Hadoop集群详情

一、hadoop集群服务器介绍

集群总共3台服务器，一台namenode节点，两台datanode节点，

，三台电脑均在路由器中，且路由器连接在内网中，使用一个内网IP，属于内网的子网（第二级网络）。

路由器详情：

路由器网关：192.168.1.1（可在网站直接访问）

路由器管理网站用户名：admin，密码：admin

Wifi名：TP-LINK\_test

密码：wislifetest

路由器内网地址：192.168.0.198，网关：192.168.0.1，DNS：192.168.0.1



路由器已经做了DMZ主机映射，把IP 192.168.0.198映射给192.168.1.102，访问0.198的请求都会转发到1.102主机中。

二、服务器详情如下：

1、Namenode198节点：192.168.1.102，主机名namenode102，系统：CentOS Linux release 8.0.1905；访问192.168.0.198可以转发至102主机上

内存：4G，硬盘：60G

1. Datanode101：192.168.1.101，主机名datanode101，系统：CentOS Linux release 7.6.1810，

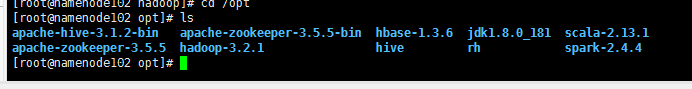
内存：8G，硬盘：1T

3、Datanode103：192.168.1.103，主机名datanode103，系统

CentOS Linux release 7.6.1810

内存：8G，硬盘：1T

集群软件路径：/opt/



Jdk-1.8.0\_181，

Hadoop：hadoop-2.6.0-cdh5.9.1，

Zookeeper：zookeeper-3.4.5-cdh5.9.1，

Hbase：hbase-1.2.0-cdh5.9.1，

hive：hive-1.1.0-cdh5.9.1，

Spark：spark-1.6.0-cdh5.9.1，（暂时没安装完成）

**Hadoop服务器列表**

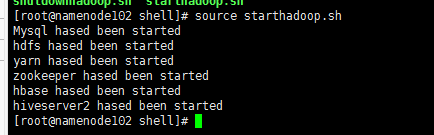
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 服务器IP | 主机名 | 任务列表 | 启用端口 | 安装目录 | 软件 |
| 192.168.1.101 | Datanode101 | DataNode  NodeManager  HQuorumPeer  Worker | 50075  8042 | /opt/hadoop-2.6.0-cdh5.9.1  /opt/spark-1.6.0-cdh5.9.1  /opt/hbase-1.2.0-cdh5.9.1  /opt/zookeeper-3.4.5-cdh5.9.1 | HDFS |
| 192.168.1.102 | Namenode198 | NameNode  SecondaryNamenode  ResourceManager  Master  HMaster  HQuorumPeer | 50070  50090  8088  9000  60010 | /opt/hadoop-2.6.0-cdh5.9.1  /opt/spark-1.6.0-cdh5.9.1  /opt/hbase-1.2.0-cdh5.9.1  /opt/zookeeper-3.4.5-cdh5.9.1  /opt/hive-1.1.0-cdh5.9.1 | HDFS  Yarn |
| 192.168.1.103 | Datanode103 | DataNode  NodeManager  HQuorumPeer  Worker | 50075  8042 | /opt/hadoop-2.6.0-cdh5.9.1  /opt/spark-1.6.0-cdh5.9.1  /opt/hbase-1.2.0-cdh5.9.1  /opt/zookeeper-3.4.5-cdh5.9.1 | HDFS |

