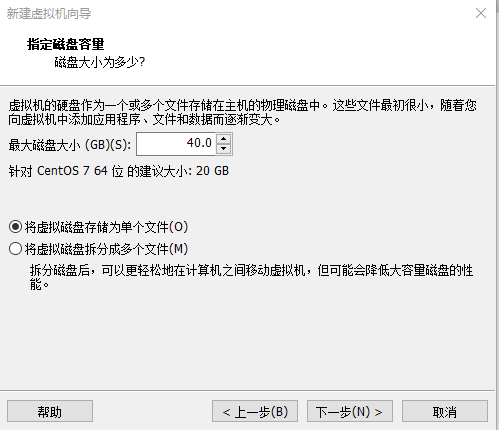
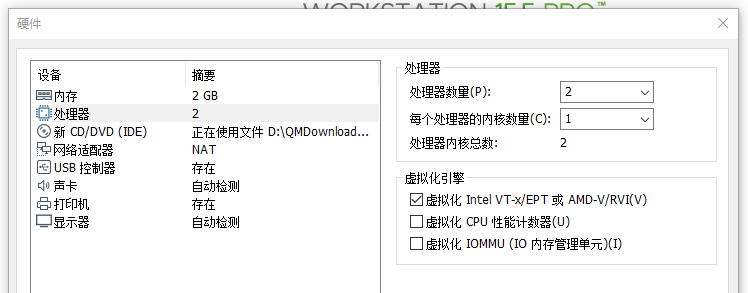
大数据第二周

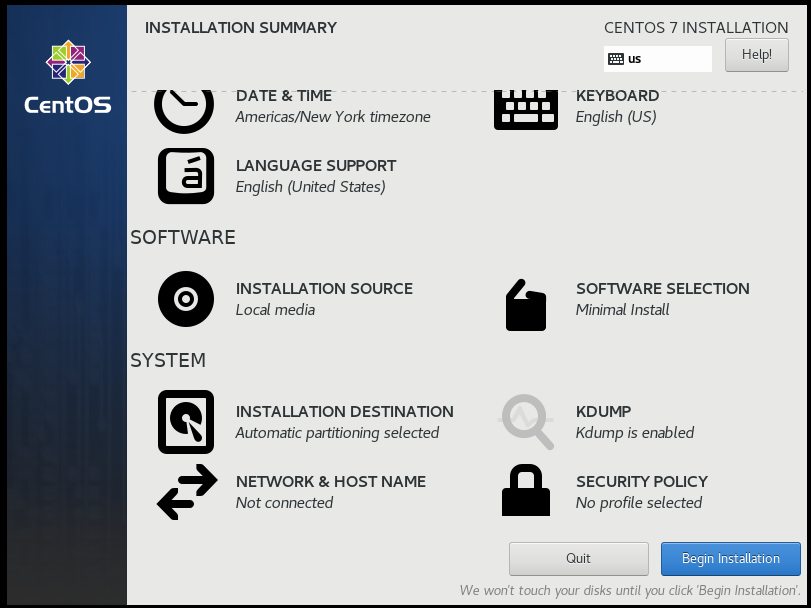
1. 安装CentOS7最小安装版



硬盘容量不小于30G。



自定义硬件中，勾选“虚拟化Intel VT-x/EPT或AMD-V/RVI(V) ”。



在此界面中，“SOFTWARE SELECTION”保持默认，因为默认就是最小安装。

安装完成后，配置网络。

进入/etc/sysconfig/network-scripts目录，使用命令

cd /etc/sysconfig/network-scripts

使用ls命令查看与网卡号相关的文件。使用ip a命令查看网卡号。

Vi编辑器的使用：有三种模式，刚进入时是一般命令模式，还有编辑模式和底行命令模式。

一般命令模式下，按i或s键进入编辑模式，按ESC进入一般命令模式。在一般命令模式下，按冒号键进入底行命令模式。

在相应目录下，执行sudo vi ifcfg-ens33，进入编辑页面，按i键，进入编辑模式，把“ONBOOT=no”改成“ONBOOT=yes”，然后按ESC进入一般命令模式，再按冒号，进入底行命令模式，输入wq，回车退出。重启虚机：sudo reboot。

ONBOOT=no表示启动时不激活，改成yes后，就变成启动时激活。

重启后用ping命令检查是否能够上网，例如：ping [www.zufe.edu.cn](http://www.zufe.edu.cn)。

Vmware虚拟机网络配置：

打开编辑菜单里的“虚拟网络编辑器”，正常情况是这样的：



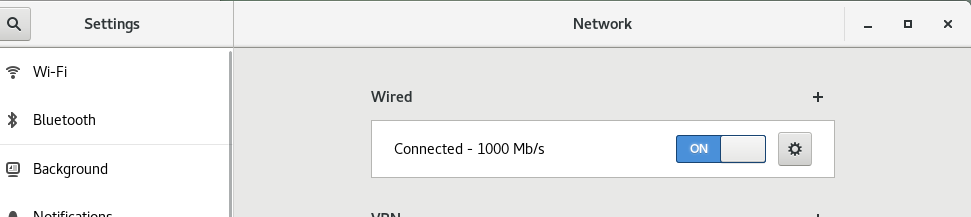
如果异常，确定是否以管理员身份启动的vmware，如果是，点击“还原默认设置”。

