**大厂学苑-大数据&人工智能**

**学习环境准备**

版本：V1.0

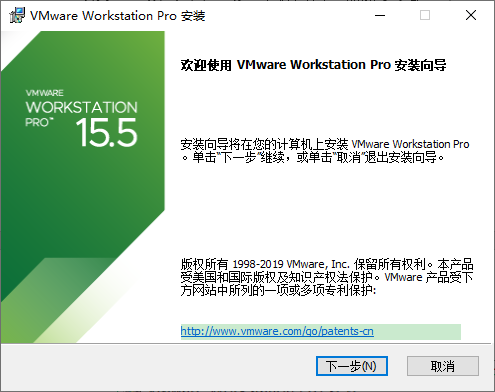


# VM与Linux的安装

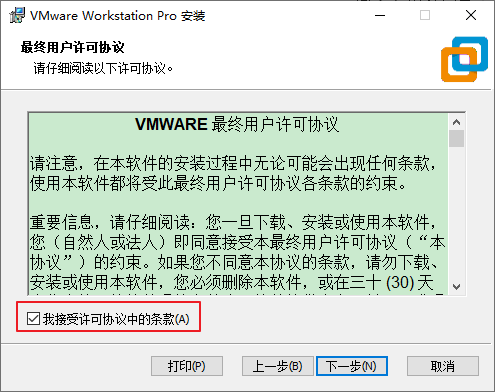
学习大数据，为了能够在本机模拟真实环境，需要在本机安装虚拟机软件(VMware)和虚拟系统（CentOS7）。相关软件在资料文件夹中，请大家下载后，进行安装。

