

《Hadoop的安装与配置》

- 学院：电子信息工程学院
- 专业：数据科学与大数据专业
- 学号：1851804
- 姓名：苗成林
- 指导教师：李文根
- 时间：2021.11.11

《Hadoop的安装与配置》

实验目的及要求：

实验过程：

软件环境

安装与配置虚拟机

搭建Hadoop集群

实验目的及要求：

1. 熟悉Hadoop分布式环境的安装
2. 熟悉Hadoop分布式环境的安装
3. 启动Hdfs和Yarn功能和Hadoop集群

实验过程：

软件环境

软件	版本	安装包	备注
Linux OS	CentOS6.8	CentOS-6.8-x86_64-bin-DVD1.iso	64位
JDK	1.8+	jdk-8u151-linux-x64.rpm	64位
VMware	11	VMware-workstation-full-11.0.0-2305329.exe	
Hadoop	2.6.5	hadoop-2.6.5.tar.gz	已编译好的安装包
Eclipse	4.5.1	eclipse-jee-mars-1-win32-x86_64.zip	64位
Eclipse Hadoop**插件**	2.6.0	hadoop-eclipse-plugin-2.6.0	
SSH**连接工具**	5	Xme5.exe	

安装与配置虚拟机

•集群规模：3台虚拟机

–修改配置信息：主机名、ip地址、主机与ip的映射

–192.168.128.130 master

–192.168.128.131 slave1

–192.168.128.132 slave2

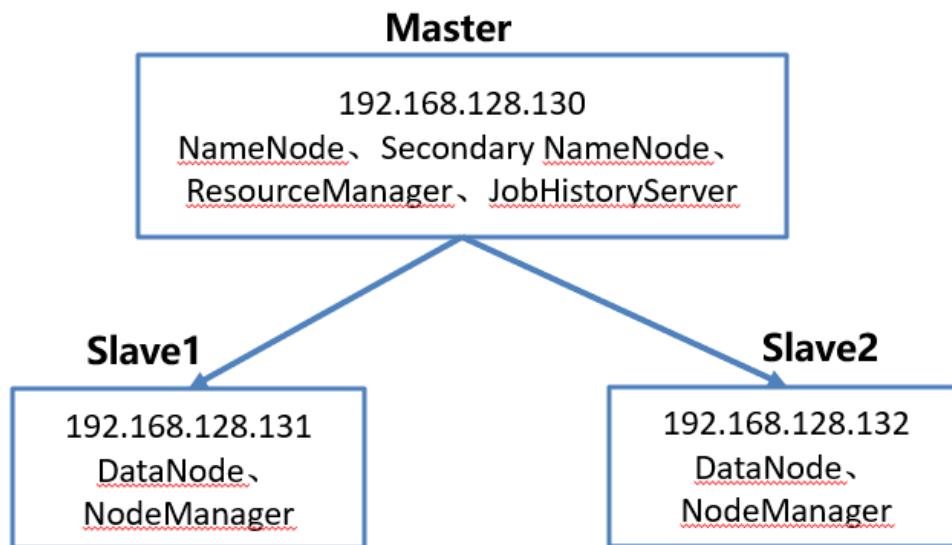
•配置SSH无密码登录

–SSH（Secure Shell的缩写）是建立在TCP/TP协议的应用层和传输层基础上的安全协议。SSH保障了远程登录和网络传输服务的安全性，起到防止信息泄露等作用

