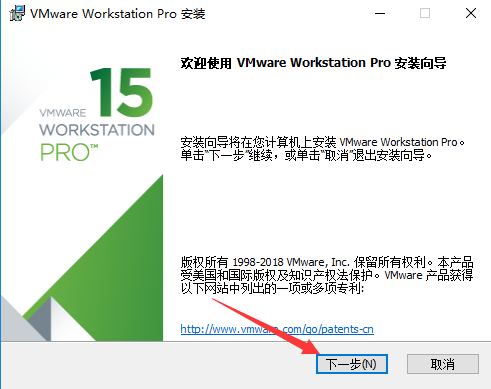
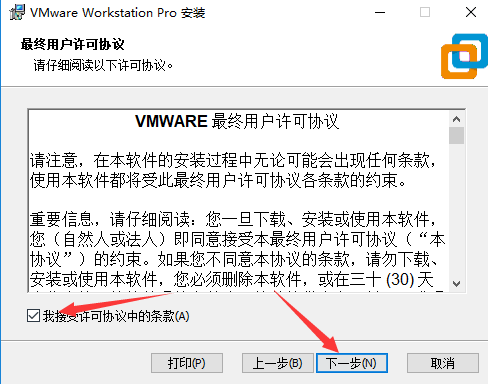
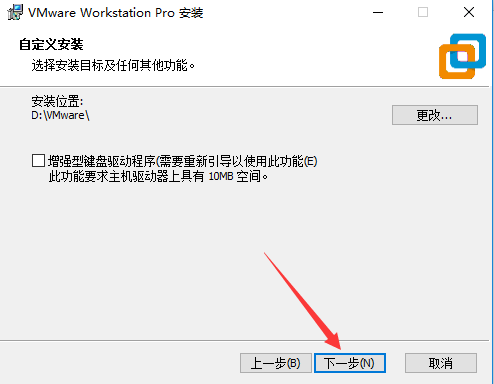
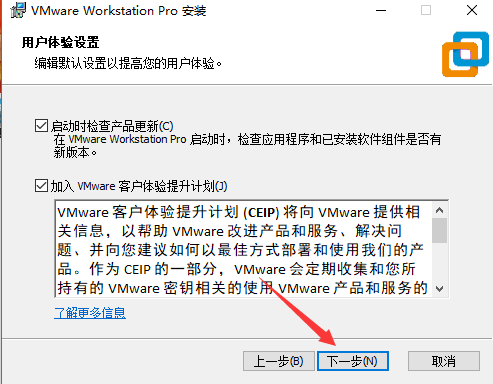
搭建Hadoop伪分布式大数据环境

