大数据平台技术课程实验报告

**实验一：开发与运行环境安装**

姓名：\_\_\_杨佳森\_\_\_\_\_

学号：\_\_\_2112080106\_

专业：\_\_\_数据科学与大数据技术

年级：\_ 2021

主讲教师：\_ 刘春 \_

实验时间： 2023.2.17 至 2023.2.19

实验内容与要求：

[1虚拟机的创建](#_Toc43287683)

[1.1虚拟化软件的安装](#_Toc43287684)

[1.2虚拟机的创建](#_Toc43287685)

[1.3 Vmware tools的安装](#_Toc43287686)

[2 JDK的安装](#_Toc43287688)

[3 IDEA+Maven的安装](#_Toc43287689)

[3.1 IDEA的安装](#_Toc43287690)

[3.2 Maven的安装](#_Toc43287691)

[3.3 在IDEA项目中配置JDK和Maven](#_Toc43287692)

[4 Hadoop的安装](#_Toc43287693)

[4.1 SSH的安装](#_Toc43287694)

[4.2 Hadoop的安装](#_Toc43287695)

[4.3 伪分布式环境配置](#_Toc43287696)

**特别提醒：部署过程需要有截图以及相应的文字介绍，欢迎大家把自己在实践过程中碰到的问题以及解决方法也记录下来**

1. **虚拟机的创建**

[**1.1虚拟化软件的安装**](#_Toc43287684)

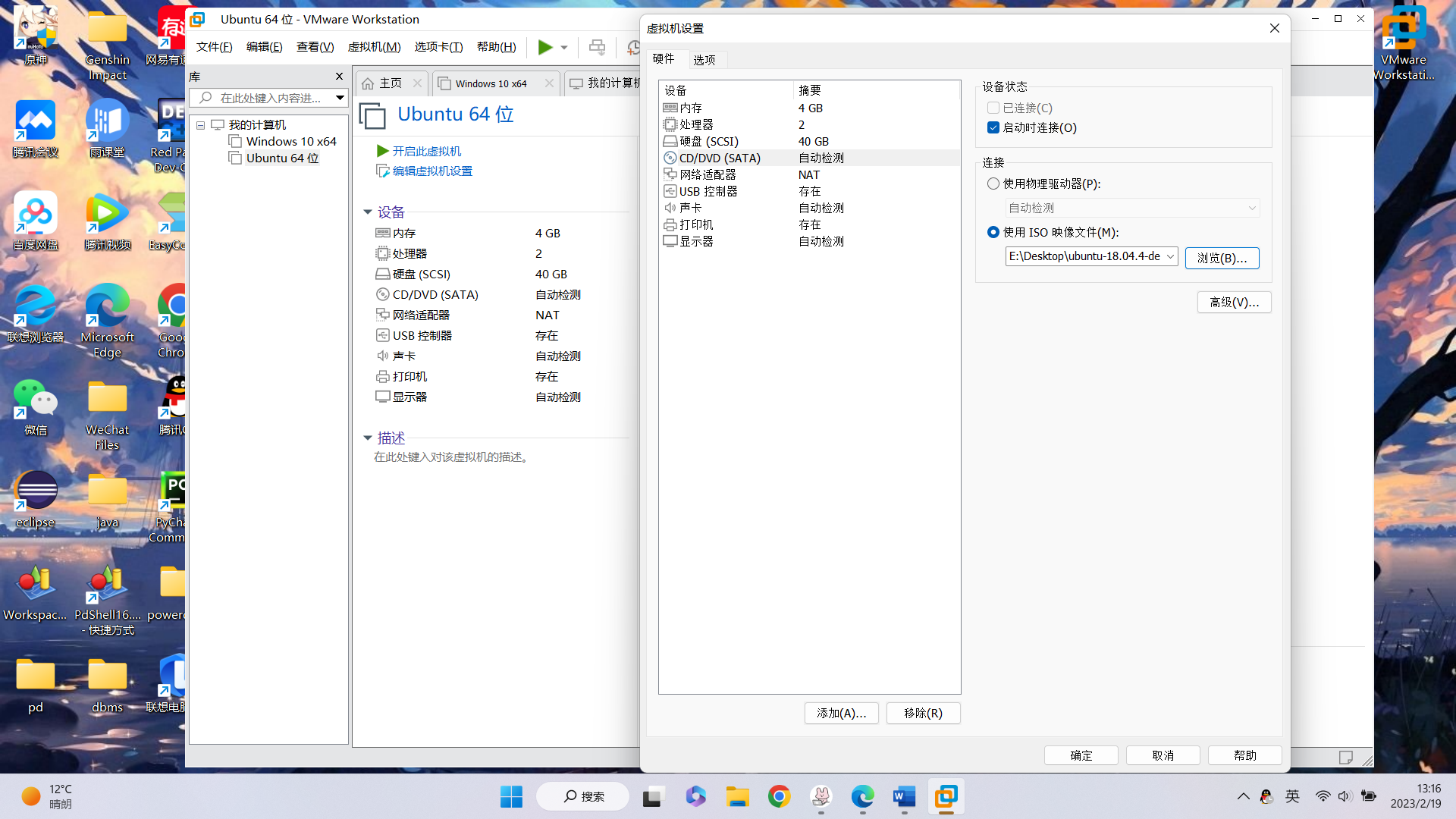
在VM官网[Search VMware Download](https://customerconnect.vmware.com/en/downloads/search?query=vmware%20workstation%20pro%2016)下载VMware Workstation Pro 16.0