•配置时间同步服务

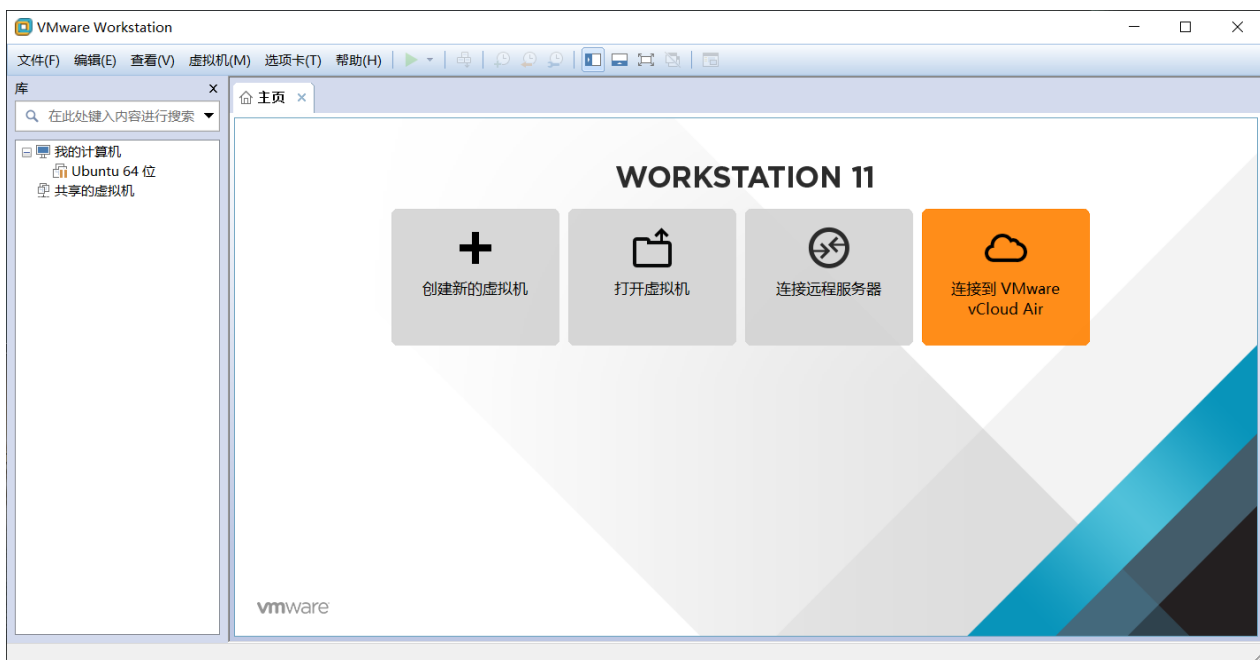
–NTP是用来使计算机时间同步化的一种协议，它可以使计算机对其服务器或时钟源做同步化，提供高精度度的时间校正。Hadoop集群对时间要求很高，主节点与各从节点的时间都必须要同步。配置时间同步服务主要是为了进行集群间的时间同步