## 1.1 VMWare安装

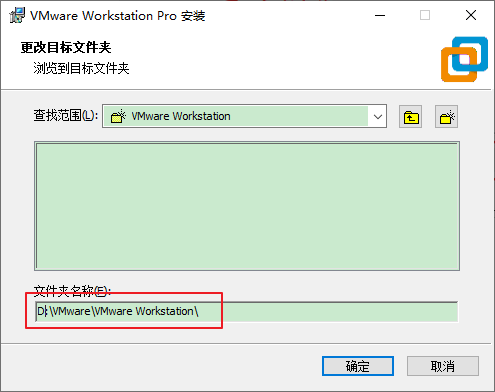
1）VMware Workstation Pro安装向导



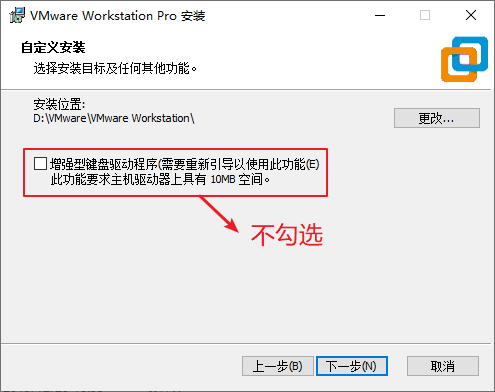
2）VMware Workstation安装的许可协议。



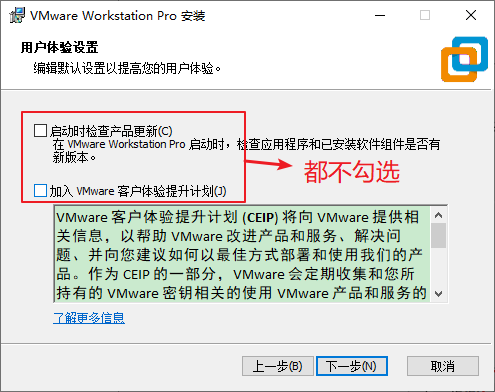
3）VMware Workstation安装路径。



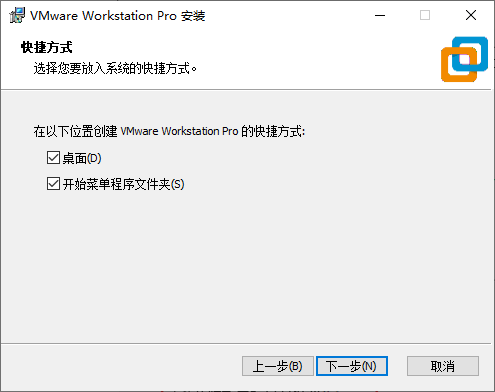
4）VMware Workstation增强型键盘功能。



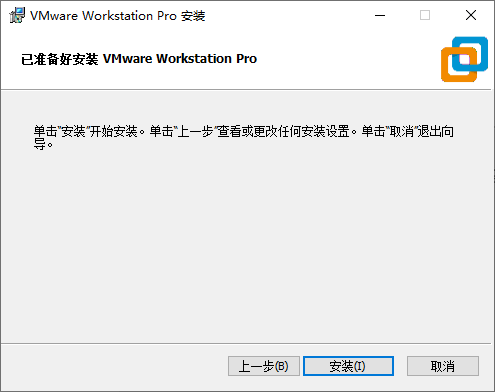
5）VMware Workstation软件检查更新和帮助完善



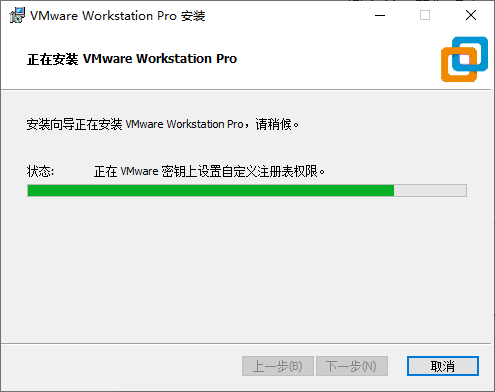
6）VMware Workstation快捷方式。



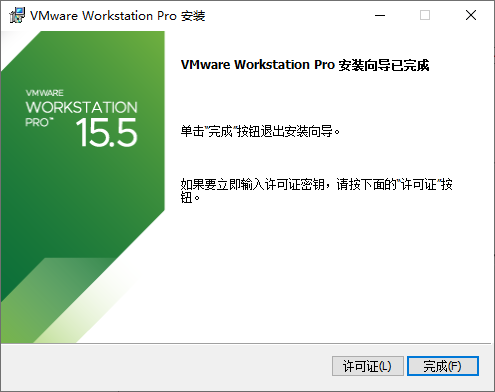
7）VMware Workstation 执行请求



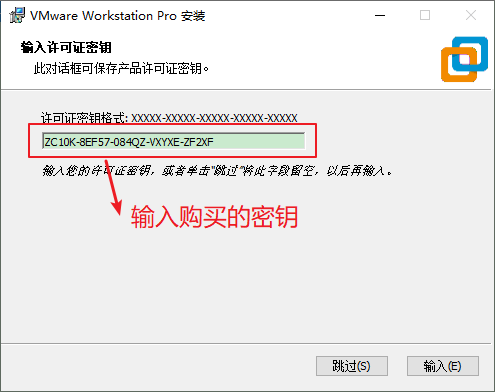
8）VMware Workstation 正在执行请求。



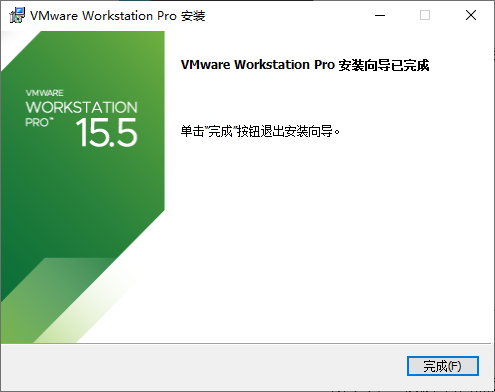
9）VMware Workstation 点击许可证。



10）VMware Workstation 输入许可证密钥。

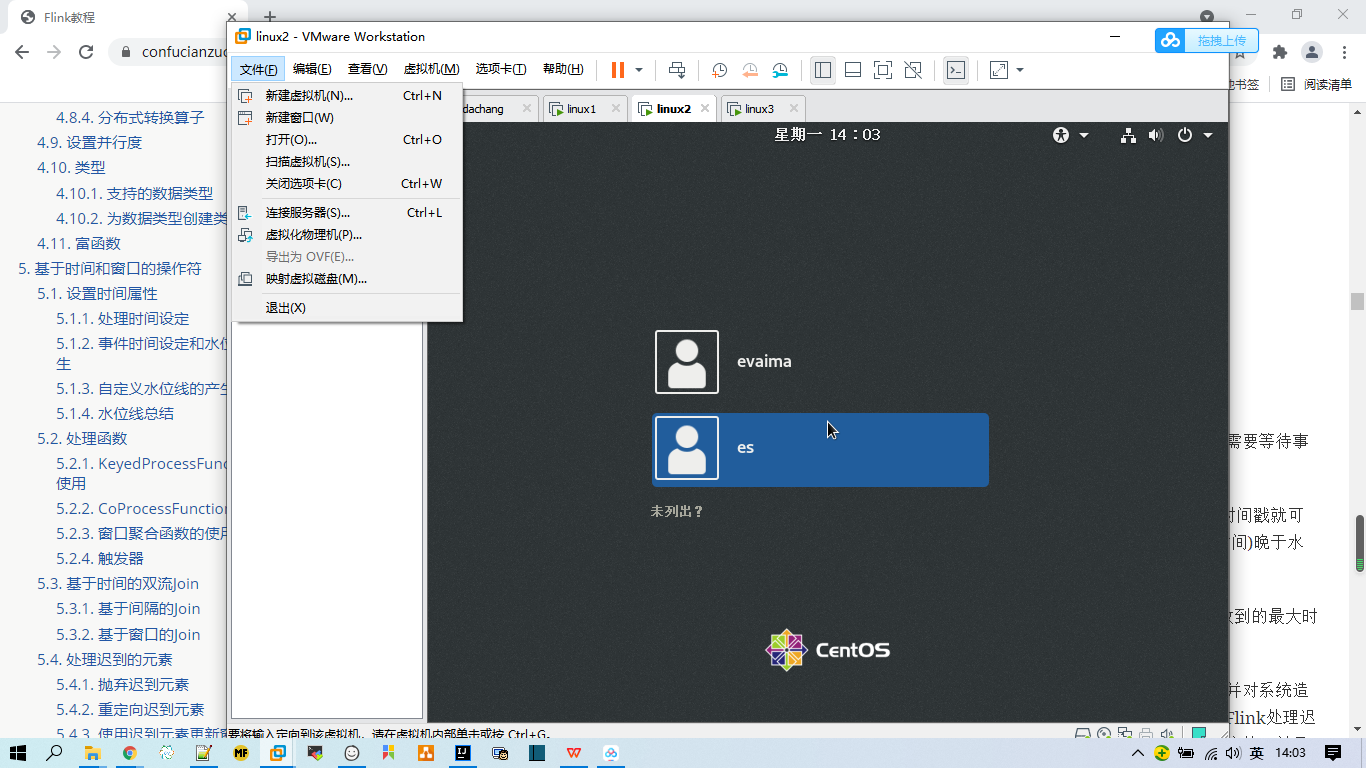


11）VMware Workstation 安装向导完成



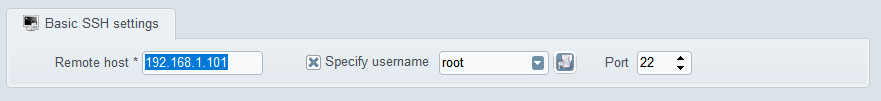
## 1.2 导入Linux系统

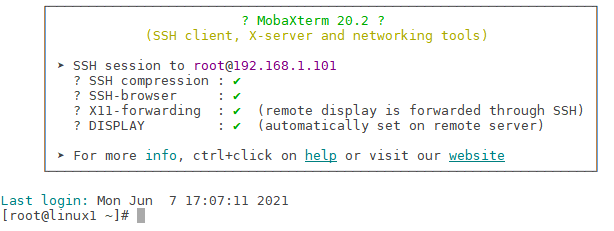
1）点击【文件/打开】菜单，选择资料文件目录中的3个虚拟机，依次导入到vmware即可。