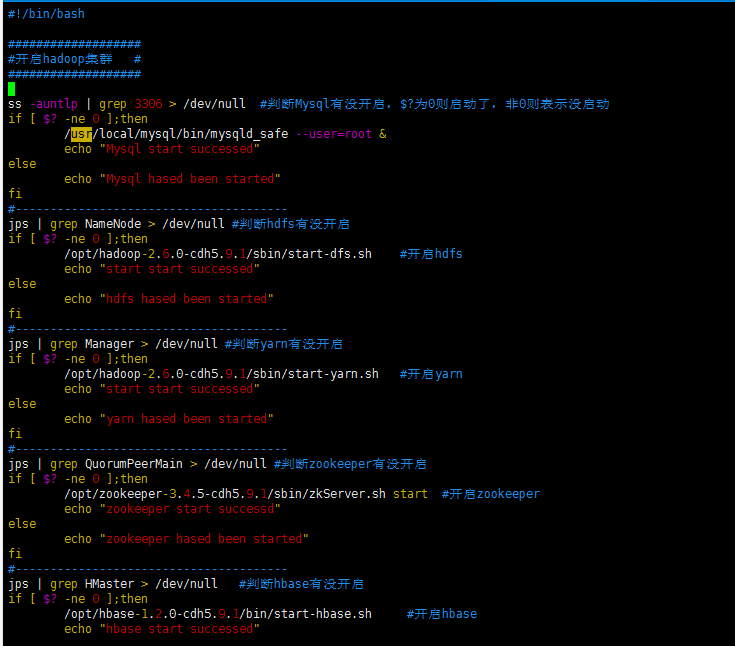
1. 集群管理

在namenode102中，路径/root/shell/starthadoop.sh和/root/shell/shutdown.sh为启动和关闭集群脚本，但在/etc/rc.local中也设置了开机自启，故开机后可不用执行starthadoop.sh脚本，执行脚本也没关系，脚本会自动判断哪个组件没开启，然后单独启动那个组件！

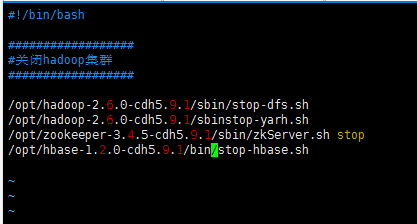
执行启动hadoop脚本：source /root/shell/starthadoop.sh