搭建Hadoop集群



•安装Vmware

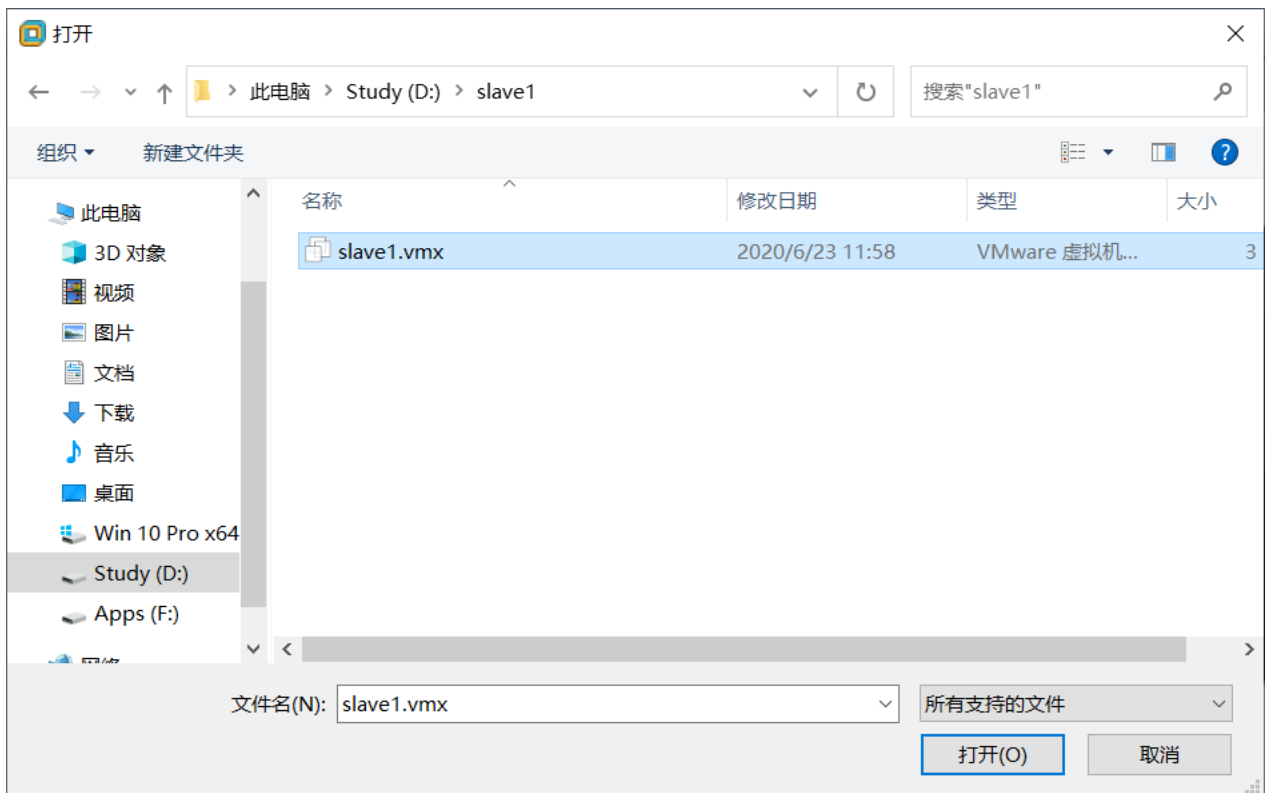
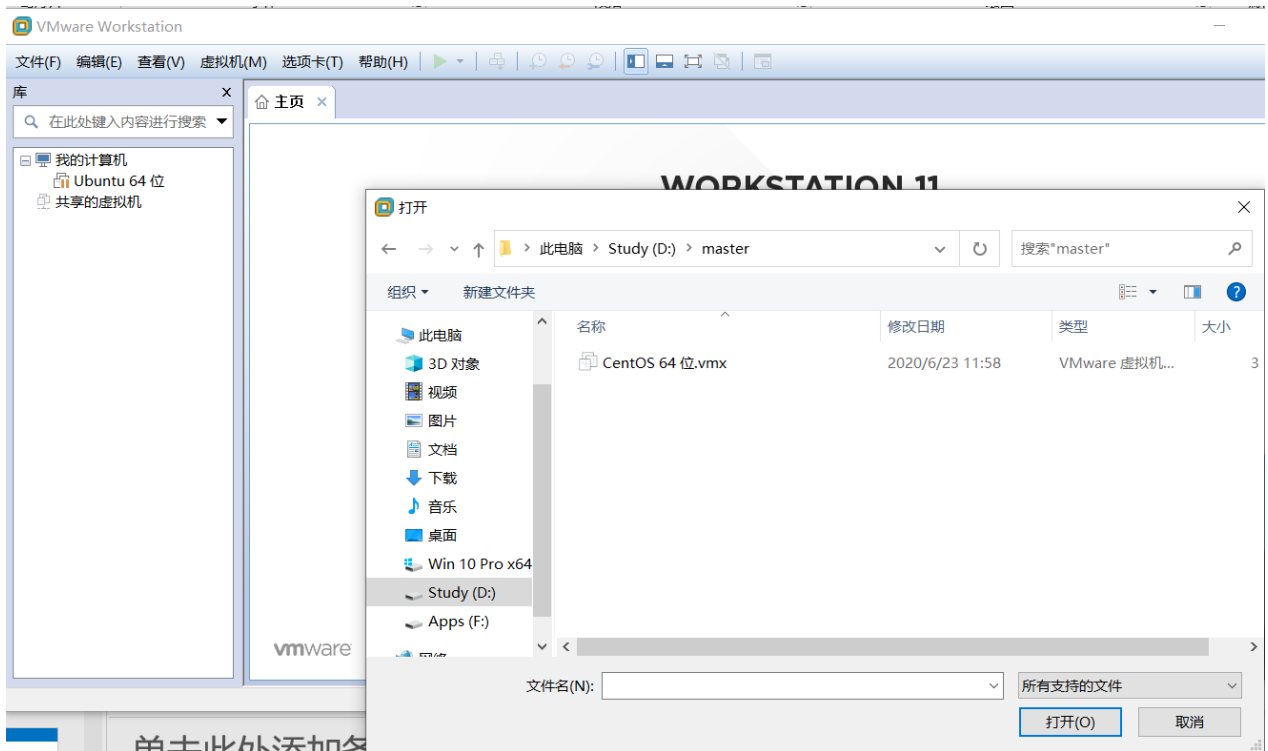
–本地电脑安装好 VMware，版本要为 11 或者以上。以下界面皆为 VMware11的操作界面，使用的安装包为 VMware-workstation-full-11.0.0-2305329.exe