2）导入成功后，启动虚拟机。使用账号root，密码root登录

3）使用任意远程工具访问虚拟机，能正常访问即可，资料文件夹中的MobaXterm\_Installer\_v20.2就是远程工具，也是讲师使用的工具。





# 第2章 基础环境测试

## 2.1 主机名

环境中的3台虚拟机已经配置好了对应的主机名，分别为

linux1

linux2

linux3

### 2.2.1 查看主机名

[root@linux1 ~]# hostname



### 2.2.2 修改主机名

* 编辑**/etc/hostname**文件

[root@linux1 ~]# vim /etc/hostname

如果不想使用当前的主机名，修改为我们想要修改的主机名即可

* 修改**etc/hosts**

[root@linux1 ~]# vim /etc/hosts

添加内容

192.168.1.101 linux1

192.168.1.102 linux2

192.168.1.103 linux3

## 2.2 IP地址

环境中的3台虚拟机已经配置好了对应的IP地址，分别为：

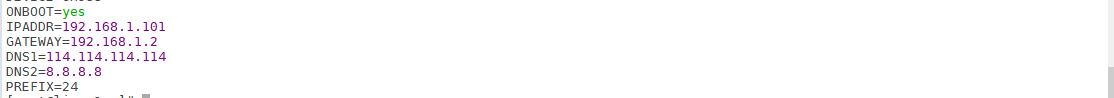
192.168.1.101 linux1

192.168.1.102 linux2

192.168.1.103 linux3

修改IP地址

[root@linux1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-ens33



## 2.3 防火墙

虚拟机中以及配置好了防火墙

# 关闭防火墙的自动启动

[root@linux1 ~]#systemctl disable firewalld

# 开启防火墙的自动启动

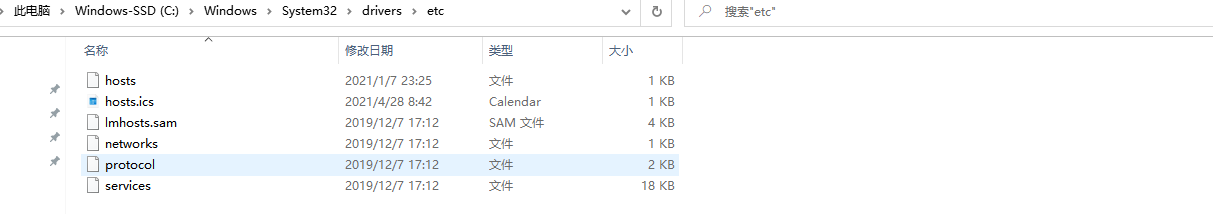
[root@linux1 ~]#systemctl enable firewalld

# 查看防火墙状态

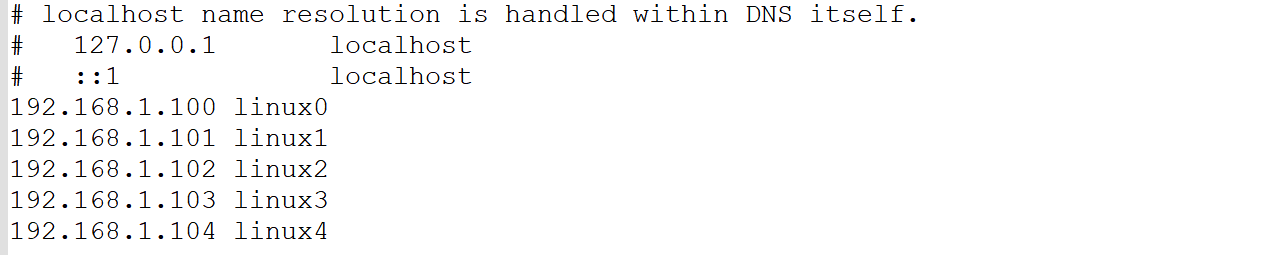
[root@linux1~]#systemctl is-enabled firewalld

## 2.4 HOST

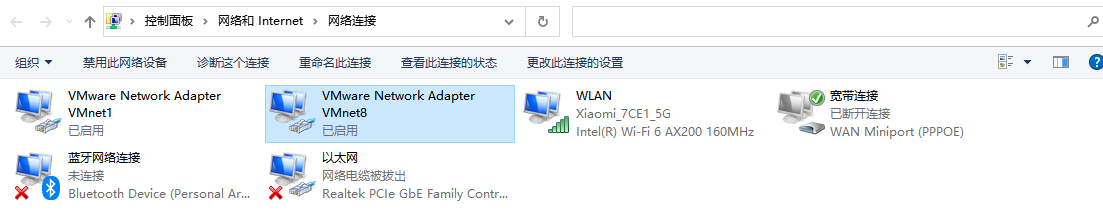
环境中的3台虚拟机已经配置好了对应的IP地址和主机名，为了能够在Windows环境中能正确访问，需要配置Windows的域名解析