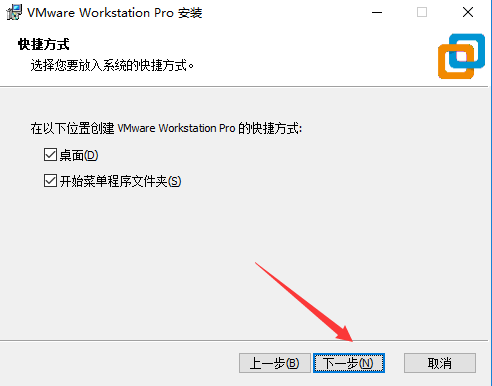
1. 安装VMWare虚拟机

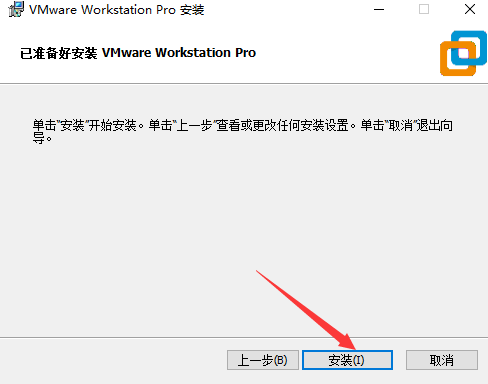


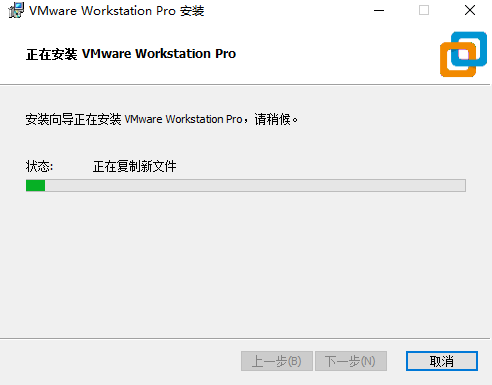














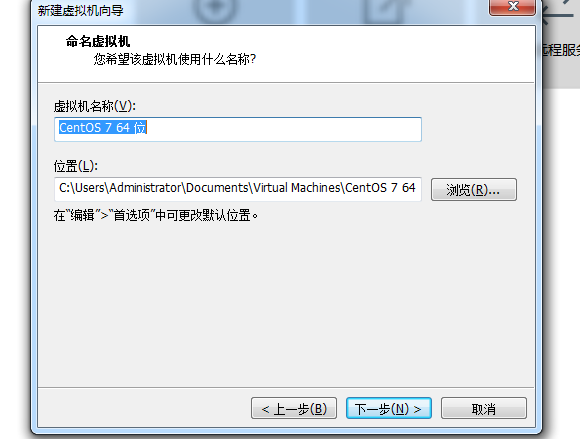
1. 在虚拟机环境中安装Linux操作系统

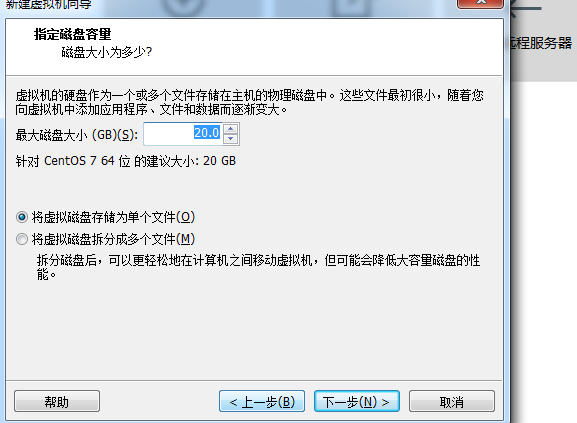
打开VMware选择新建虚拟机



选择镜像文件

设置虚拟机名称和安装位置

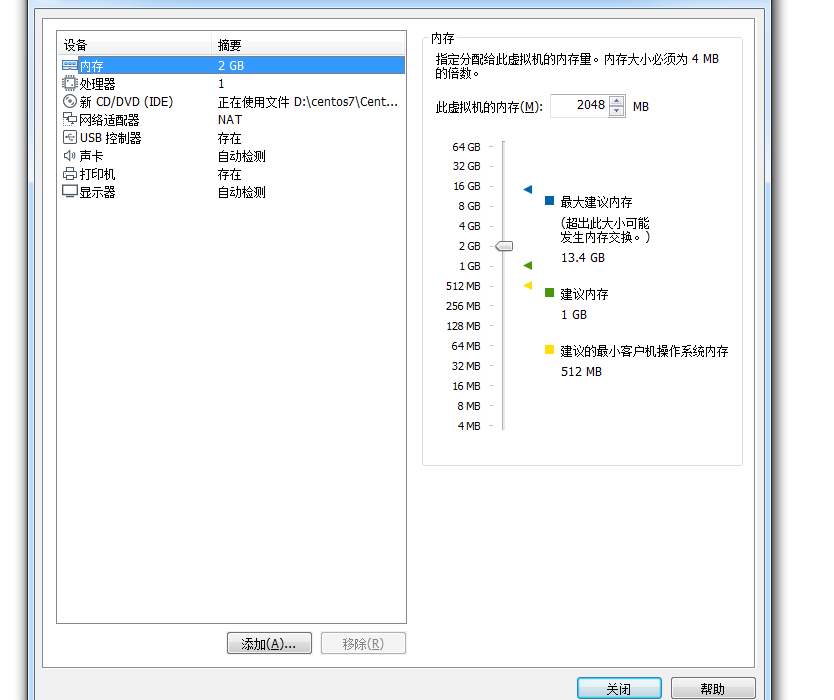




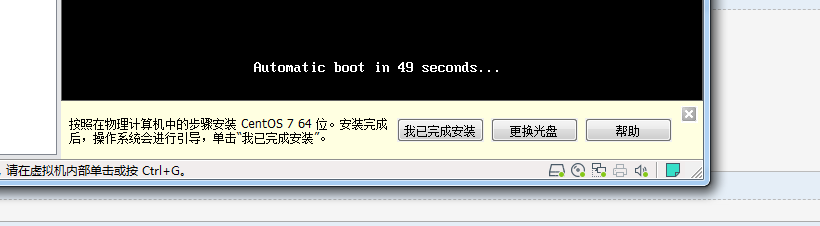
点击自定义配置



 内存根据自己电脑的内存来我这里设置成2g,然后点击关闭



点击我已安装,然后等待



选择中文然后继续



1. 在Linux系统中安装JDK并完成环境变量配置

此次安装的CentOS版本自带jdk1.8 无需再次安装

1. 安装Hadoop，实现伪分布式配置

安装hadoop

在官网下载

[hadoop-2.7.7.tar.gz](https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-2.7.7/hadoop-2.7.7.tar.gz)