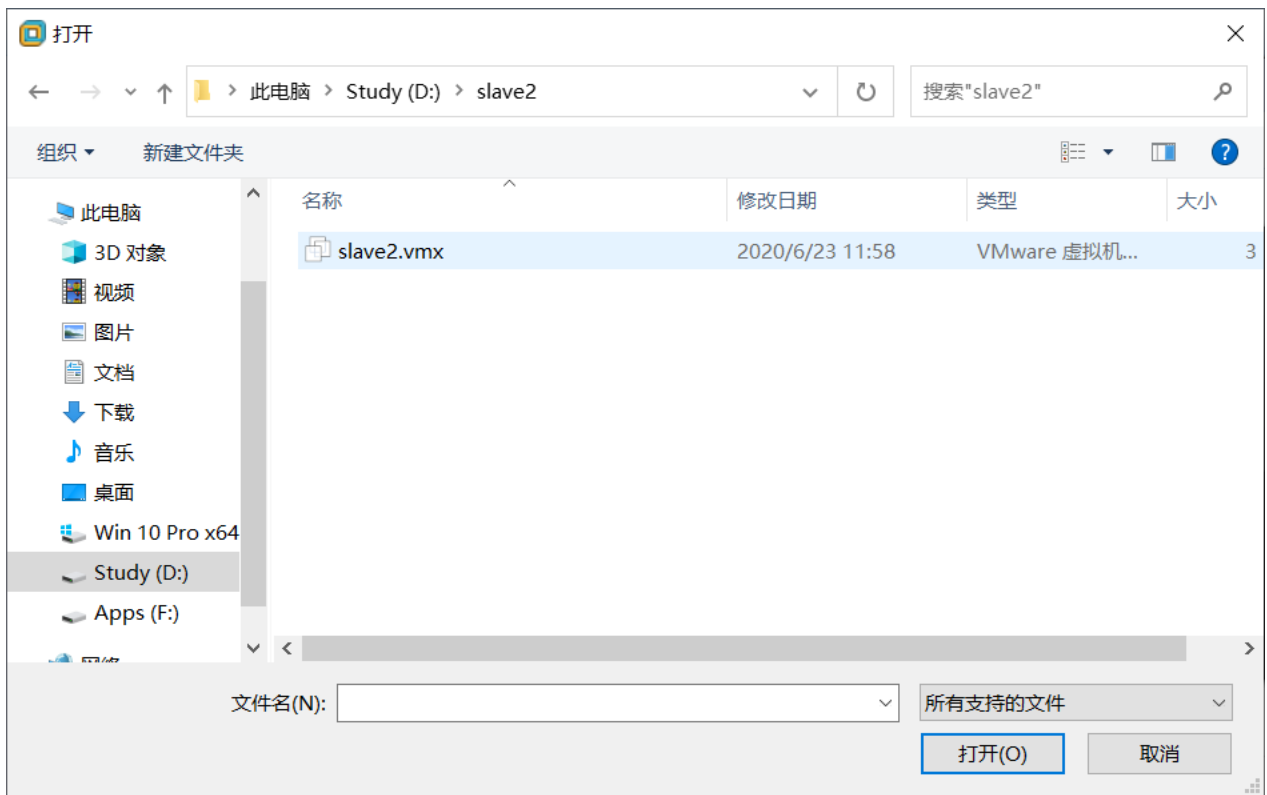


•虚拟机导入

–解压缩 master 压缩包和另外 2 个 slave 压缩包。以**管理员身份**打开 VMware

【打开】后，找到解压缩的 master 目录下，选中后点击【打开】。两个slave以类似的形式打开。





•修改虚拟网络编辑器设置

-导入完成后打开【编辑】中的【虚拟网络编辑器】，修改相关设置

修改 VMnet8 的子网 IP、子网掩码、NAT 设置和 DHCP 设置

虚拟网络编辑器

✕

名称	类型	外部连接	主机连接	DHCP	子网地址
VMnet0	桥接模式	自动桥接	-	-	-
VMnet1	仅主机...	-	已连接	已启用	192.168.44.0
VMnet8	NAT 模式	NAT 模式	已连接	已启用	192.168.128.0

<

>

添加网络(E)...

移除网络(O)

VMnet 信息

☐ 桥接模式(将虚拟机直接连接到外部网络)(B)

桥接到(T): 自动

自动设置(U)...

☒ NAT 模式(与虚拟机共享主机的 IP 地址)(N)

NAT 设置(S)...

☐ 仅主机模式(在专用网络内连接虚拟机)(H)

☒ 将主机虚拟适配器连接到此网络(V)

主机虚拟适配器名称: VMware 网络适配器 VMnet8

☒ 使用本地 DHCP 服务将 IP 地址分配给虚拟机(D)

DHCP 设置(P)...