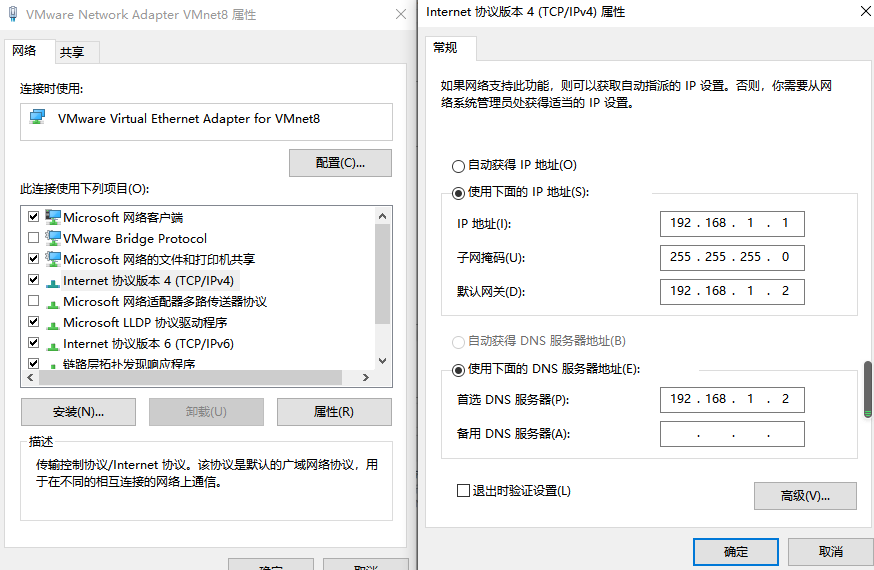


修改hosts文件



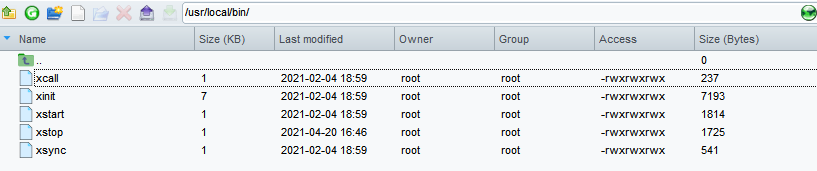
修改本机VMware的Ipv4配置，配置如下图：



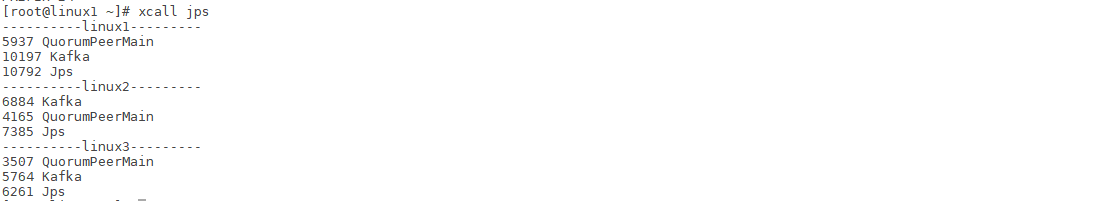


## 2.5 基础脚本

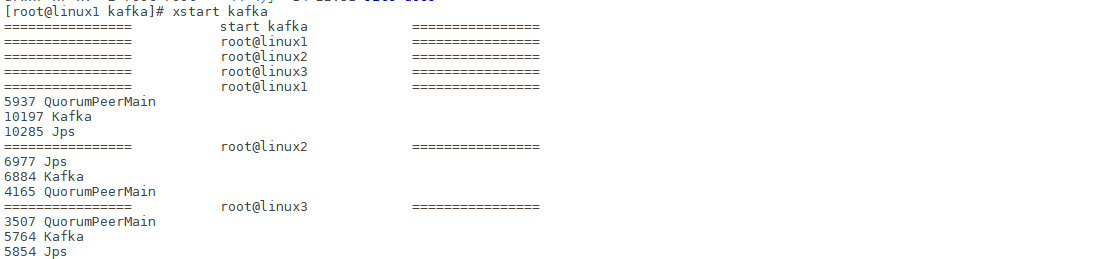
虚拟机中有很多软件和命令需要执行，每一次都要输入大量的指令，非常的麻烦，在掌握了基本的语法后，可以通过脚本文件直接执行相关的命令，为此，系统中封装了一些简单的脚本文件



**xcall** 用于执行全局系统指令：**xcall jps**



**xstart, xstop** 用于执行全局软件启动，关闭指令：**xstart zk,kafka**



**xsync**用于执行文件同步指令：**xsync kafka**

## 2.6 SSH无密登录配置

虚拟机之间想要互相访问，可以通过ssh命令进行操作，但是需要输入密码，比较麻烦，所以咱们的3台虚拟机之间已经配置了无密登录

### 2.6.1 ssh连接另外一台电脑

ssh linux2

### 2.6.2 无密登录

* 生成公钥和私钥：

ssh-keygen -t rsa

然后敲（三个回车），就会生成两个文件id\_rsa（私钥）、id\_rsa.pub（公钥）

* 将公钥拷贝到要免密登录的目标机器上

ssh-copy-id linux1

ssh-copy-id linux2

ssh-copy-id linux3

注意：

还需要在linux2上采用root账号，配置一下无密登录到linux1、linux2、linux3

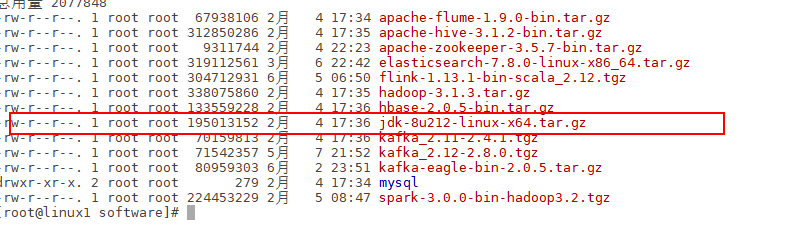
还需要在linux3上采用root账号，配置一下无密登录到linux1、linux2、linux3