（之前已经装过，不再进行安装了。）

[**1.2虚拟机的创建**](#_Toc43287685)

**乌班图的安装过程参考了博客**[(200条消息) 【慕伏白教程】在Vmware中安装Ubuntu流程\_vmware安装ubuntu\_慕伏白的博客-CSDN博客](https://blog.csdn.net/qq_43377653/article/details/126877889)，其中在选择最大磁盘大小时与教材有所分歧，我选择了教材上的将虚拟磁盘存储为单个文件（O）。

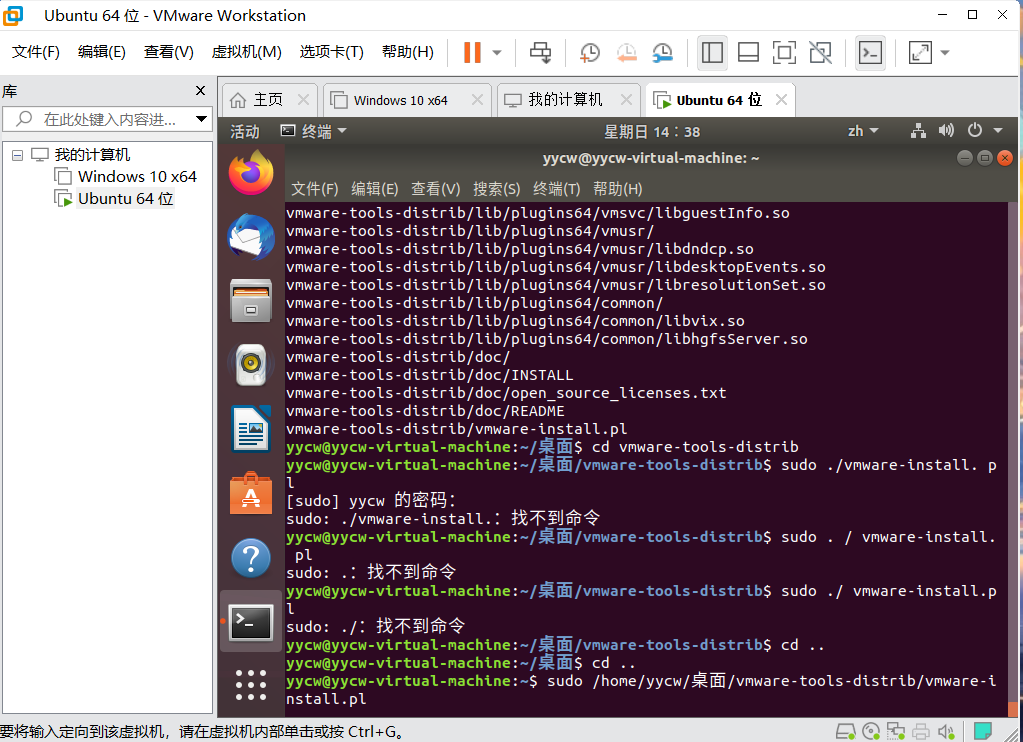


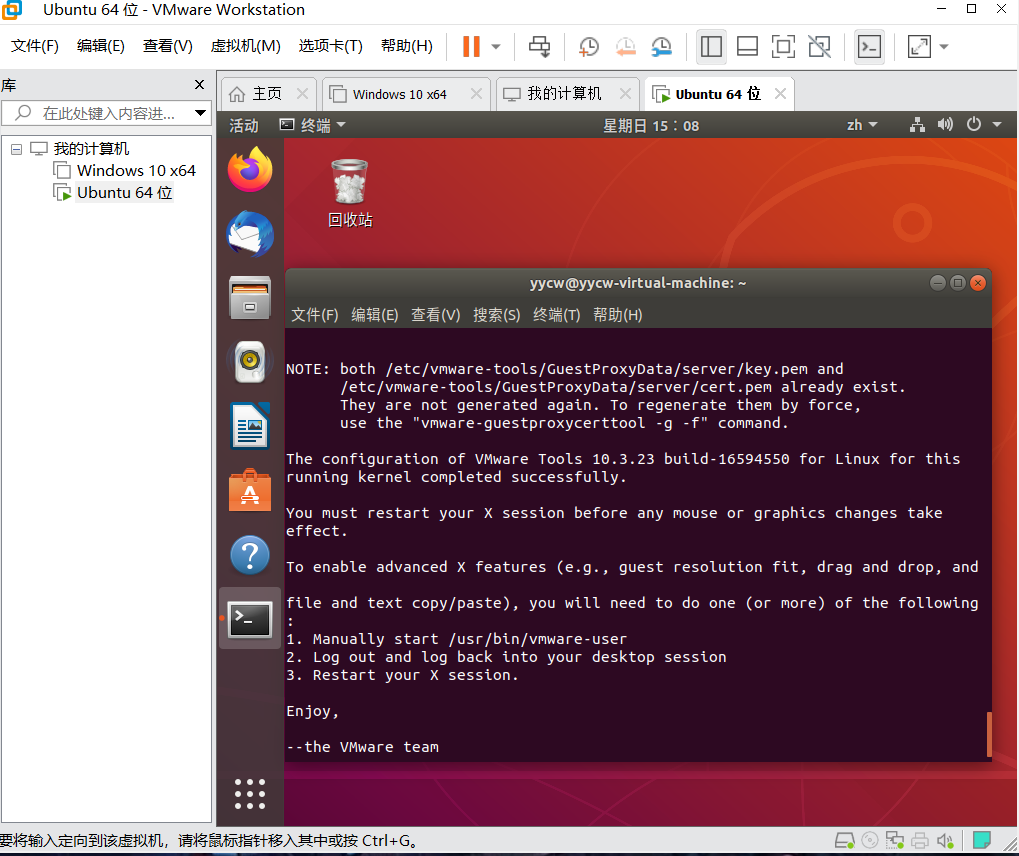


* 1. [**Vmware tools的安装**](#_Toc43287686)