子网 IP (I): 192 . 168 . 128 . 0

子网掩码(M): 255 . 255 . 255 . 0

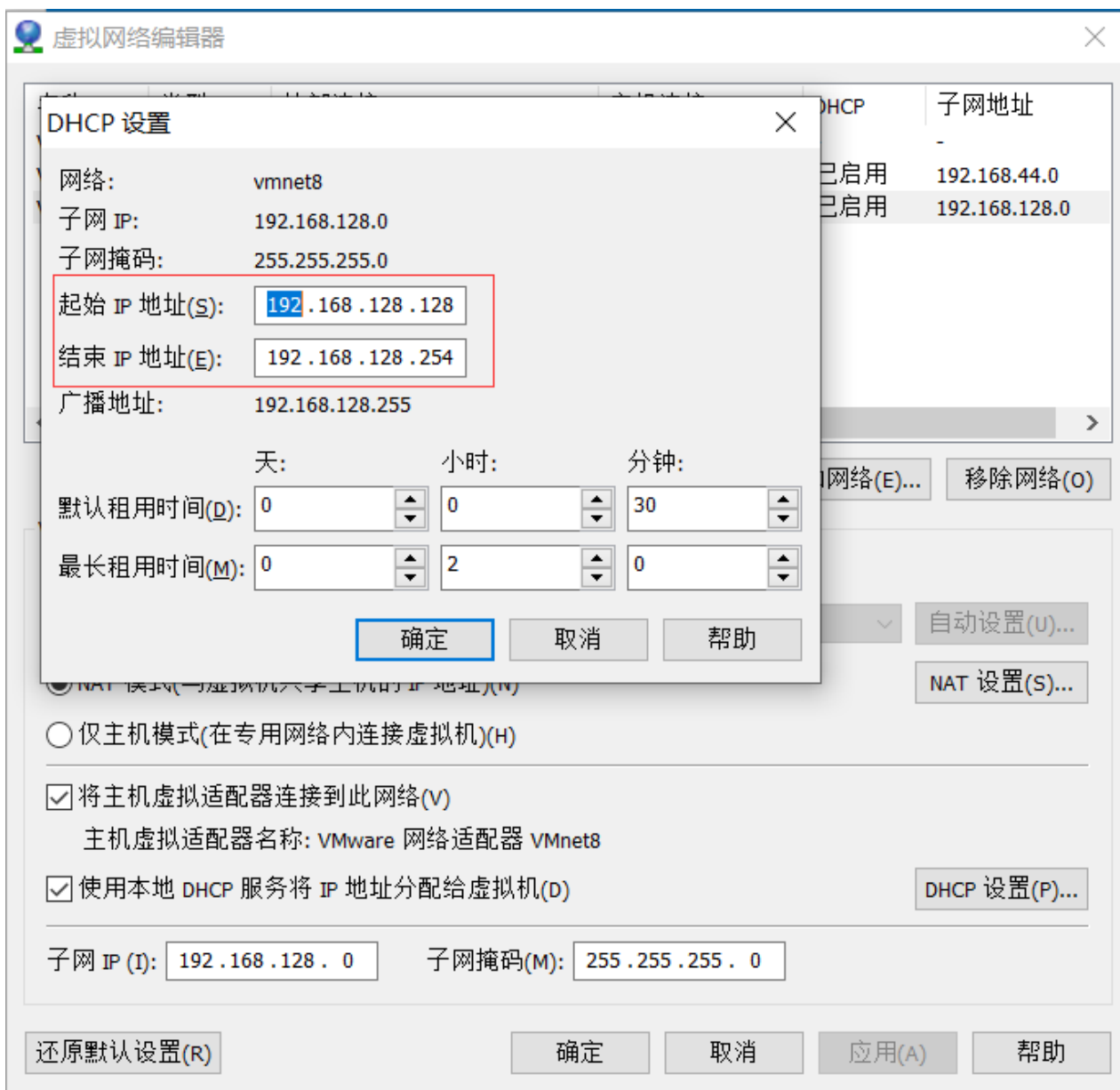
还原默认设置(R)

确定

取消

应用(A)