#可反复执行，但执行时需要保证datanode101和datanode103网络通讯，开机时最好是先开这两台，最后开这台namenode102（宝贝窗主机）

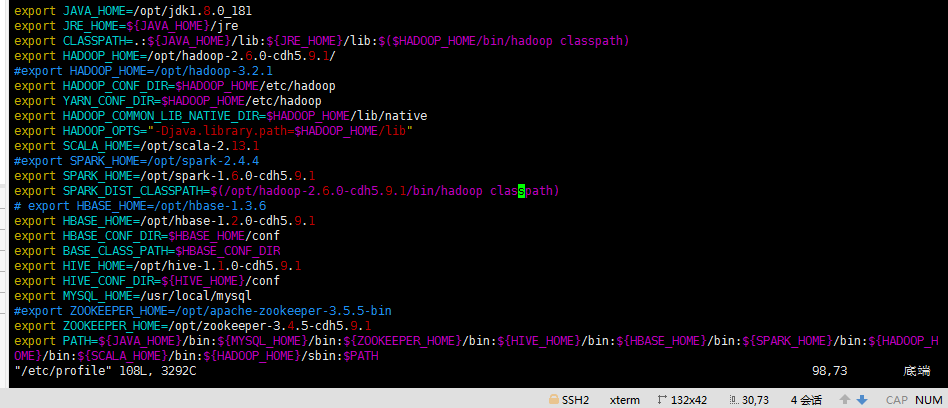




执行关闭hadoop脚本:source /root/shell/shutdownhadoop.sh



环境变量在：/etc/profile 文件中



1、使用命令检测集群情况：

hdfs dfsadmin -report #集群任意节点执行都可以,如图所示



1. **启动HDFS**

Hdfs启动：start-dfs.sh #启动hdfs服务，在198上执行即可

Hdfs关闭：stop-dfs.sh #启动hdfs服务，在198上执行即可

用命令jps，查看各台机器上对应进程已经起来：

namenode102主机：NameNode、SecondaryNameNode；

datanode101、103主机：DataNode，（9000端口是连接的）

