

**Hadoop大數據處理技術及應用基礎**

**實驗報告一**

**專 業：** 數據科學

**授課教師：** 应作斌

**學生姓名：** 刘传

（一）安装VMware workstation 16



**图1 VMware workstation 16界面图**

注：在之前安装VMware workstation 14出现如下若干问题：

启动虚拟机出现蓝屏，经过查阅资料得知：

①需关闭Hyper-V所有服务，并在服务中禁用；

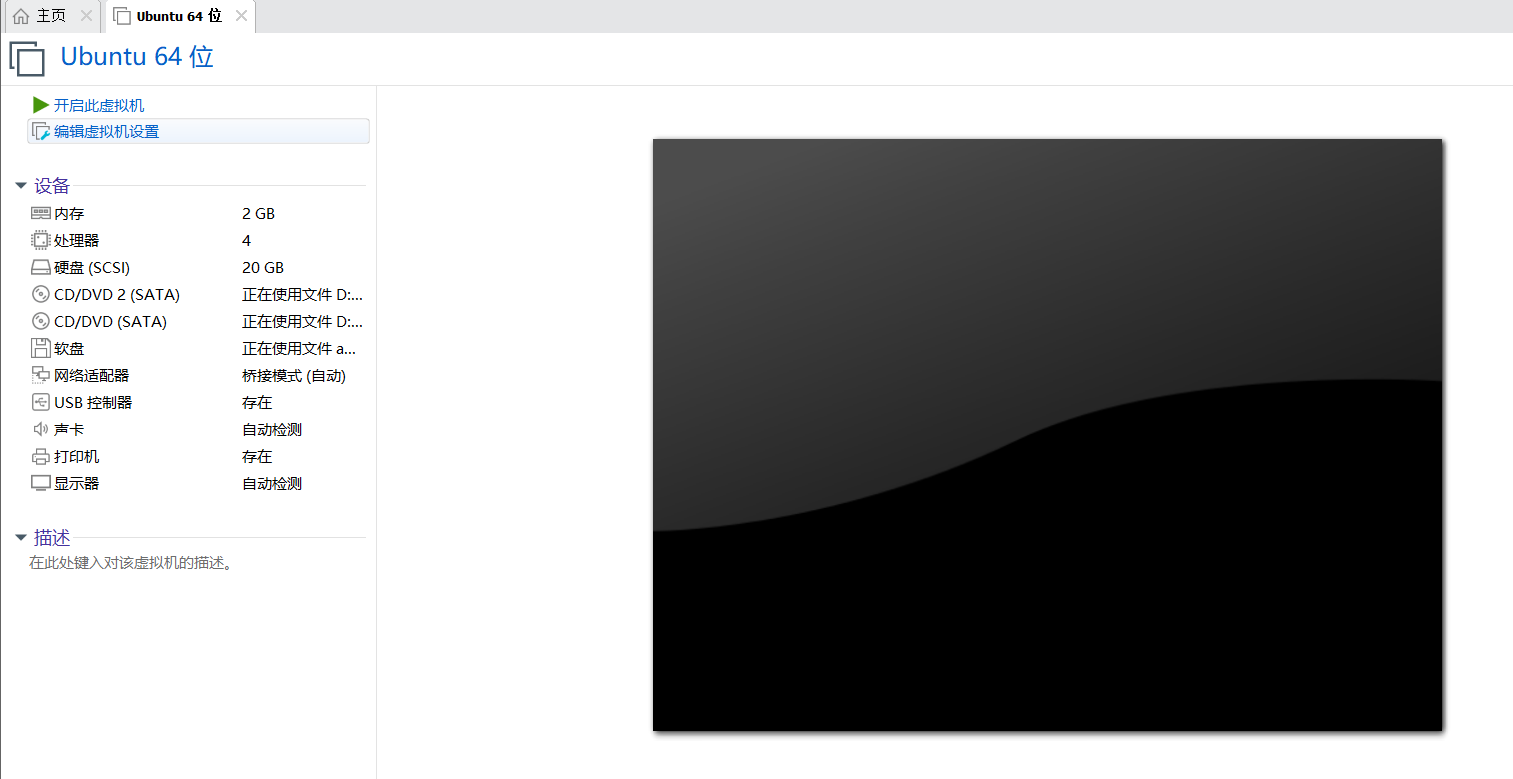
②打开Windows虚拟机监控平台，虚拟机平台

③打开服务，禁用HV主机服务，并将Hyper-V相关服务全部禁用。

④win+X（Windows PowerShell选择管理员），输入 bcdedit /set hypervisorlaunchtype off，再重启电脑。

通过以上步骤的设置，依然存在问题，后发现VMware workstation14貌似与Win10不兼容，依然蓝屏，改用VM16成功解决蓝屏问题。

（二）安装Ubuntu22.04



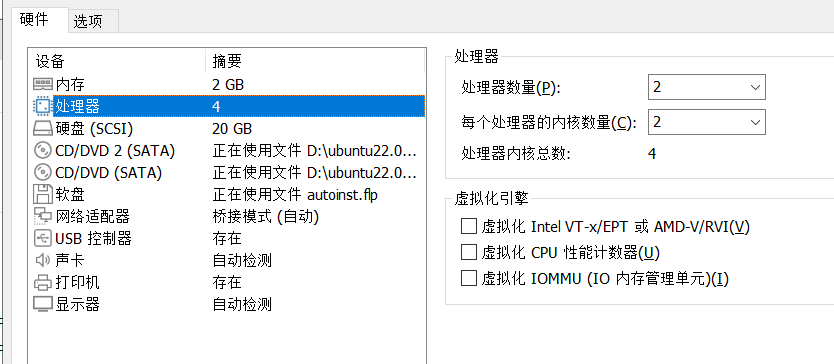