在终端输入命令时遇到了问题，那就是如何在终端输入汉字，查阅相关资料得知在ubuntu中切换中英文快捷键是win+空格键。

另外教材上命令显示不够仔细，应注意tar后还要再加上一个空格，然后再跟-zxvf，而且在解压文件时教材的文件名与我的不一样，应当用自己本机显示的文件名。





以及双向复制粘贴的问题，[(200条消息) Windows和VMware中的ubuntu22.04两者之间不可以复制、粘贴\_珍惜时间\_多思考的博客-CSDN博客\_unbuntu22.04 不让从外粘贴](https://blog.csdn.net/qq_41570500/article/details/125756573?spm=1001.2101.3001.6661.1&utm_medium=distribute.pc_relevant_t0.none-task-blog-2~default~CTRLIST~Rate-1-125756573-blog-125905122.t5_download_comparev1&depth_1-utm_source=distribute.pc_relevant_t0.none-task-blog-2~default~CTRLIST~Rate-1-125756573-blog-125905122.t5_download_comparev1&utm_relevant_index=1)和[(200条消息) Ubuntu 22.04 vmware 复制文件问题\_闫有尽意无琼的博客-CSDN博客](https://blog.csdn.net/yanchenyu365/article/details/125905122?ops_request_misc=&request_id=&biz_id=102&utm_term=Ubuntu22.04%20vmtools&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-7-125905122.142%5Ev50%5Econtrol,201%5Ev3%5Eadd_ask&spm=1018.2226.3001.4187)也得以解决

1. [**JDK的安装**](#_Toc43287688)

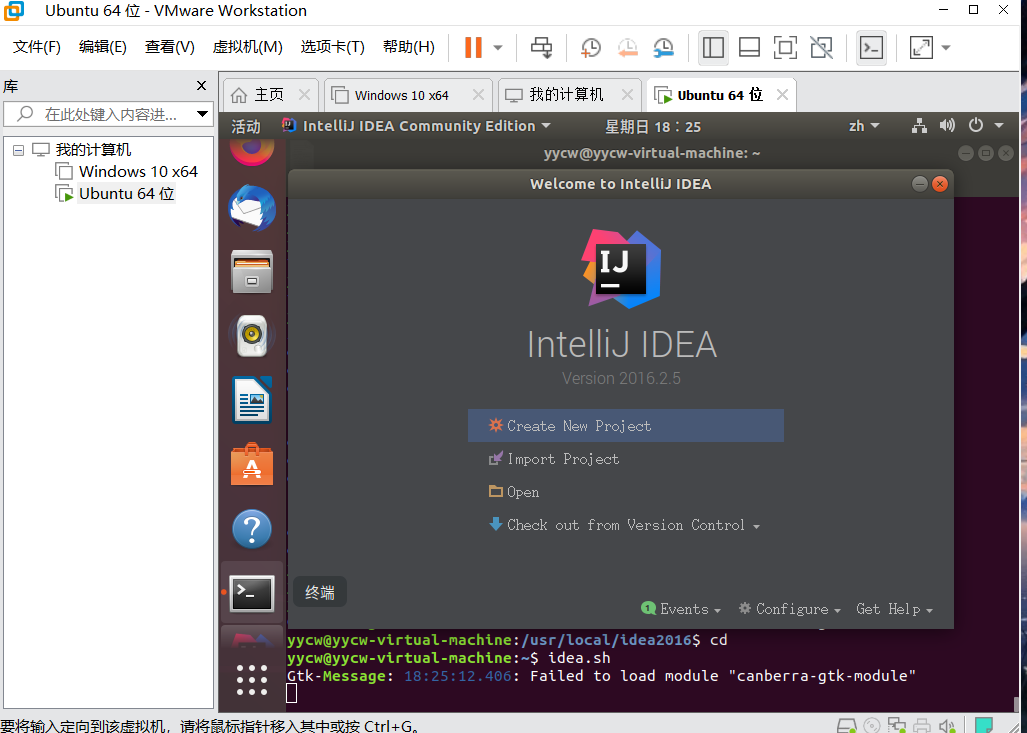
# 在配置换将变量时由于输入命令不正确导致ls cd 等命令不能使用，提示：没有那个文件或目录...然后参考了 [(200条消息) linux中环境变量PATH设置错误，导致ls cd 等命令不能使用，提示：没有那个文件或目录...\_メイ的博客-CSDN博客](https://blog.csdn.net/weixin_30785593/article/details/95733862?spm=1001.2101.3001.6650.1&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1-95733862-blog-83717739.pc_relevant_recovery_v2&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1-95733862-blog-83717739.pc_relevant_recovery_v2&utm_relevant_index=2) 的博客，问题得以解决。

