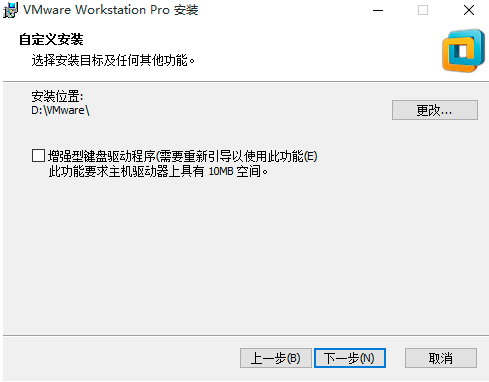
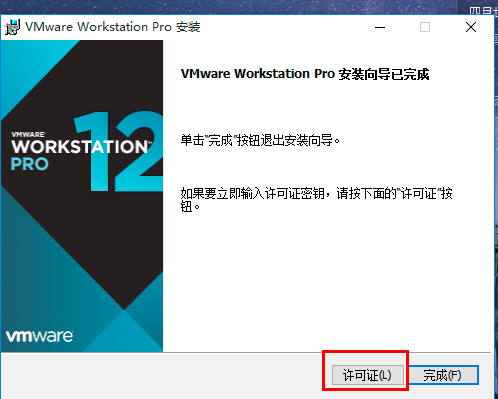
**搭建Hadoop伪分布式大数据环境**