**3、启动Yarn**

Yarn启动：start-yarn.sh #在198上执行即可

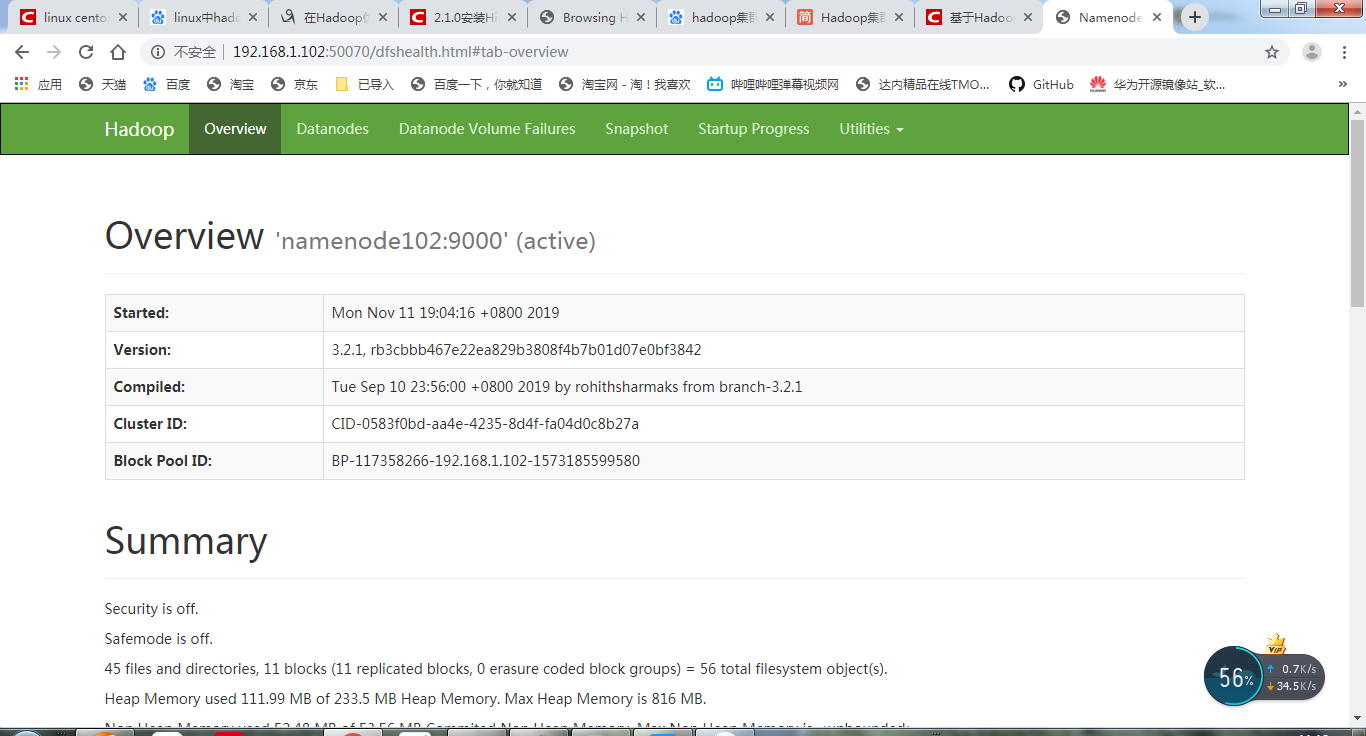
Yarn关闭：stop-yarn.sh #在198上执行即可

jps查看各台机器上对应进程已经起来：

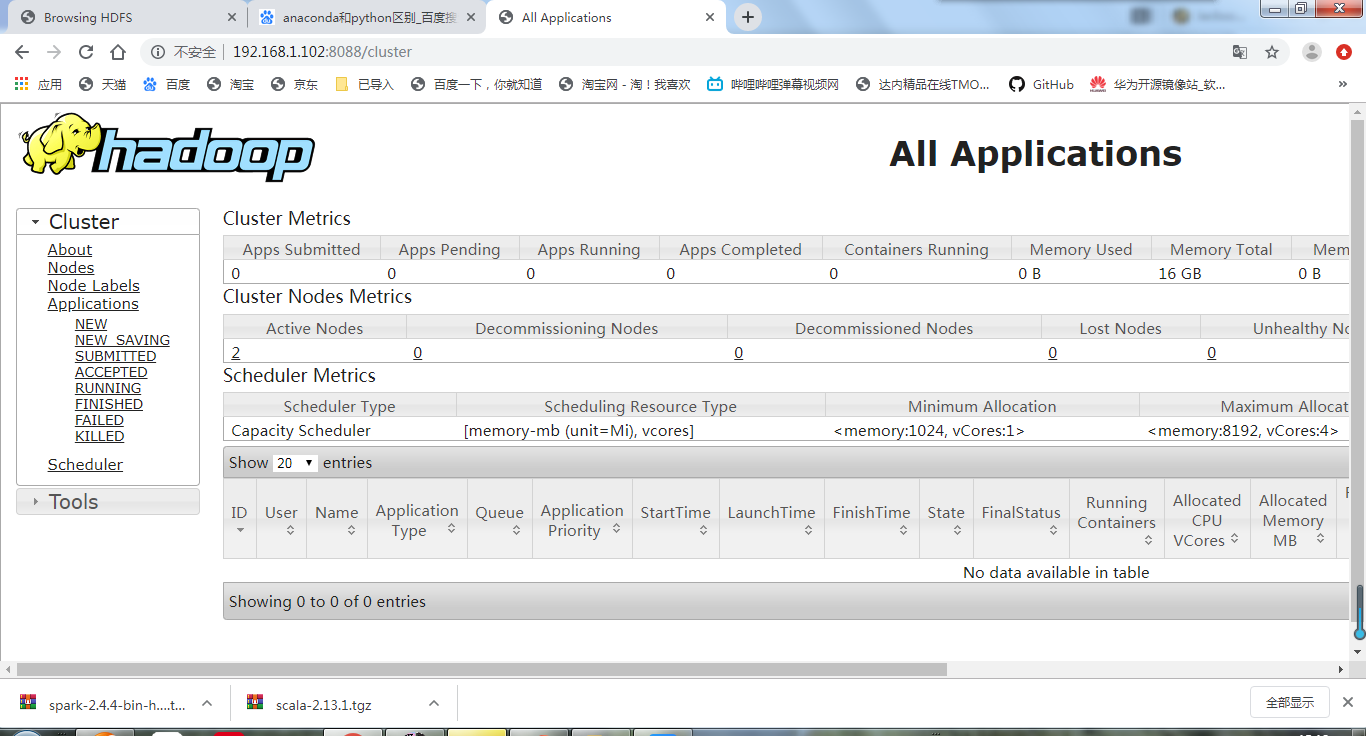
namenode102主机新增服务：ResourceManager；

datanode101、103主机：NodeManager；

http://192.168.0.198:50070/ //--namenode198 web页面，浏览器上登录该网址即可（namenode198）

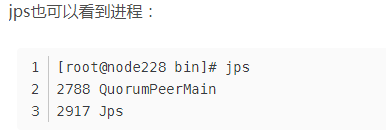


[http://192.168.0.198:8088/cluster](http://192.168.1.102:8088/cluster) （查看数据节点datanode和集群使用情况，浏览器上登录该网址即可）