# 

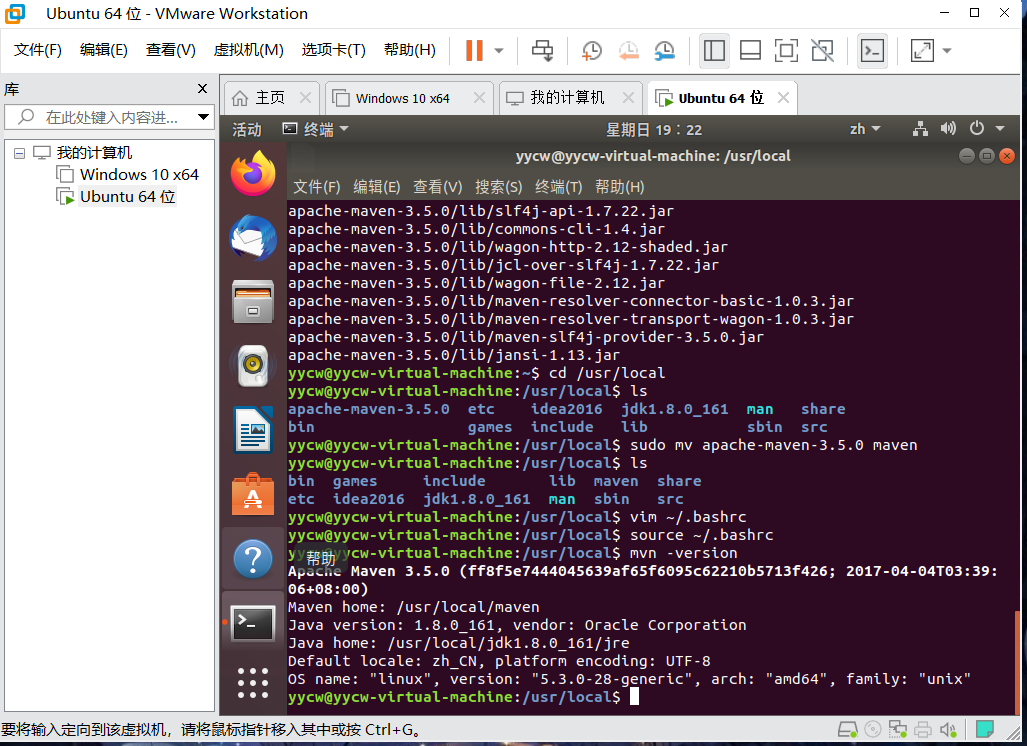
# 

1. [**IDEA+Maven的安装**](#_Toc43287689)
   1. [**IDEA的安装**](#_Toc43287690)