如果要在NAT模式修改配置IP地址，选中NAT模式：

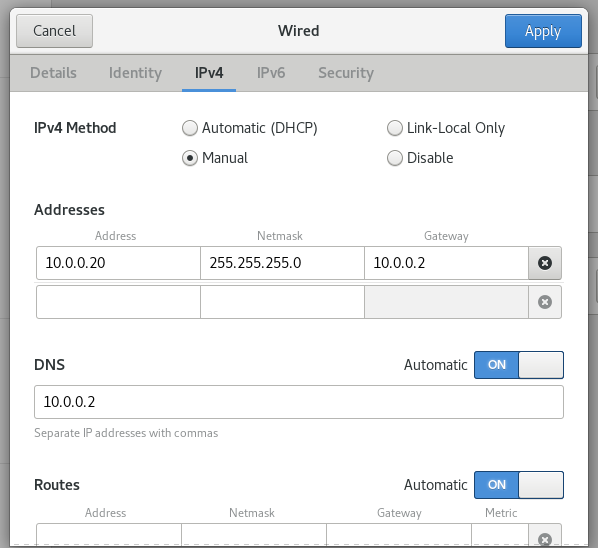


在子网IP栏内输入自己想要的网络地址。

设置静态ip（在GUI下），打开如下界面：



点击齿轮，进入设置页面，选中IPV4：



按图示输入，点击“Apply”，回到上上图，关闭网络再打开，ip地址就生效了。

打开网络配置文件，内容如下：

TYPE=Ethernet

PROXY\_METHOD=none

BROWSER\_ONLY=no

BOOTPROTO=none

DEFROUTE=yes

IPV4\_FAILURE\_FATAL=no

IPV6INIT=yes

IPV6\_AUTOCONF=yes

IPV6\_DEFROUTE=yes

IPV6\_FAILURE\_FATAL=no

IPV6\_ADDR\_GEN\_MODE=stable-privacy

NAME=ens33

UUID=485a313b-0242-4d61-b54f-4ce88077a7fd

DEVICE=ens33

ONBOOT=yes

IPADDR=10.0.0.20

PREFIX=24

GATEWAY=10.0.0.2

DNS1=10.0.0.2

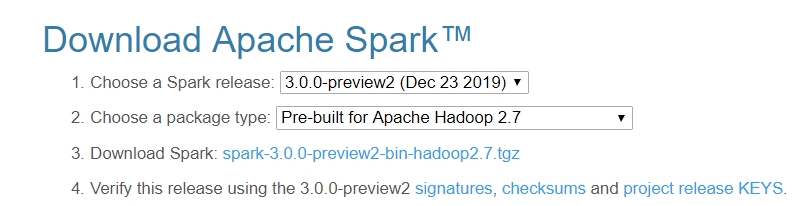
观察文件内容，红色部分是有变化的。

Namenode的IP地址是10.0.0.20

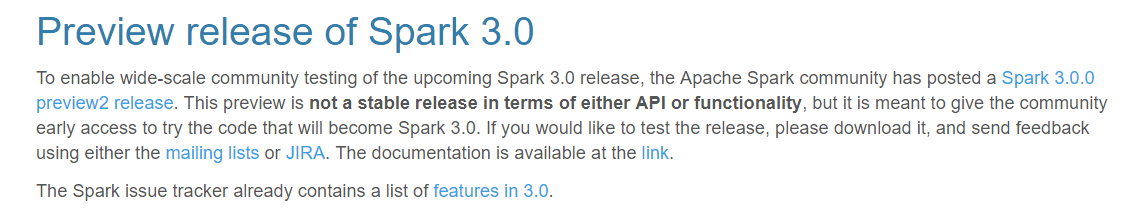
如果要设置datanode的ip地址是10.0.0.21，如何作？

1. 确定Spark、Hadoop、Java版本

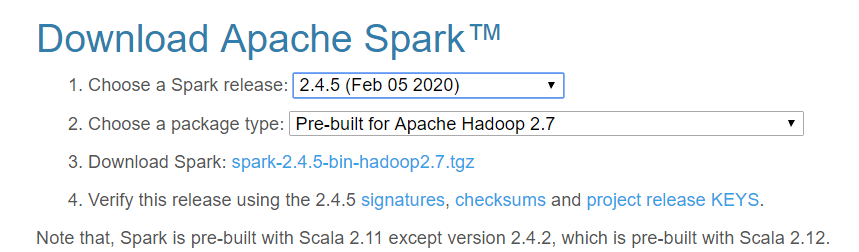
进入Spark官网：<http://spark.apache.org/>，官网首页没有版本信息，根据经验，点击“download”，进入下载页面：<http://spark.apache.org/downloads.html>，页面信息如下：