**4、Zookeeper启动：zkServer.sh start #3台电脑都需要执行该命令**

**查看状态：zkServer.sh status**



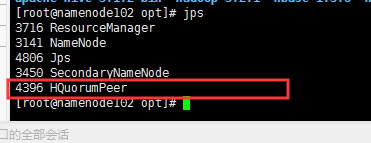
会新增服务：QuorumPeerMain

1. Hbase集群

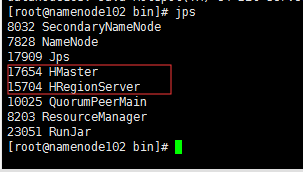
Hbase集群启动：start-hbase.sh #在198服务器上执行

停止：stop-hbase.sh #在198服务器上执行

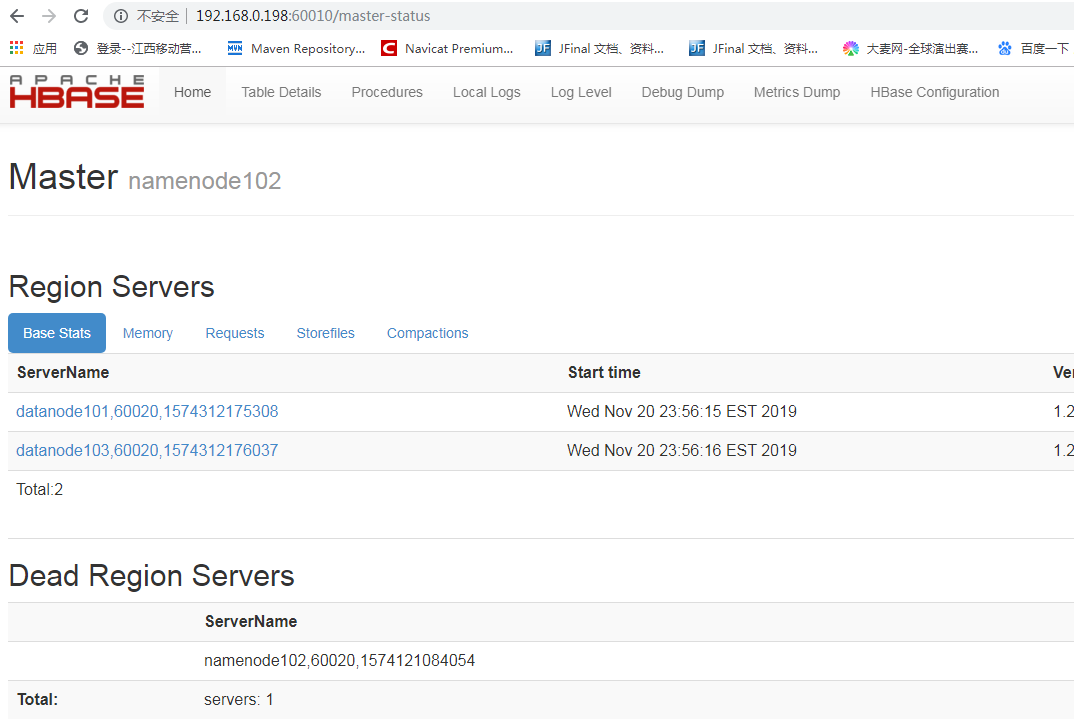
3台主机新增服务：HQuorumPeer



管理电脑还会新增服务：HMaster



查看hbase网址：http://192.168.0.198:60010



#若集群时间不同步，各节点会自动关闭

在CentOS8.0中默认不再支持ntp软件包，时间同步将由chrony来实现，像我这种习惯了ntp同步时间的，一时难以去适应chrony。

本文将通过wlnmp提供的源，来安装ntp服务

添加wlnmp源

|  |  |
| --- | --- |
| 1 | rpm -ivh http://mirrors.wlnmp.com/centos/wlnmp-release-centos.noarch.rpm |

安装ntp服务

|  |  |
| --- | --- |
| 1 | yum install wntp |

时间同步

|  |  |
| --- | --- |
| 1 | ntpdate ntp1.aliyun.com |

Hive默认启动，但是若需要用第三方软件连接hive，需要启动metastore和hiveserver2服务，启动10000端口连接

Beeline是Hive新的命令行客户端工具。

HiveServer2 支持一个新的命令行Shell，称为Beeline，它是基于SQLLine CLI的JDBC客户端。

Beeline支持嵌入模式(embedded mode)和远程模式(remote mode)。在嵌入式模式下，运行嵌入式的Hive(类似Hive CLI)，而远程模式可以通过Thrift连接到独立的HiveServer2进程上

nohup hive --service metastore & #在 201 节点上运行 jps 应该会有RunJar 进程则启动成功

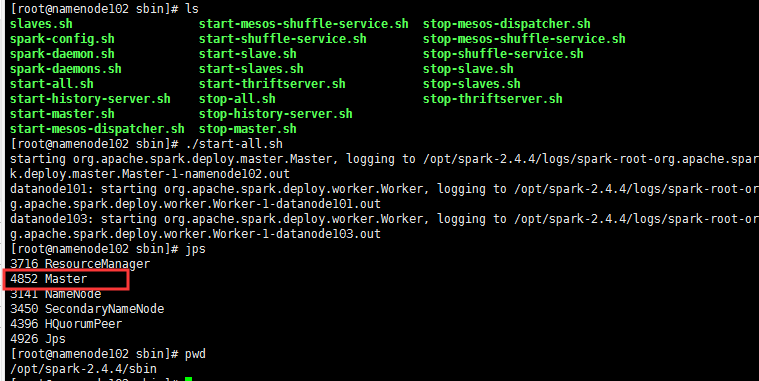
nohup hive --service hiveserver2& #在 201 节点上运行 jps 应该会有Master 进程则启动成功，端口是10000

6、Spark集群（暂时未安装）

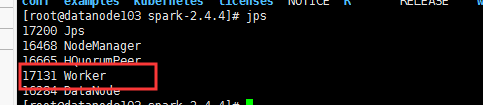
启动spark：/opt/spark-2.4.4/sbin/start-all.sh #在198上执行

