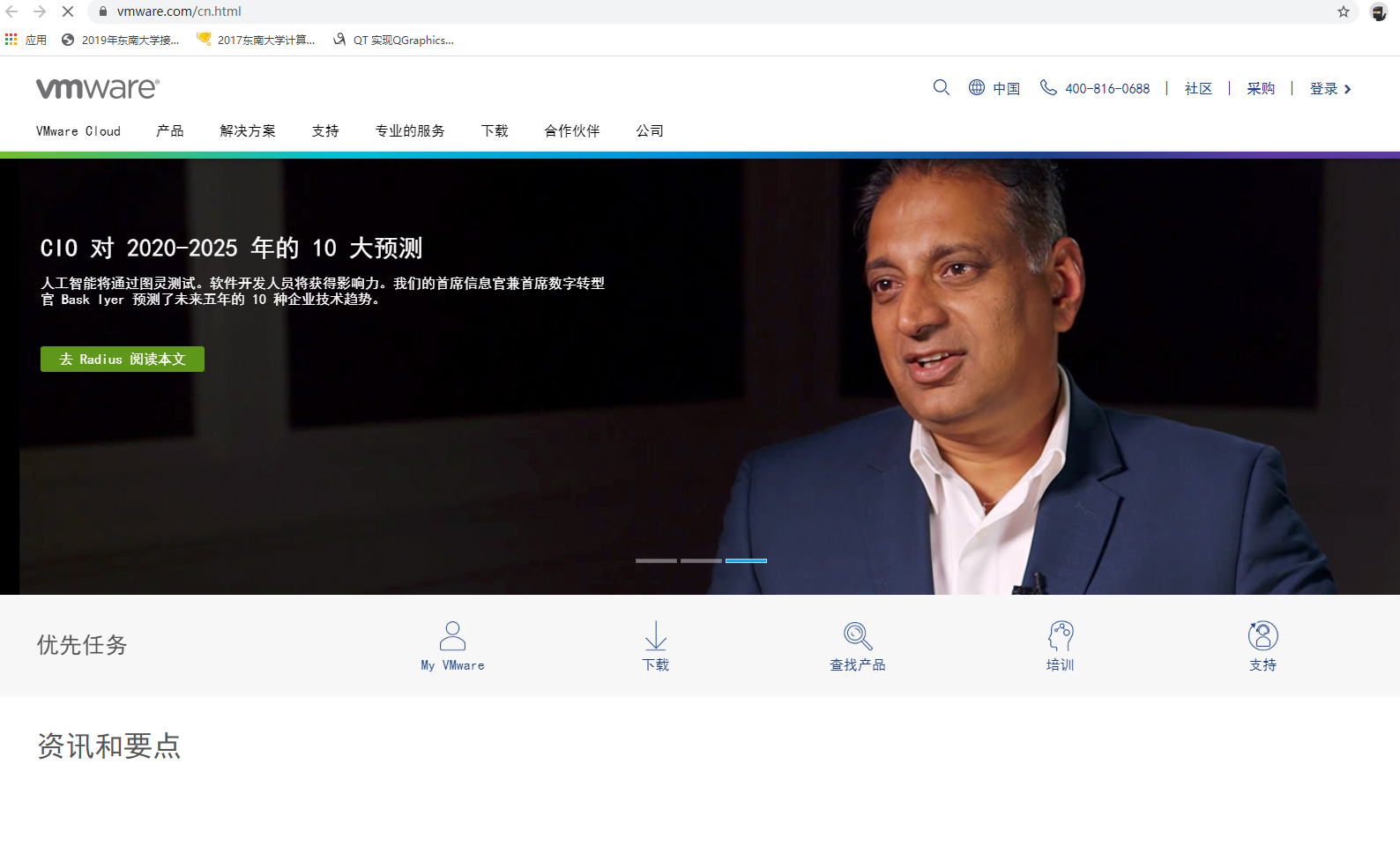
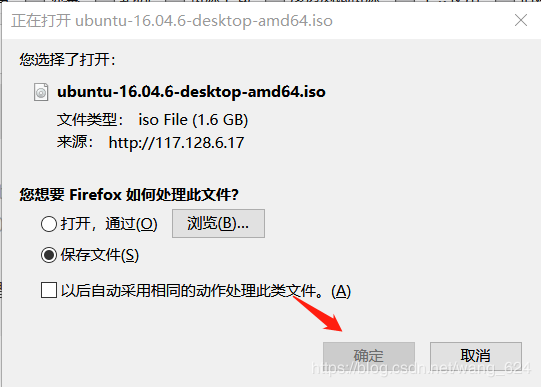
在官网上下载VMWare软件。点击下载按钮，等待下载完成。