**图2 创建Ubuntu虚拟机界面图**

1.为虚拟机分配内存



**图3 分配内存界面图**

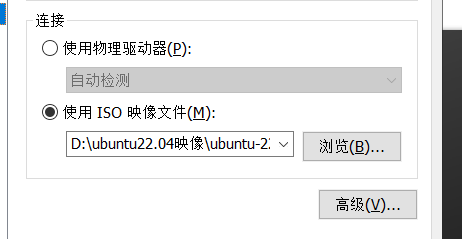
2.设置处理器数量及每个处理器的内核数量



**图4 配置处理器界面图**

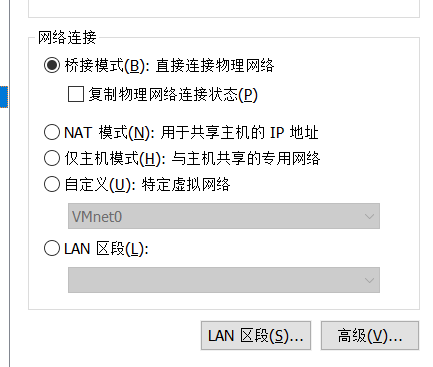
3.设置硬盘大小（20GB）

4.使用ISO映像文件



**图5 导入ISO映像文件界面图**

5.设置网络连接模式



**图6 设置网络连接模式界面图**

出现的问题：在之前为Centos7配置网络IP时，将子网IP与子网掩码修改后，可能导致了Ubuntu在NAT模式无法正常连接网络；后来改为桥接模式直接连接物理网络，可连接网络。

6.建立名为‘akim’的用户，并设置密码，成功安装Ubuntu。

7.成功安装Ubuntu



**图7 Ubuntu界面图**

出现的问题：因为无法联网，导致无法通过sudo apt-get install open-vm-tools，安装vm-tools，也无法下载输入法以及与windows正常复制粘贴操作的工具的下载，导致影响了安装hadoop的进度。