# 第3章 软件环境测试

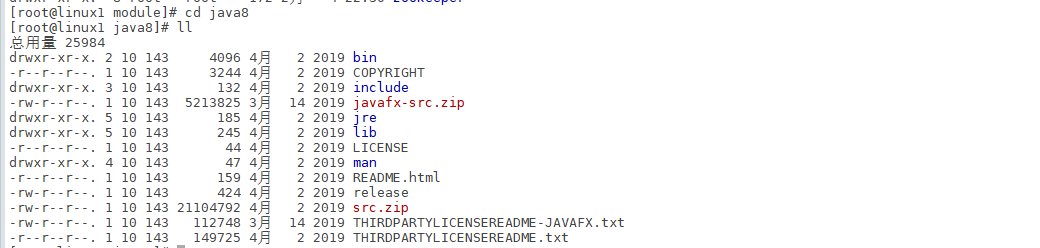
在课程中，需要使用很多的大数据软件，为了更好的学习效果和教学体验，课程中需要的一些基础软件已经事先在虚拟机中安装好了。在使用前，需要对已经安装的软件进行测试。

## 3.1 Java

Java软件是整个大数据软件的语言基础。软件在linux系统的的/opt/module/software中。



软件安装目录在**/opt/module/java8**中



环境变量配置在**/etc/profile.d/my\_env.sh**中

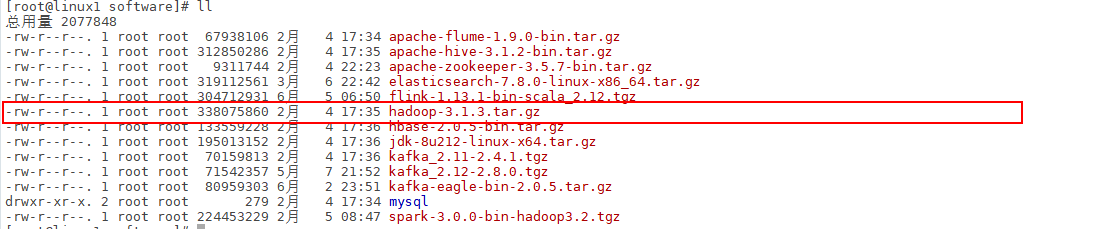


命令行中输入指令：**java -version**

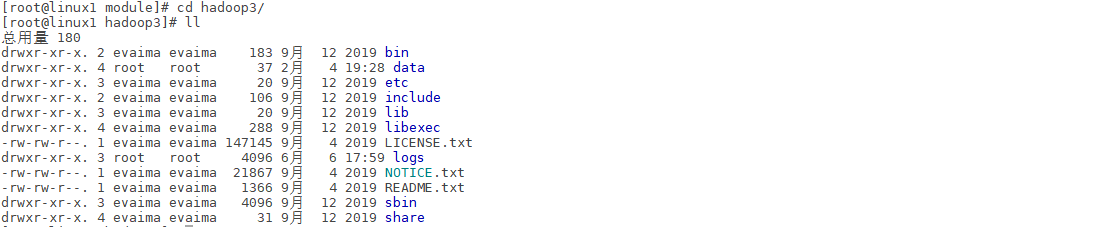


## 3.2 Hadoop

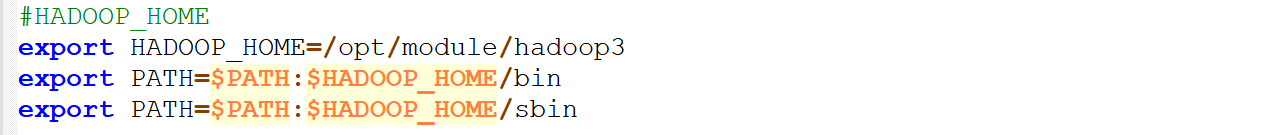
Hadoop框架为大数据的基础框架，在当前环境中必须安装。软件在linux系统的的**/opt/module/software**中