停止spark：/opt/spark-2.4.4/sbin/stop-all.sh #在198上执行

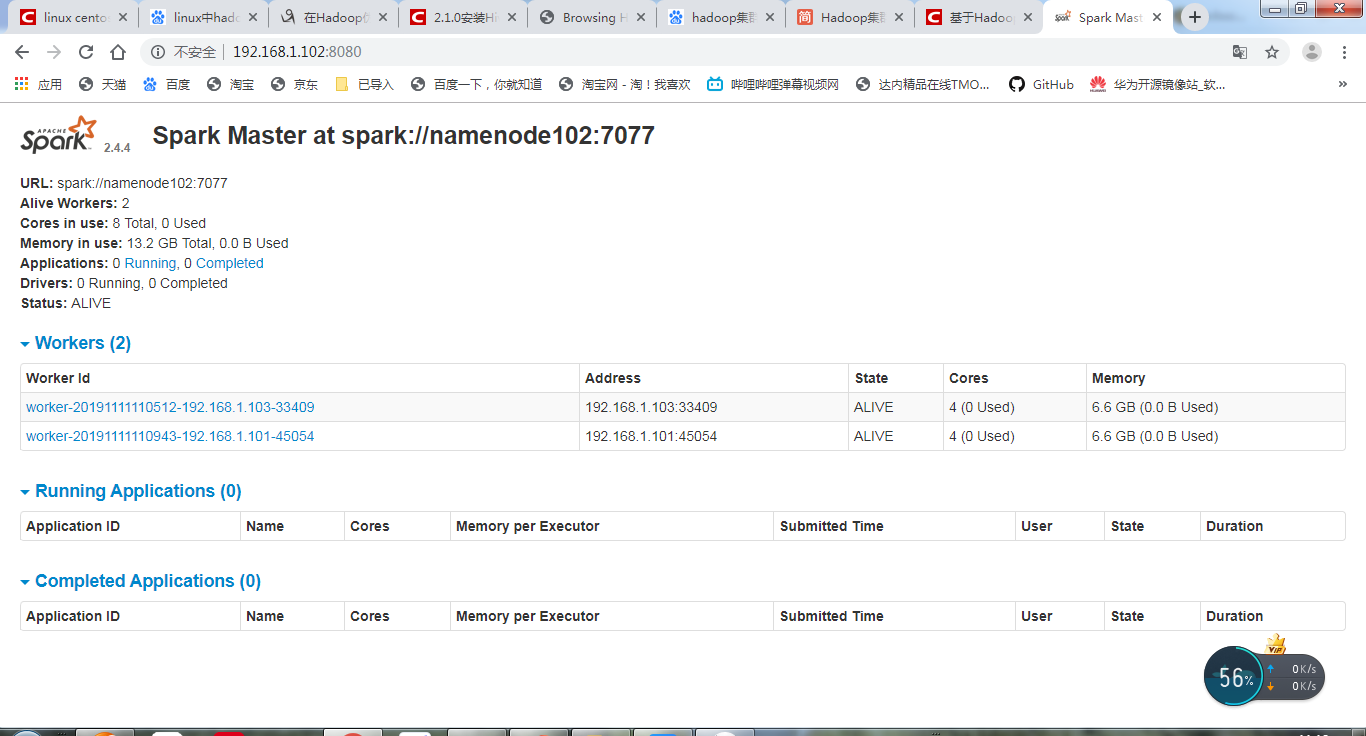
jps新增服务：Master

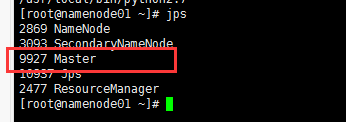


Datanode101、103新增服务：Worker



**Spark访问网址：**





**7、数据库Mysql-5.7.23管理（hive集群连的数据库为mysql，故mysql需确保启动）**

<https://www.cnblogs.com/lovling/p/7928674.html>

安装在**192.168.0.198**主机上

路径：/usr/local/mysql ,数据库保存路径:/usr/local/mysql/data

**启动 mysql ：**

**/usr/local/mysql/bin/mysqld\_safe --user=root & #默认使用/etc/my.cnf配置文件**

**(或者指定配置文件的方式启动 => ./bin/mysqld\_safe --defaults-file=/etc/my.cnf &)**

**关闭 mysql 使用该命令 ：**

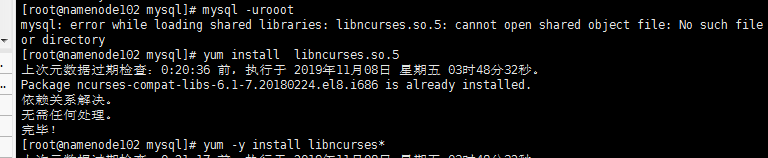
**/usr/local/mysql/bin/mysqladmin -uroot -p shutdown**

**或者用ss -anutlp | grep 3306找出pid号，再kill掉**

**如果是真实情况下, 一般需要创建 mysql 用户和 用户组, 来代替 root, 方便管理，直接使用 root 会引发安全问题，这里只是测试**

**测试连接 ：mysql -uroot -p**

**遇到问题**



解决办法，安装依赖库：yum install libncurses\*

数据库用户名和密码：root，123456

登录数据库：mysql -uroot -p123456