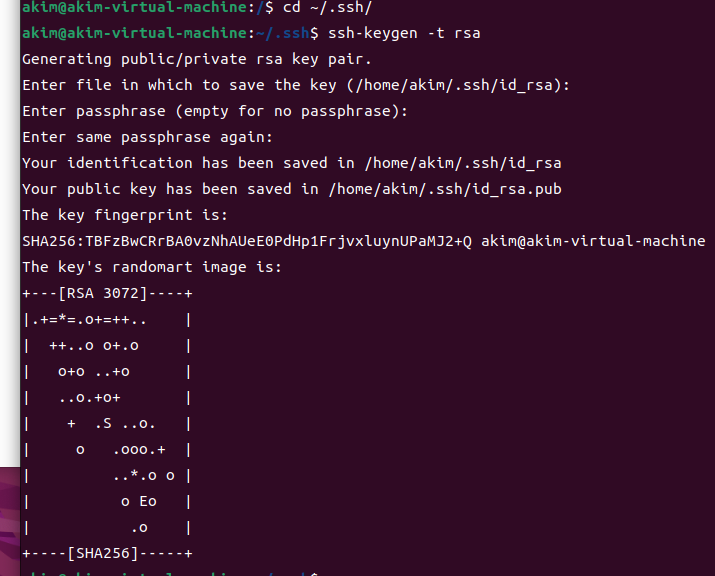
8.更新apt

在shell终端，执行**$sudo apt-get update**

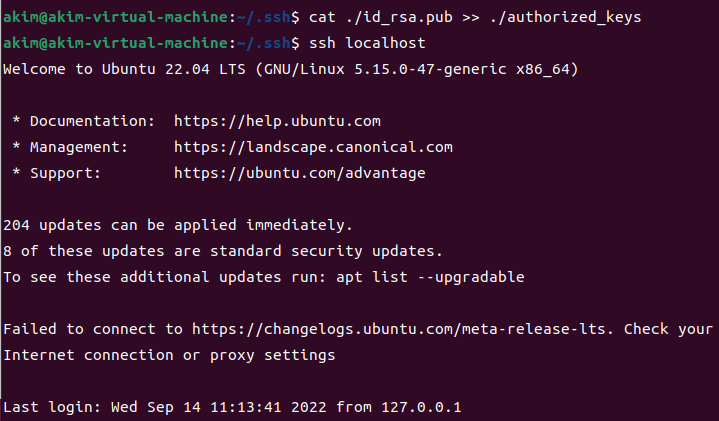
9.安装SSH、配置SSH无密码登录



**图8 安装并配置SSH界面图1**



**图9 安装并配置SSH界面图2**



**图10 SSH登陆成功界面图**

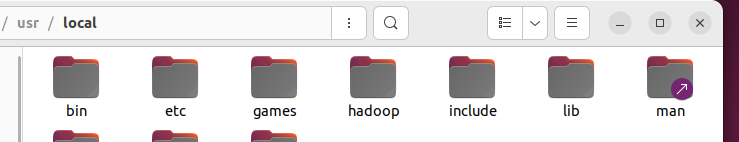
10.安装JAVA环境

出现的问题：在配置java环境变量时，仅通过命令行进行设置，在对hadoop安装进行测验时（输入hadoop显示没有java）出现错误，后来通过**gedit ~/.bashrc**在该文件中添加了JAVA\_HOME，成功解决了该问题。

（三）安装hadoop

1.下载hadoop安装包

2.将安装包解压至/usr/local并安装



**图11 解压hadoop成功界面图**

3.对hadoop进型文件夹改名，权限修改等操作

4.hadoop伪分布式配置

①修改配置文件core-site.xml

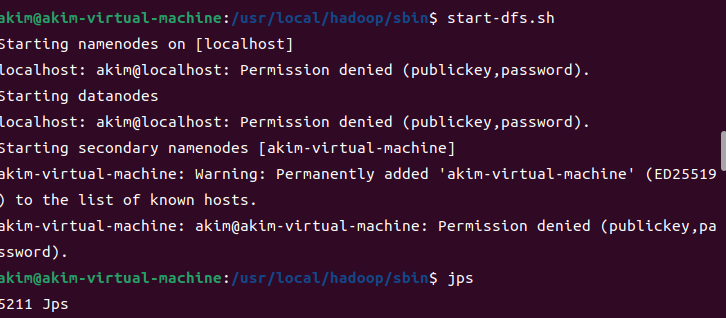
②修改hdfs-site.xml

③执行NameNode格式化，通过如下代码：

**$./bin/hdfs namenode -format**

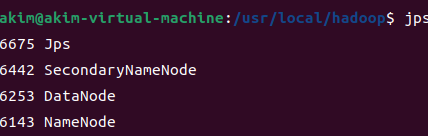
④开启NameNode和DataNode守护进程

出现的问题：在配置hadoop的NameNode和DataNode时，出现Permission denied(publickey,password)，经过查阅资料发现，在之前缺少了配置SSH无密码登陆导致的，如下图所示。