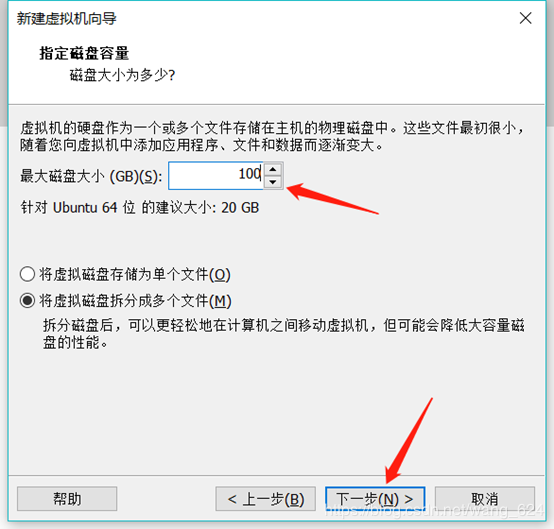


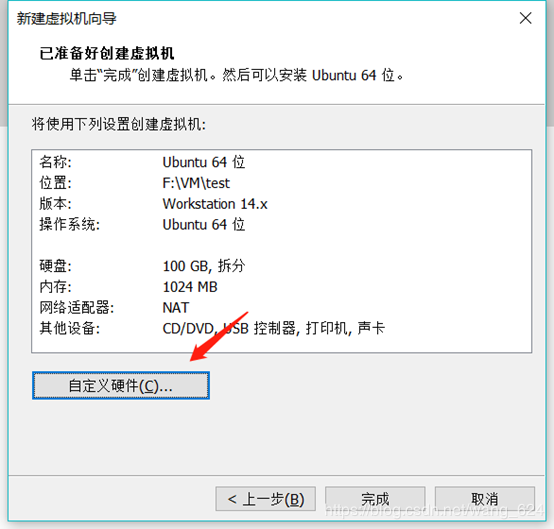
1. 虚拟机环境中安装Linux操作系统：
2. 下载镜像：选择64位的（根据自己的系统选择）