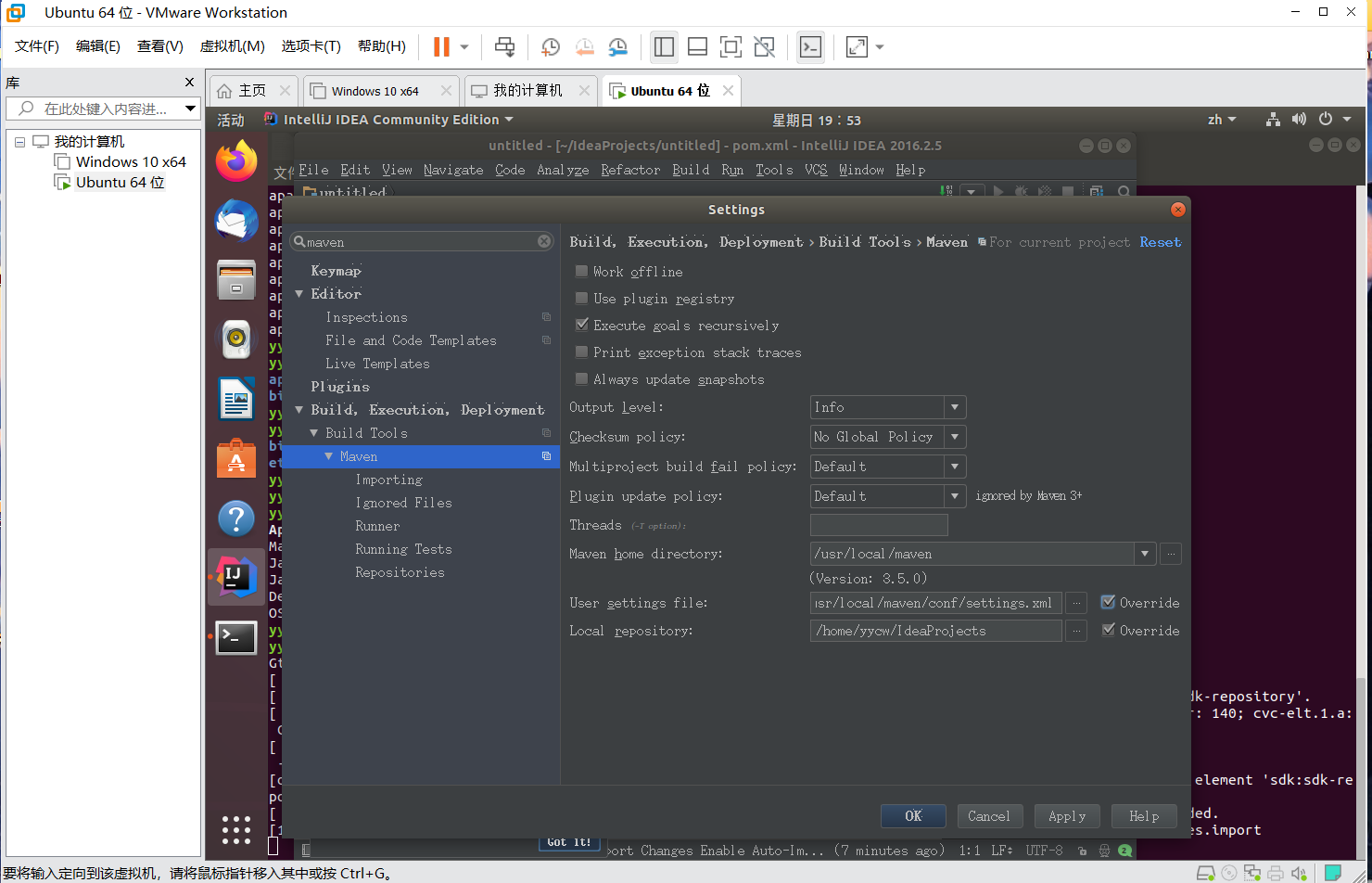
在解压缩idea时遇到了（*tar (child): ideaIC-2016.2.5.tar.gz：无法 open: 没有那个文件或目录）*报错，应在压缩包指令前加入绝对路径。命令如下：sudo tar -zxvf /home/yycw/桌面/ideaIC-2016.2.5.tar.gz -C /usr/local （yycw是我的用户名，下同）



* 1. [**Maven的安装**](#_Toc43287691)