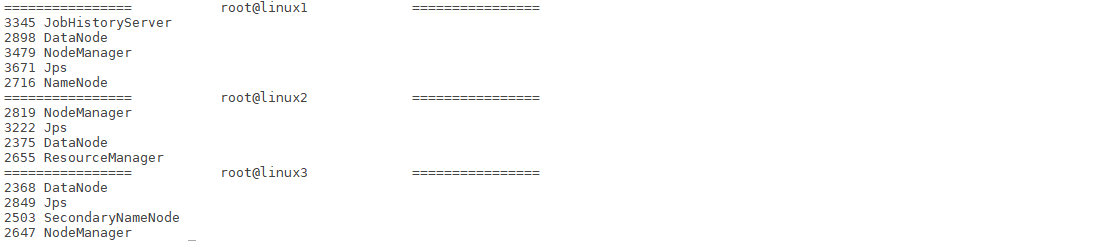
软件安装目录在**/opt/module/hadoop3**中



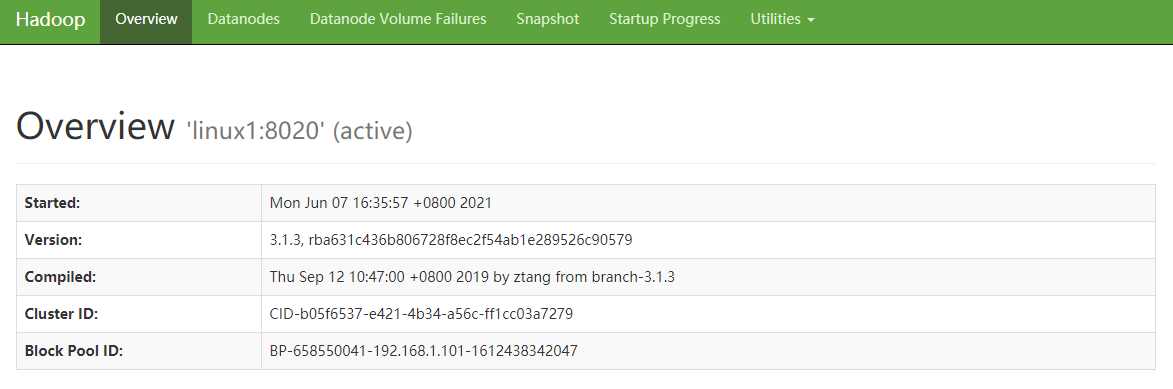
环境变量配置在**/etc/profile.d/my\_env.sh**中



使用群起脚本命令**启动**hadoop软件 ：**xstart hadoop**



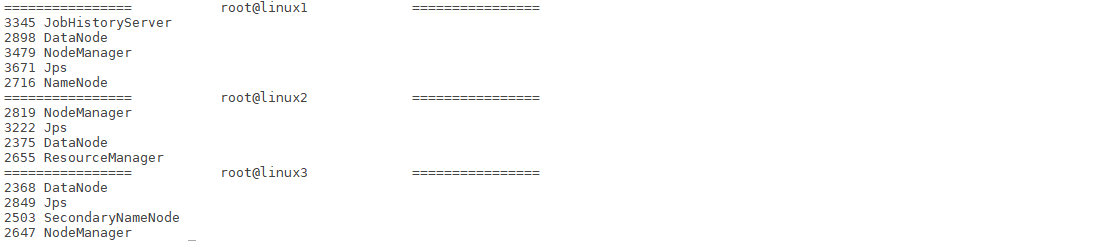
在浏览器中输入网址查看UI页面：<http://linux1:9870>



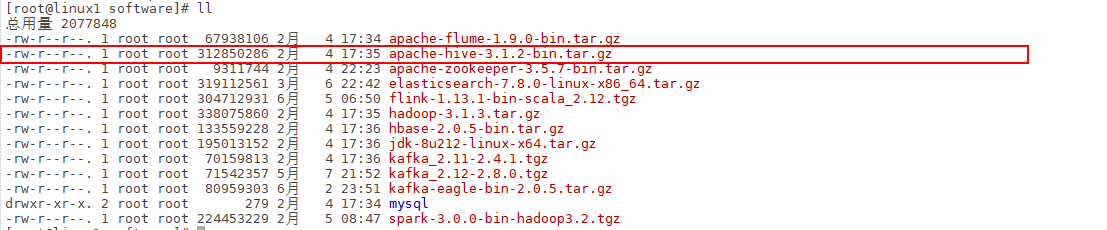
使用群起脚本命令**关闭**hadoop软件 ：**xstop hadoop**

## 3.3 Hive

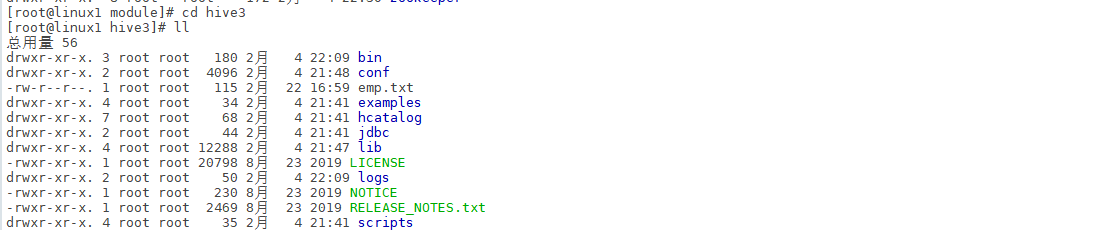
Hive软件为基于Hadoop软件的SQL操作工具软件，所以使用前，必须保证hadoop可以正确启动



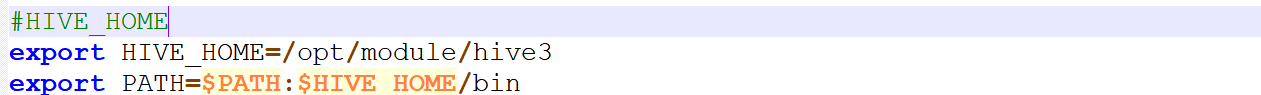
软件在linux系统的的**/opt/module/software**中



软件安装目录在**/opt/module/hive3**中



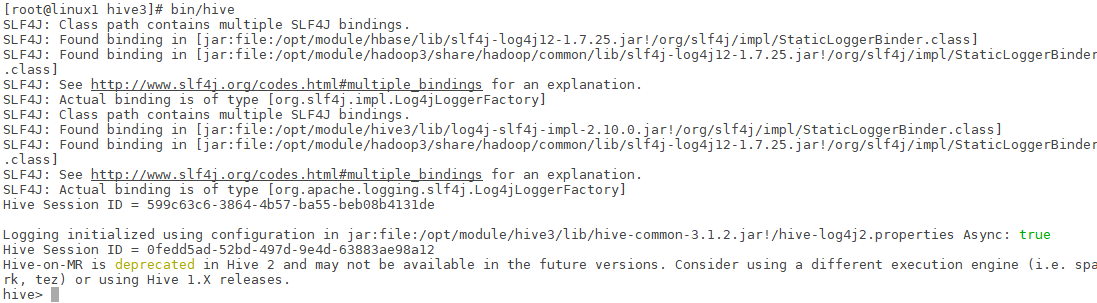
环境变量配置在**/etc/profile.d/my\_env.sh**中



使用命令行**启动**Hive：**hiverserverices.sh start**



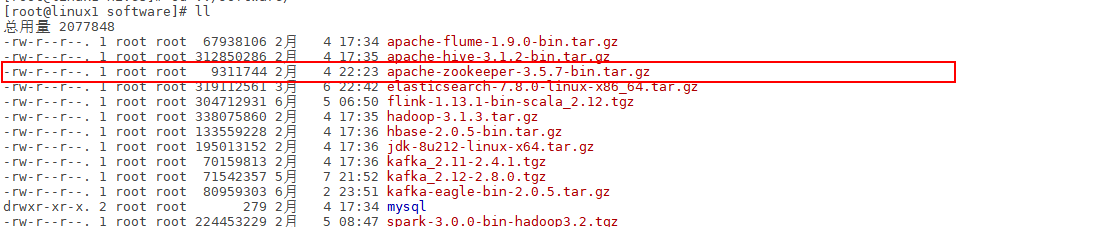
启动后，执行命令：**bin/hive**进入Hive中查询数据 : **show tables;**