通过WinScp上传文件到虚拟机中

解压



**配置hadoop-env.sh**

export JAVA\_HOME=/usr/java/jdk1.8.0\_171-amd64

**配置core-site.xml**

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

**配置hdfs-site.xml**

<configuration>

<property>

<name>dfs.data.dir</name>

<value>/usr/hadoop/hadoop-2.7.7/datanode</value>

</property>

</configuration>

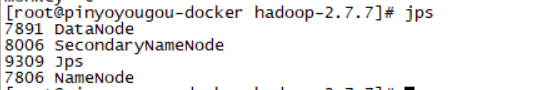
**格式化 namenode**



**启动**

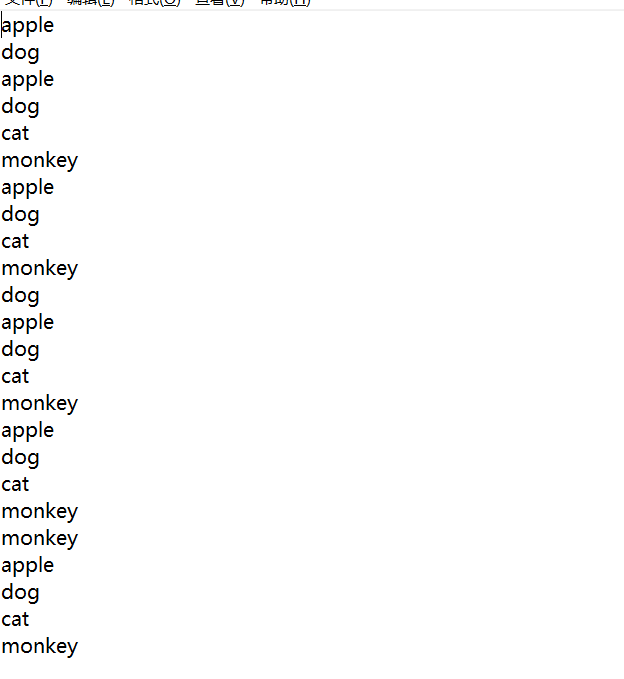


**验证 输入jps 查看进程**



1. 执行wordcount程序，统计指定目录下的词频

**创建一个文本文档如下所示**



**将此文件传到hdfs上**



**执行wordcount**



**查看输出结果**