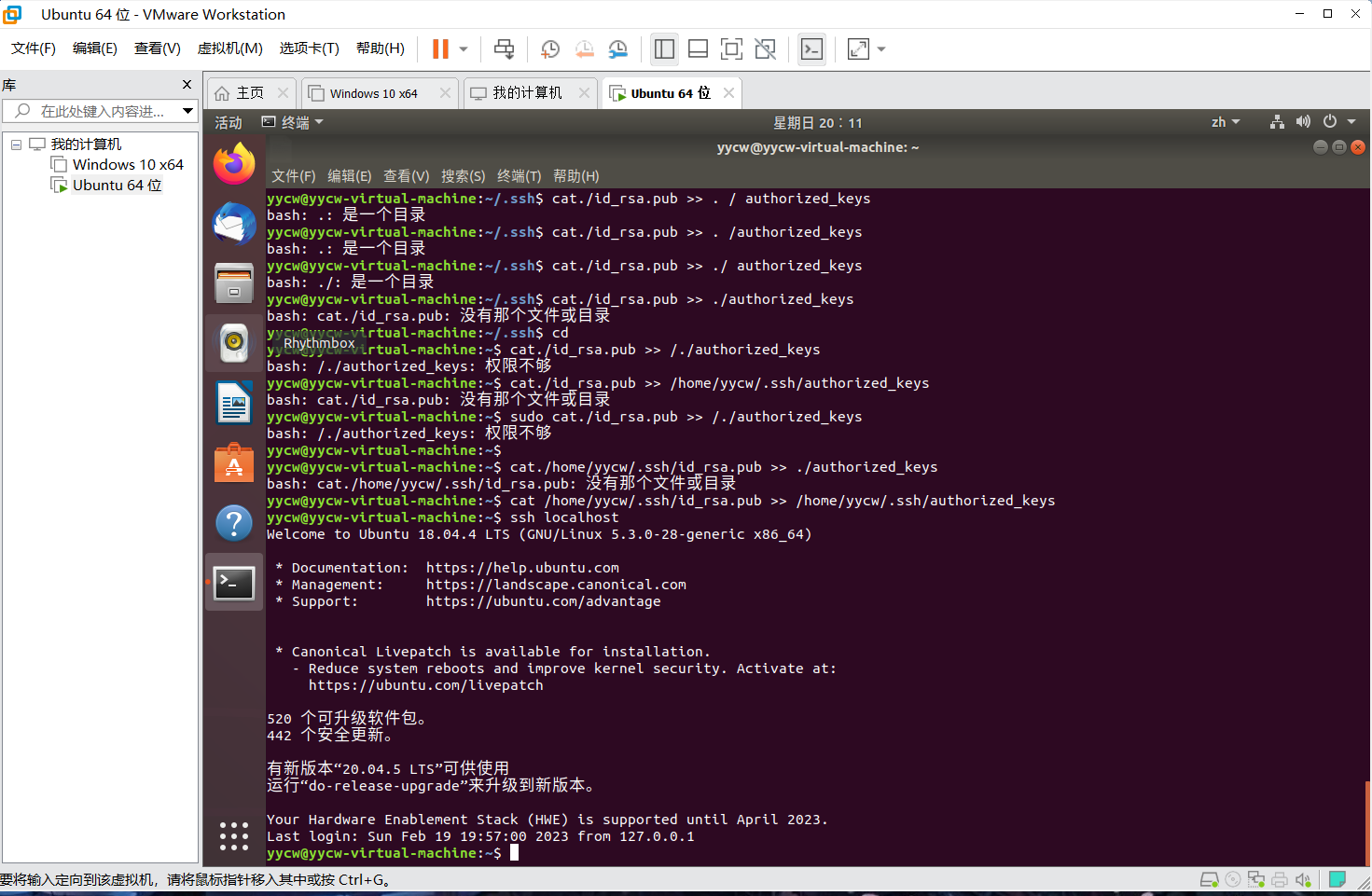
****

* 1. [**在IDEA项目中配置JDK和Maven**](#_Toc43287692)



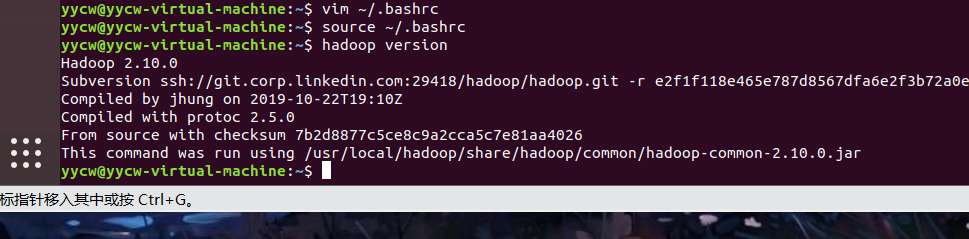
1. [**Hadoop的安装**](#_Toc43287693)
   1. [**SSH的安装**](#_Toc43287694)

教材上添加密钥的命令并不适用，所以我用命令如下：cat /home/yycw/.ssh/id\_rsa.pub >> /home/yycw/.ssh/authorized\_keys



* 1. [**Hadoop的安装**](#_Toc43287695)

赋予Hadoop用户使用当前目录下hadoop目录的权限应该使用命令如下：sudo chown -R yycw ./hadoop



* 1. [**伪分布式环境配置**](#_Toc43287696)

在对NameNode进行格式化时用的命令是：hdfs namenode -format

