore，看到 metastore 就很熟悉，上面遇到过，于是我把上面注释掉的代码解除注释，并在 master（与配置文件里的 ip 对应） 上运行 **hive --service metastore** 命令，重新启动 spark-shell，问题解决。

nohup hive --service metastore 2>/dev/null &（运行命令）



（hive 正常）



（spark 正常）

1. **hiveserver2 的启动问题**

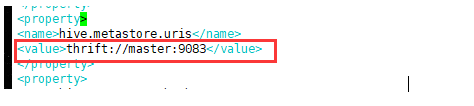
我们启动 hiveserver2 来开启 10000 端口，通过 idea 连接到 hive，方便地写建表语句，第一次启动没有任何问题，idea 成功连接，但是过了一段时间，idea 就突然连接不上去了，我到 linux 把进程先 kill 掉，然后重新运行，发现报错：

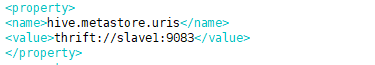


我当时从网上查了一些资料，发现与我们的情况都不符合，我们不可能是因为缺少包导致的，否则第一次就启动不起来，所以我就尝试在 master 上也启动 hiverserver2，发现报了同样的错误，但是在该错误的下面还有其他正常信息：



看起来是正常的样子，于是我尝试用 idea 连接 master 的 hive，发现连上去了，**果然和缺少Tez 没有关系，**master 可以启动 hiveSerer2，但是 slave1 不可以，二者关于 hive 的差别就是在这里（图一），master 就位于 metastore 服务的主机上，而 slave1 不是。因为我们要把 slave1 当作 hive 的主机，所以我就在三台节点上都做了更改，把 master 改成 slave1（图二）

（图一）

 （图二）

更改完配置之后，kill 掉 master 上的 hive metastore service，然后在 slave1 上启动 hive metastore service 和 hiveserver2，问题解决！