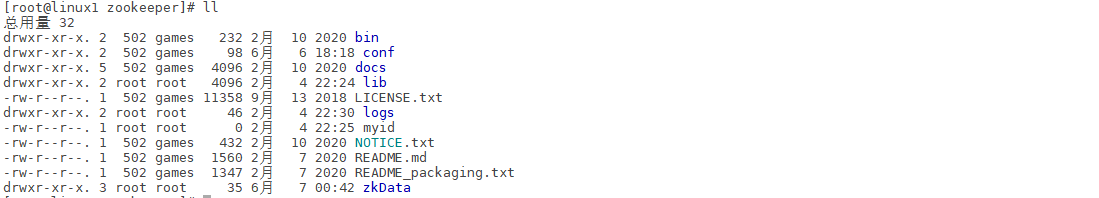
使用命令行**关闭**Hive：**hiverserverices.sh stop**

## 3.4 Zookeeper

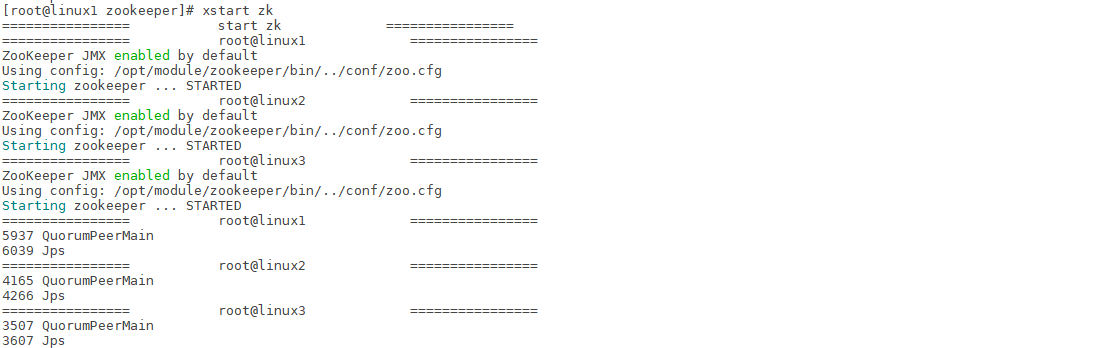
软件在linux系统的的**/opt/module/software**中



软件安装目录在**/opt/module/zookeeper**中



使用群起脚本命令**启动**zookeeper软件 ：**xstart zk**

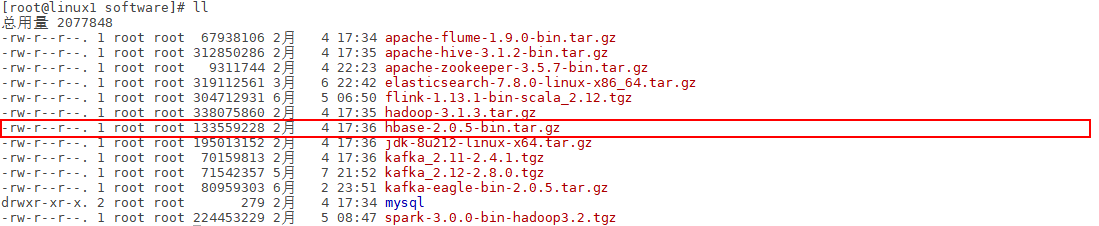


使用群起脚本命令**关闭**zookeeper软件 ：**xstop zk**

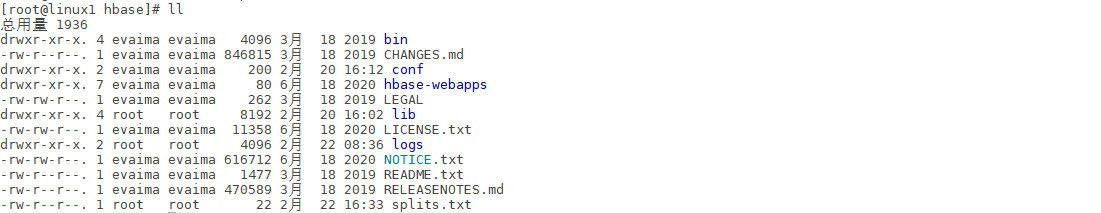
## 3.5 HBase

HBase软件为基于Hadoop软件的海量数据存储软件，所以使用前，必须保证hadoop可以正确启动。并且HBase软件基于zookeeper进行集群管理，所以也必须保证zookeeper正常启动。