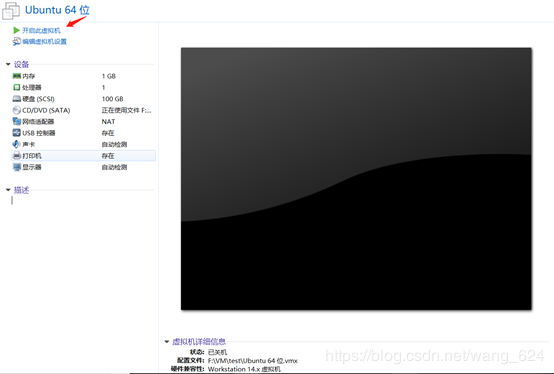


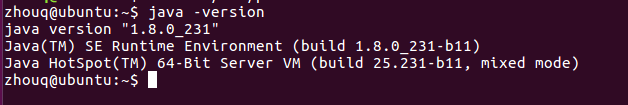
1. 点击之后就会出现下载，文件较大，需要等待一会儿



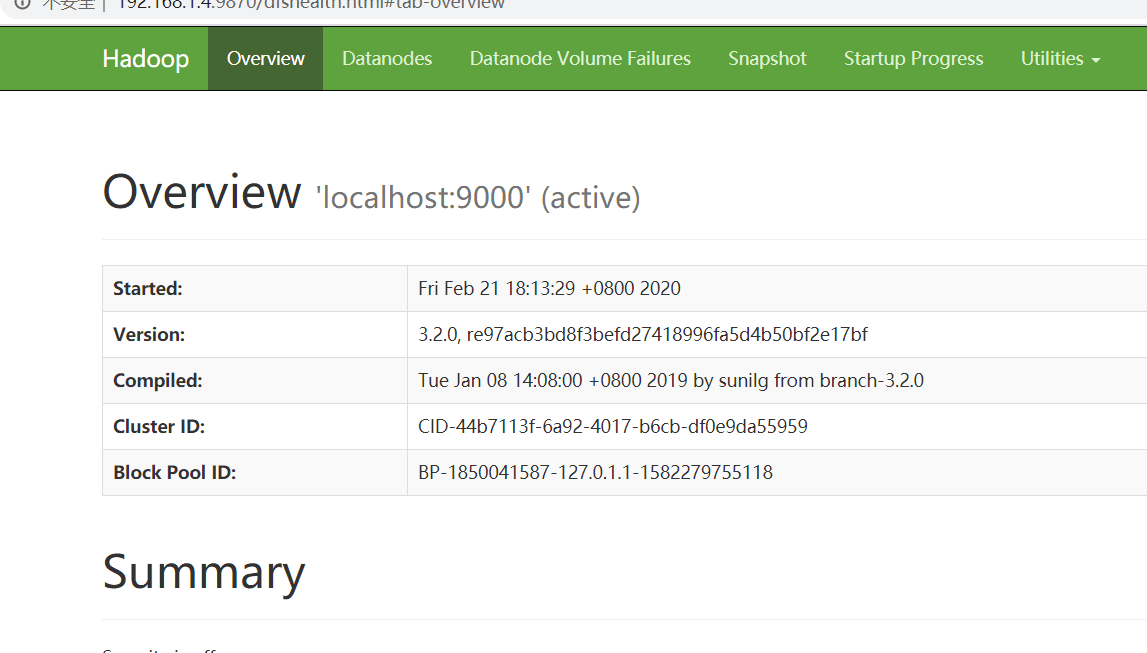
1. 安装过程：
2. 创建新的虚拟机。
3. 选择Linux的ubuntu64位（根据自己的系统情况进行选择）。
4. 选择虚拟机存放位置。
5. 指定磁盘的大小，这里的大小不会立即把你的物理机的磁盘容量霸占
6. 自定义硬件，修改设置



1. 修改内存和处理器，这个看大家的用途而定，重点是选择iso的镜像文件，将之前安装好的镜像文件选择就好了。
2. 最后开始安装。
3. 开始安装虚拟机系统，首先点击开始
4. 选择英文，点击install now。
5. 后面直接都是继续，确定默认。
6. 设置你的用户名和密码。
7. 最后等待安装完毕，然后重启。
8. 在Linux系统下安装jdk并配置环境变量
9. 安装环境：CentOS 07 64位 jdk版本：1.8 工具：Xshell6、Putty、WinSCP
10. 安装步骤：
11. 下载Linux环境下的jdk1.8安装包
12. 解压安装包
13. 解压命令进行解压
14. 解压完成后，可以在当前目录下看到一个名字为【jdk1.8.0\_131】的目录，里面存放的是相关文件，我们要将解压后的【jdk1.8.0-131】里面的所有数据移动到我们需要安装的文件夹当中，我们打算将jdk安装在usr/java当中，我们在usr目录下新建一个java文件夹。将【jdk1.8.0\_131】里的数据拷贝到java目录下，将解压后的文件移动到此目录下。
15. 编辑配置文件，修改环境变量
16. 测试一下是否安装成功了。



1. 安装Hadoop，实现伪分布式配置：
2. 创建hadoop用户，如果你安装Ubuntu的时候不是用的“hadoop”用户，那么需要增加一个名为hadoop的用户。
3. 更新apt。
4. 安装SSH、配置SSH无密码登陆。
5. 安装Java环境
6. 安装Hadoop2。
7. Hadoop伪分布式配置。Hadoop可以在单节点上以伪分布式的方式运行，节点既作为NameNode也作为DataNode，同时，读取的是HDFS中的文件。



1. 执行wordcount程序，统计指定目录下的词频：
2. 启动Hadoop

切换到hadoop安装目录下的sbin目录下执行./start-all.sh命令或执行./start-dfs-sh和./start-yarn.sh两条命令

1. 在集群中创建目录inputdata\_w。
2. 将Hadoop安装目录下的LICENSE.txt、README.txt、NOTICE.txt文件上传到集群。
3. 试用hadoop-mapreduce-examples-2.7.6.jar对上传的数据进行词频统计



1. 查看统计结果