帮助



•启动3个虚拟机并登录

-用户名: root 密码: 123456

-主节点为 master, 2 个从节点分别为 slave1和slave2

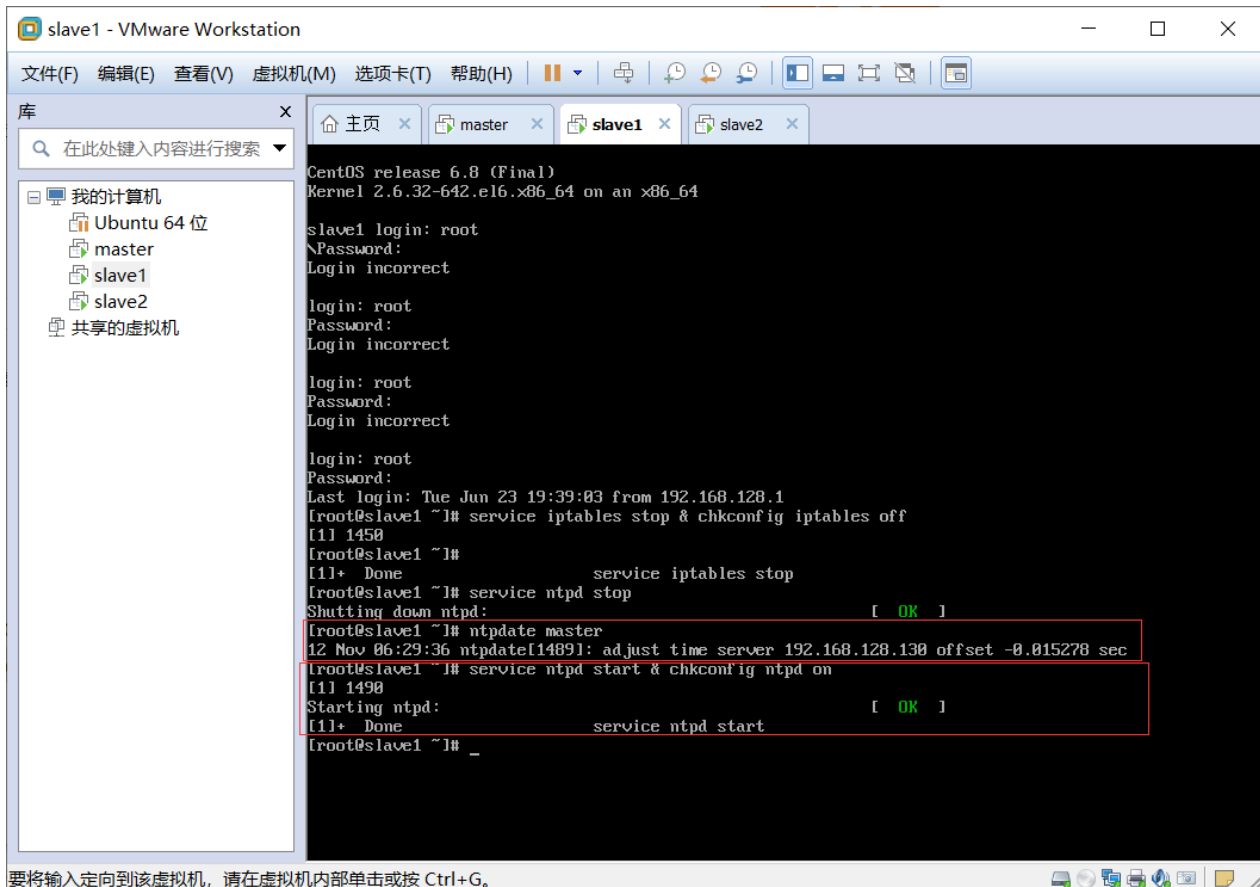
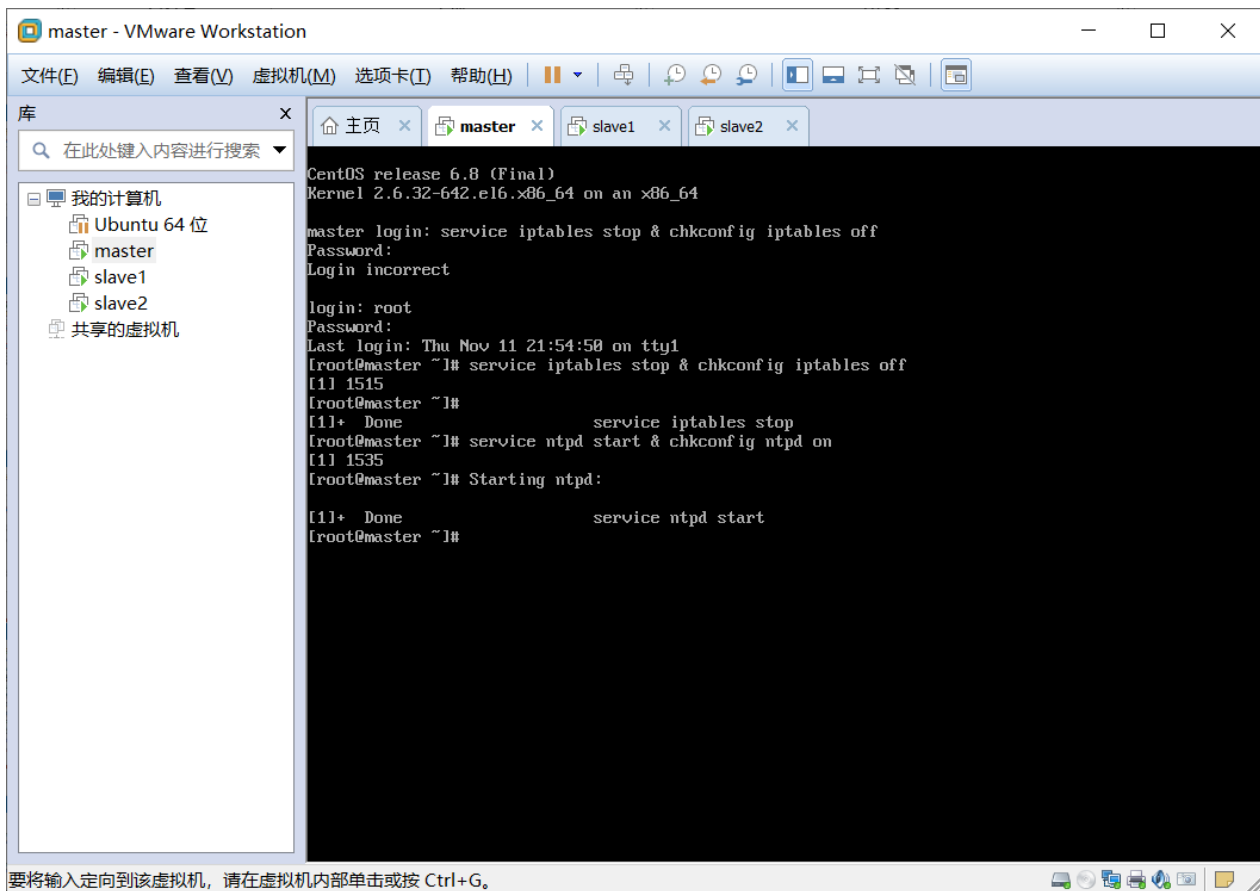
•时间同步