# 安装VMWare虚拟机

使用默认安装：

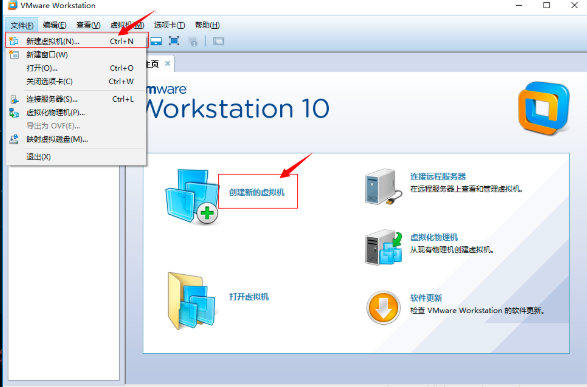




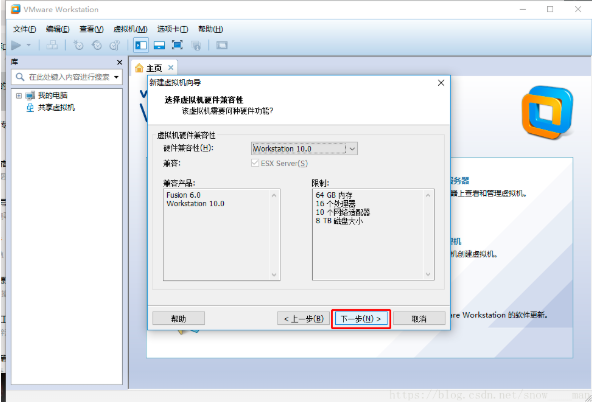
安装完成，输入许可证激活。

# 在虚拟机环境中安装Linux操作系统

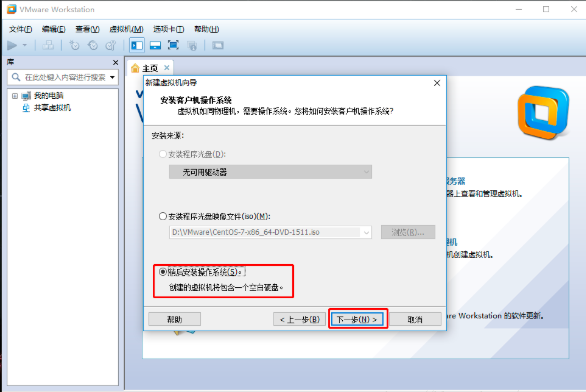
选择新建虚拟机



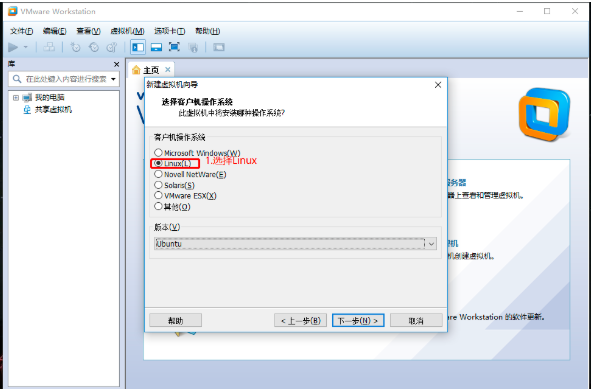
下一步



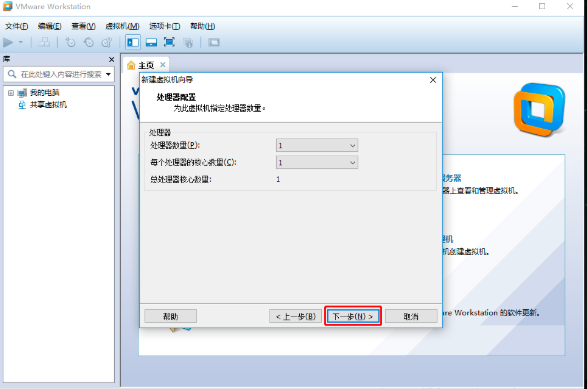
稍后安装操作系统



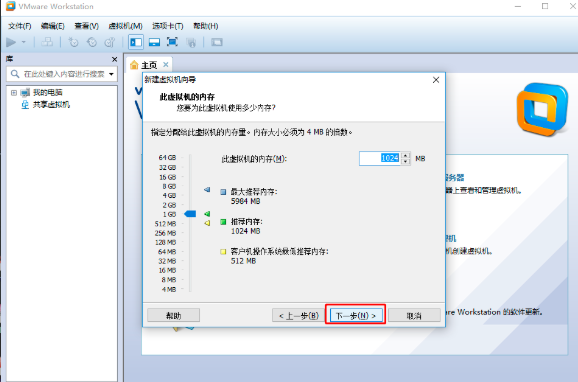
选Linux



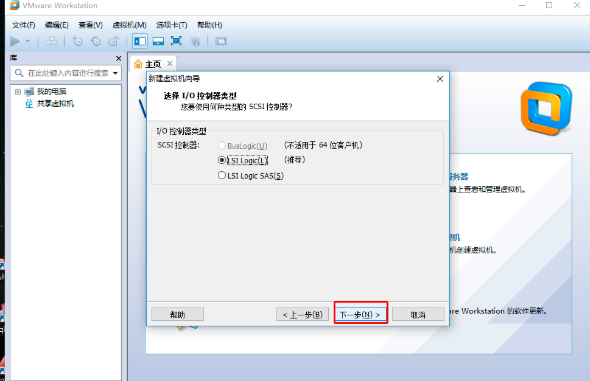
设置处理器个数



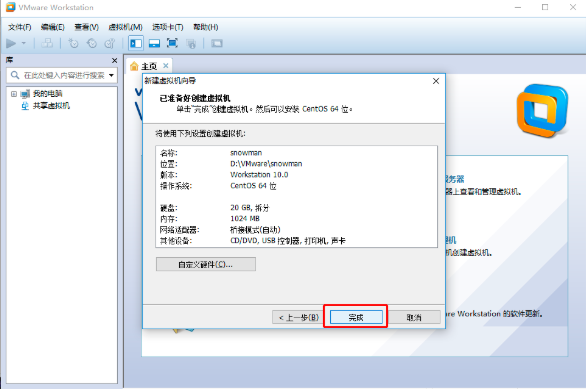
配置内存



IO类型



一直下一步，配置完成



开启centos系统，进行一些必要的配置



创建root用户和普通用户



安装完成，重启虚拟机

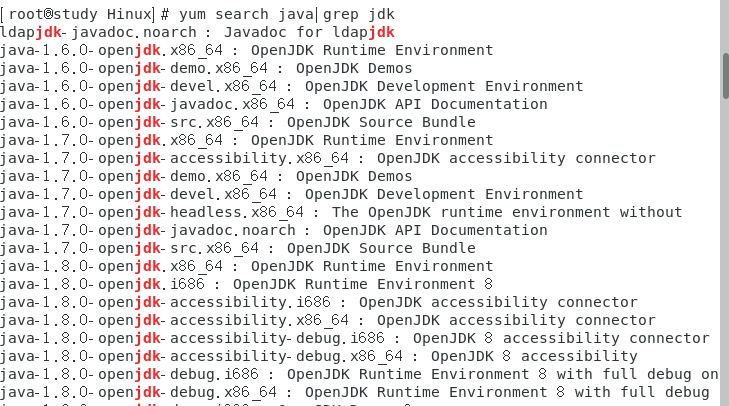


解决centos无法联网的问题：

https://blog.csdn.net/qq\_24342335/article/details/81449270

# 三、安装JDK

搜索安装包

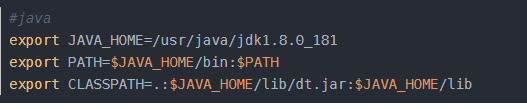


下载jdk1.8

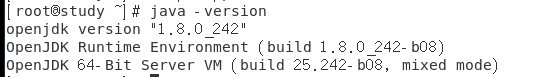


配置环境变量





测试环境

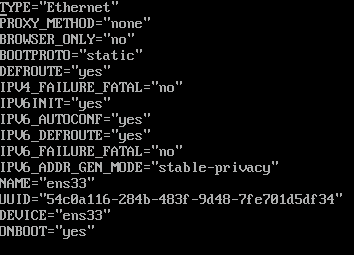


# 四、安装Hadoop，实现伪分布式配置

修改主机名



设置静态IP



添加主机与IP映射

关闭防火墙

设置ssh免登陆



解压下载的hadoop安装包

配置hadoop,分别配置以下几个文件：

配置hadoop-env.sh,修改jdk路径

配置core-site.xml

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop01:9000</value>

</property>

</configuration>

配置hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

配置maphred.site.xml

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

配置yarn-site.xml

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hadoop01</value>

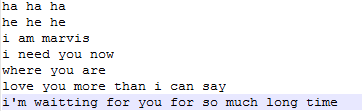
</property>

</configuration>

在浏览器地址栏中输入 192.168.91.124:8080，登录成功！,

# 五、执行wordcount程序，统计指定目录下的词频

创建统计的文件：



将文件上传至hdfs：



执行wordcount程序

查看结果：