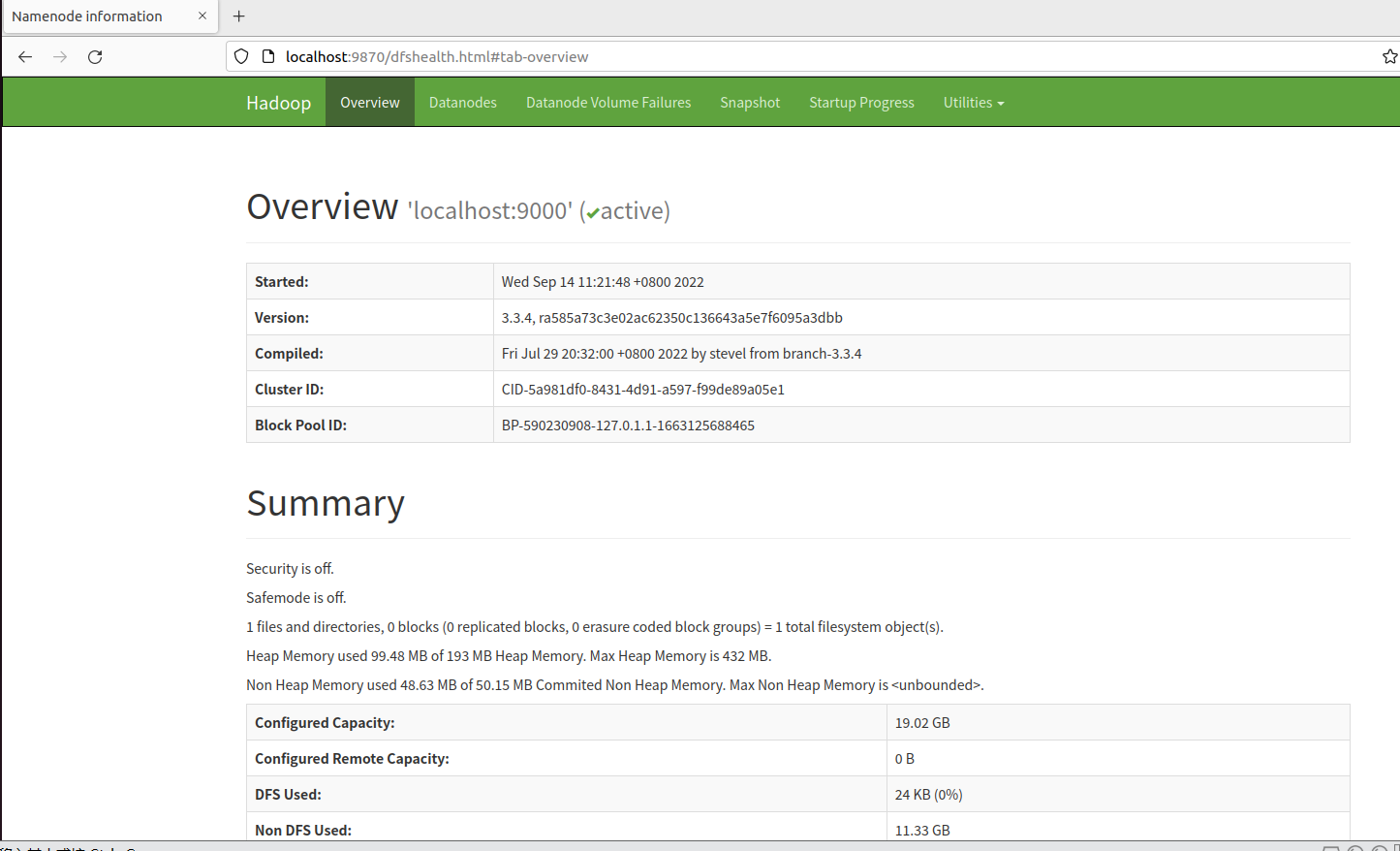
**图12 错误示意图**

在成功配置SSH密码登录后，重新进行守护进程的开启，并通过jps进行查看。



**图13 查看进程开启界面图**

⑤访问Web界面http://localhost:9870，成功显示如下图页面，hadoop安装完成。



**图14 Hadoop的Web界面图**

⑥关闭hadoop，运行如下代码：

**$./sbin/stop-dfs.sh**

下次启动hadoop时，无需进行NameNode的初始化，只需要运行 **./sbin/start-dfs.sh**即可。