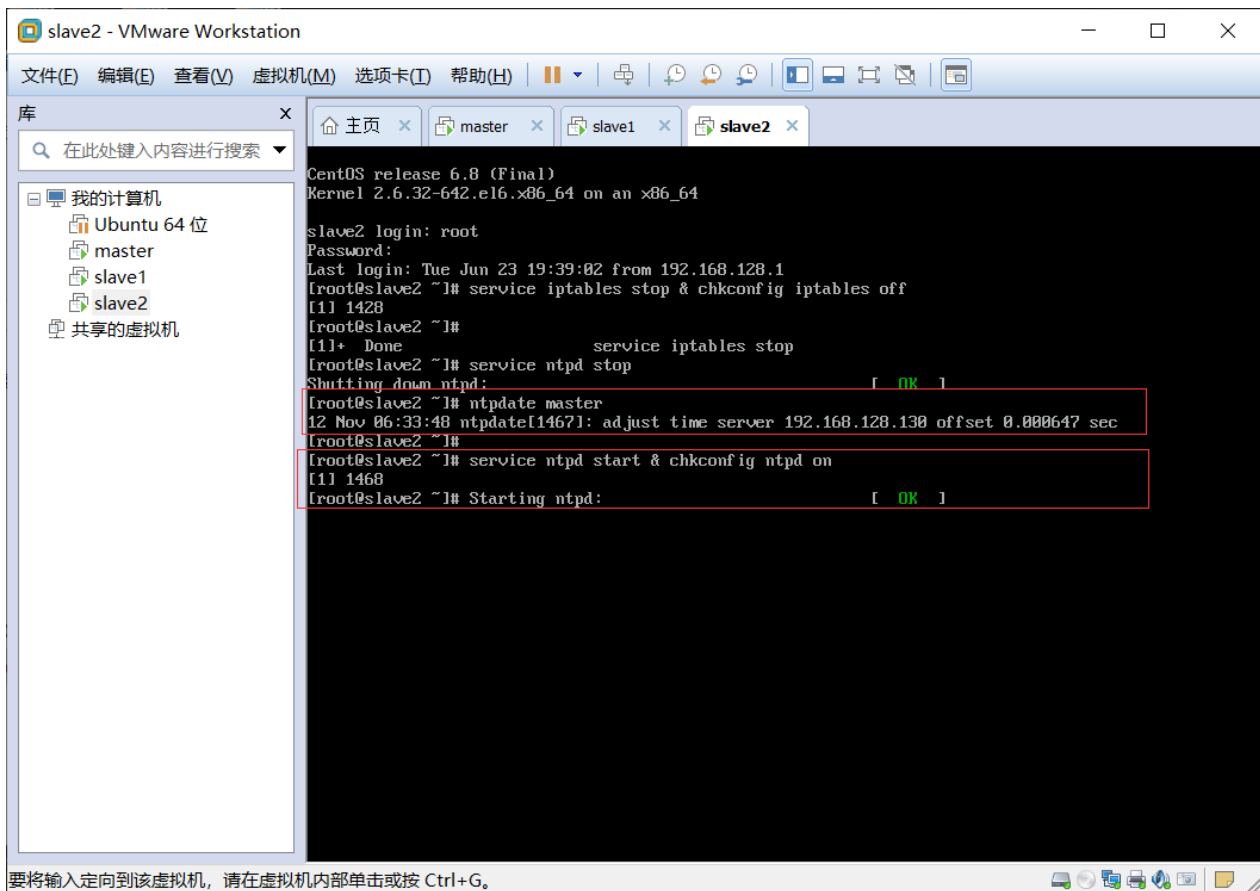
-在所有节点中关闭防火墙: 【service iptables stop & chkconfig iptables off】