软件在linux系统的的**/opt/module/software**中



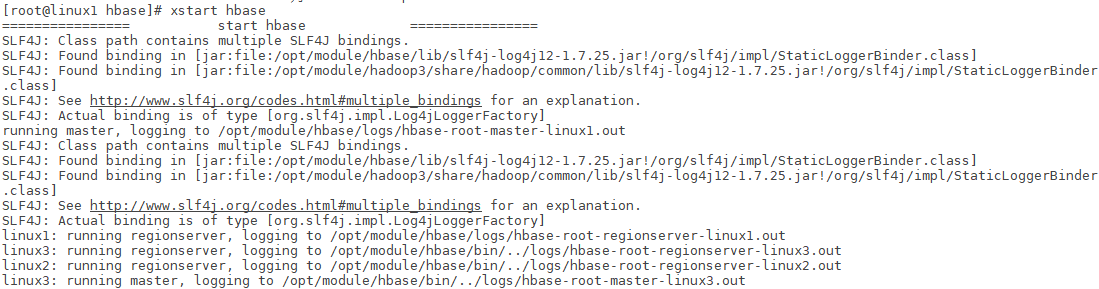
软件安装目录在**/opt/module/hbase**中



环境变量配置在**/etc/profile.d/my\_env.sh**中

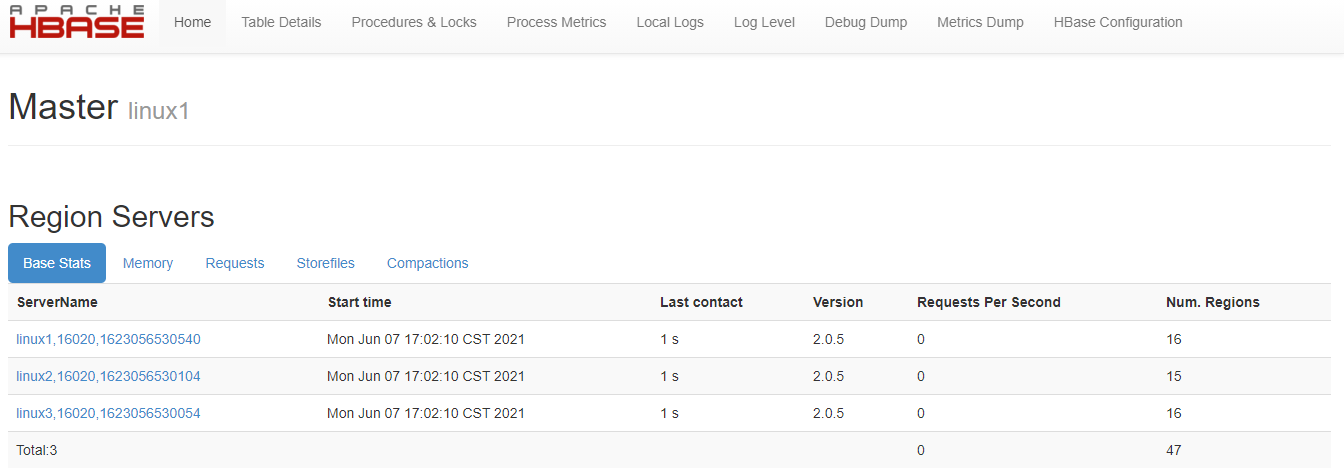


使用群起脚本命令**启动**hbase软件 ：**xstart hbase**





在浏览器中输入网址查看UI页面：[http://linux1:1](http://linux1:9870)6010



使用群起脚本命令**关闭**hbase软件 ：**xstop hbase**

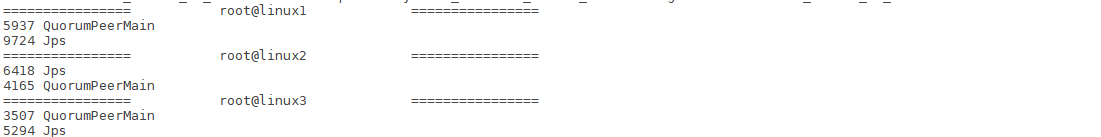
## 3.6 MySQL

mysql已经作为系统服务，系统启动后可以直接访问。

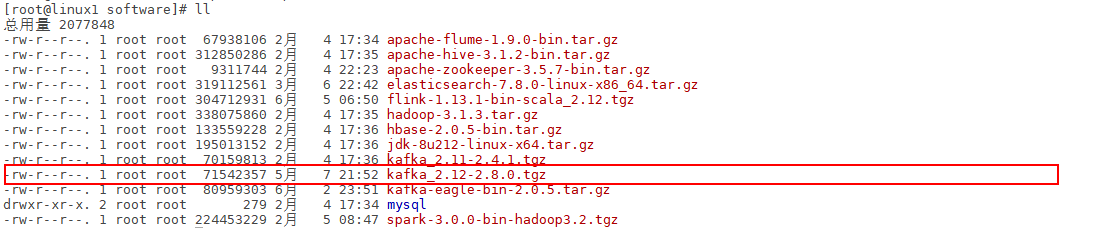


## 3.7 Kafka

Kafka软件基于zookeeper进行集群管理，所以也必须保证zookeeper正常启动。



软件在linux系统的的**/opt/module/software**中



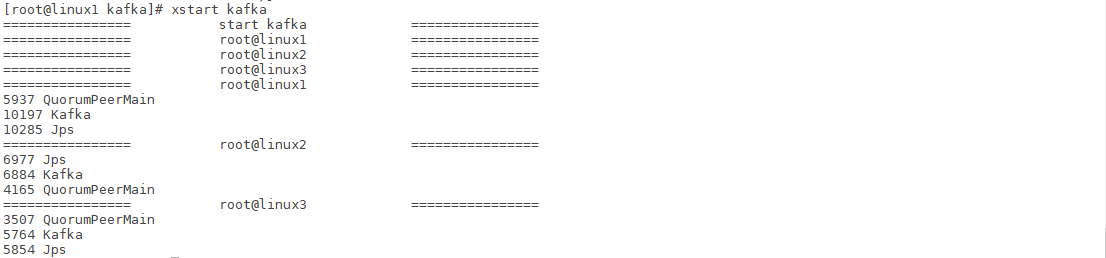
软件安装目录在**/opt/module/kafka**中



环境变量配置在**/etc/profile.d/my\_env.sh**中



使用群起脚本命令**启动kafka**软件 ：**xstart kafka**



使用群起脚本命令**关闭kafka**软件 ：**xstop kafka**