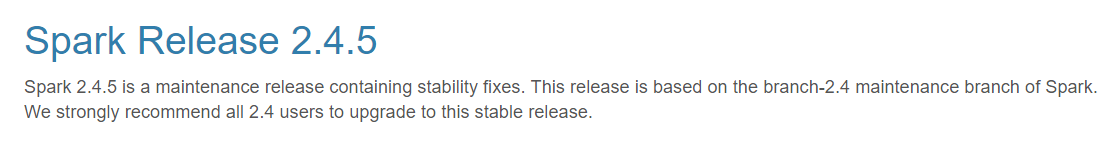
从此页面上看，似乎是应该下载Spark3.0.0，但此版本标识preview2，应该不是稳定版，作为初次学习之用，必须安装稳定版。需要确定此信息。在此页面：<http://spark.apache.org/news/spark-3.0.0-preview2.html>，



确定了Spark3.0.0不是稳定版。回到下载页面，选择其他版本：Spark2.4.5



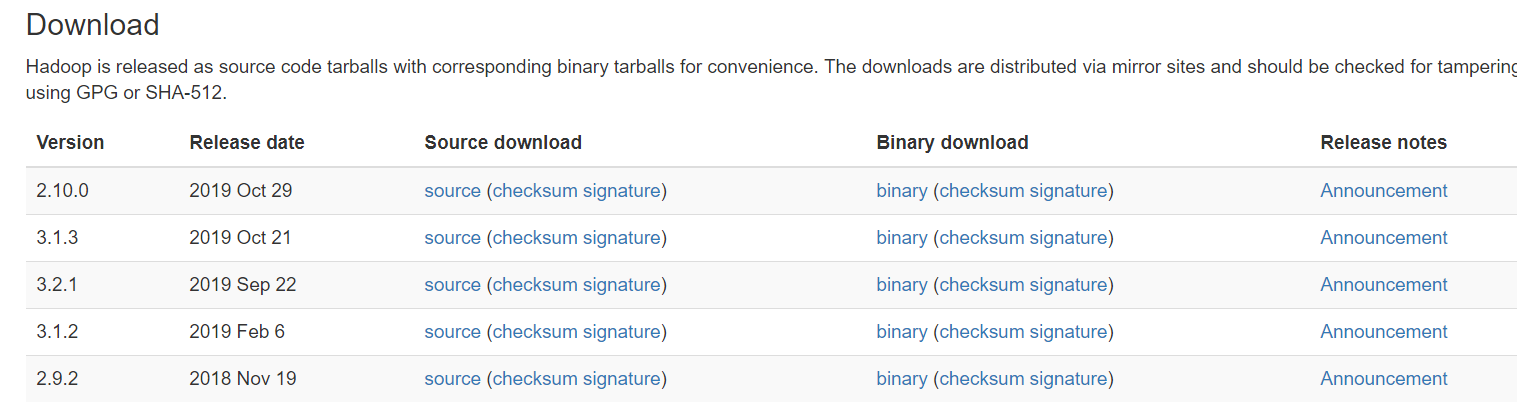
从此页面上可以看到Spark2.4.5，对应Hadoop2.7。下面核实Spark2.4.5是否稳定版。点击页面下部的Spark2.4.5的链接，进入http://spark.apache.org/releases/spark-release-2-4-5.html：



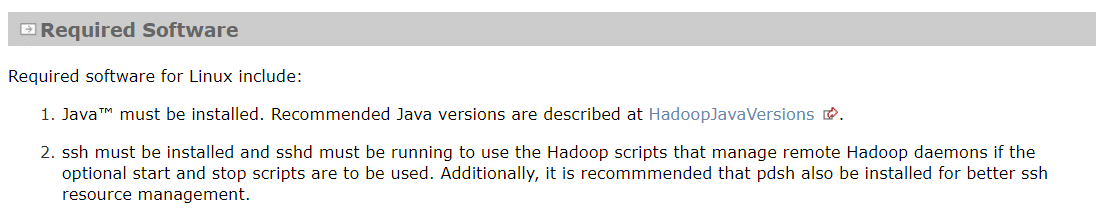
页面上明确Spark2.4.5是正在维护的稳定版。

结论：选择Spark2.4.5，Hadoop2.7。

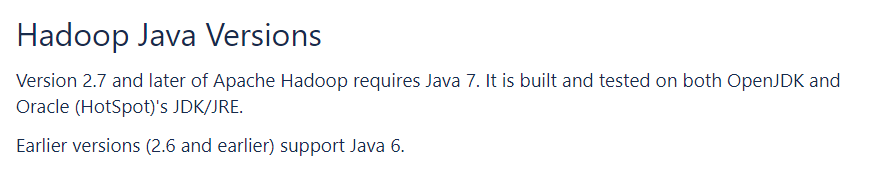
进入Hadoop主页：hadoop.apache.org，按经验，点击“download”按钮进入下载页面。



下载页面中没有我们需要的2.7版本，也没有java相关信息。回到主页。点击“getting started”：



在此项目上有java版本说明，点击进入说明



从页面说明上，需要java7。

最终结论：Spark2.4.5，Hadoop2.7，java7。

下载Hadoop2.7，由于下载页面上没有此版本，打开页面上的“mirror site”链接，里面有2.7版本下载。