-在master节点上启动ntp服务: 【service ntpd start & chkconfig ntpd on】

•注意: 启动后点击一下回车完成

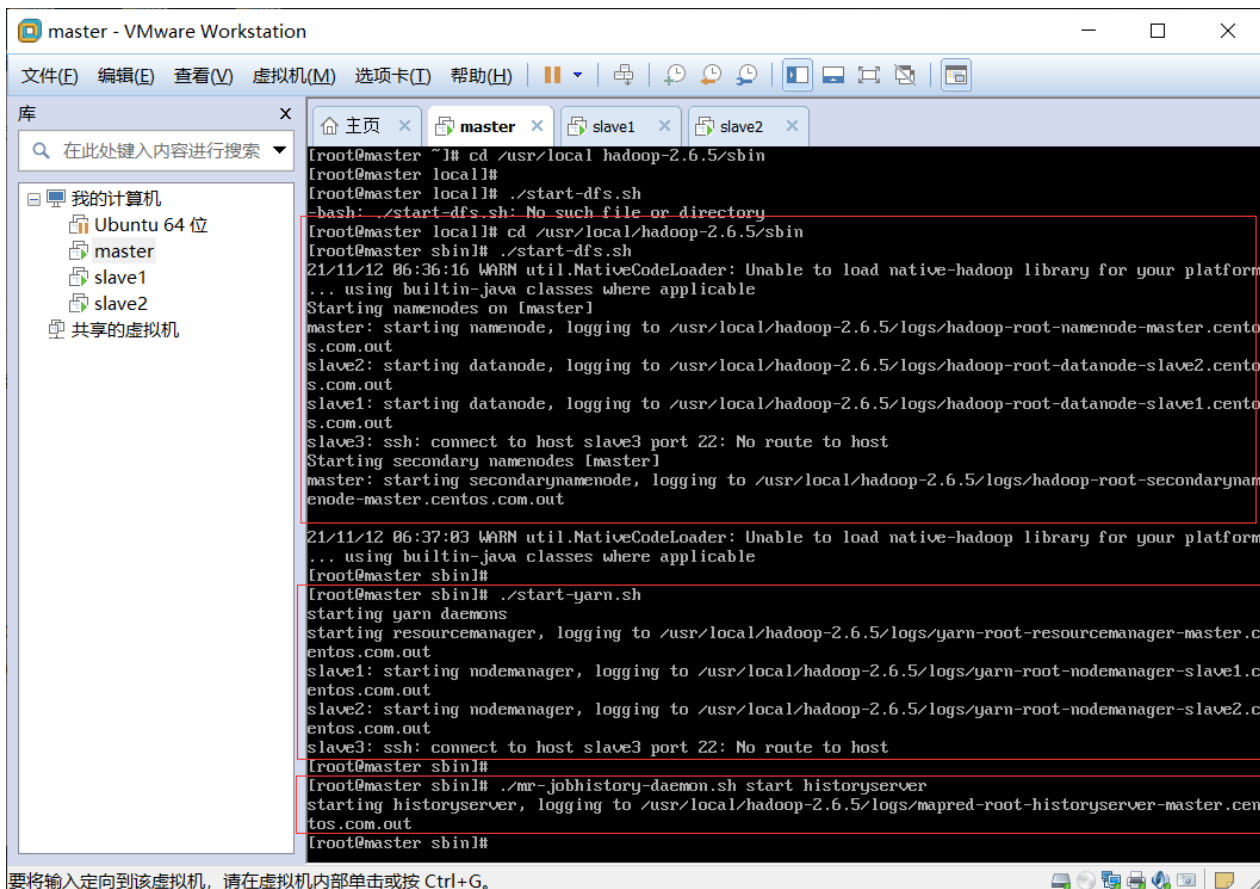
-在2 个子节点执行命令【service ntpd stop】、【ntpddate master】、【service ntpd start & chkconfig ntpd on】启动与主节点 master 的时间同步服务





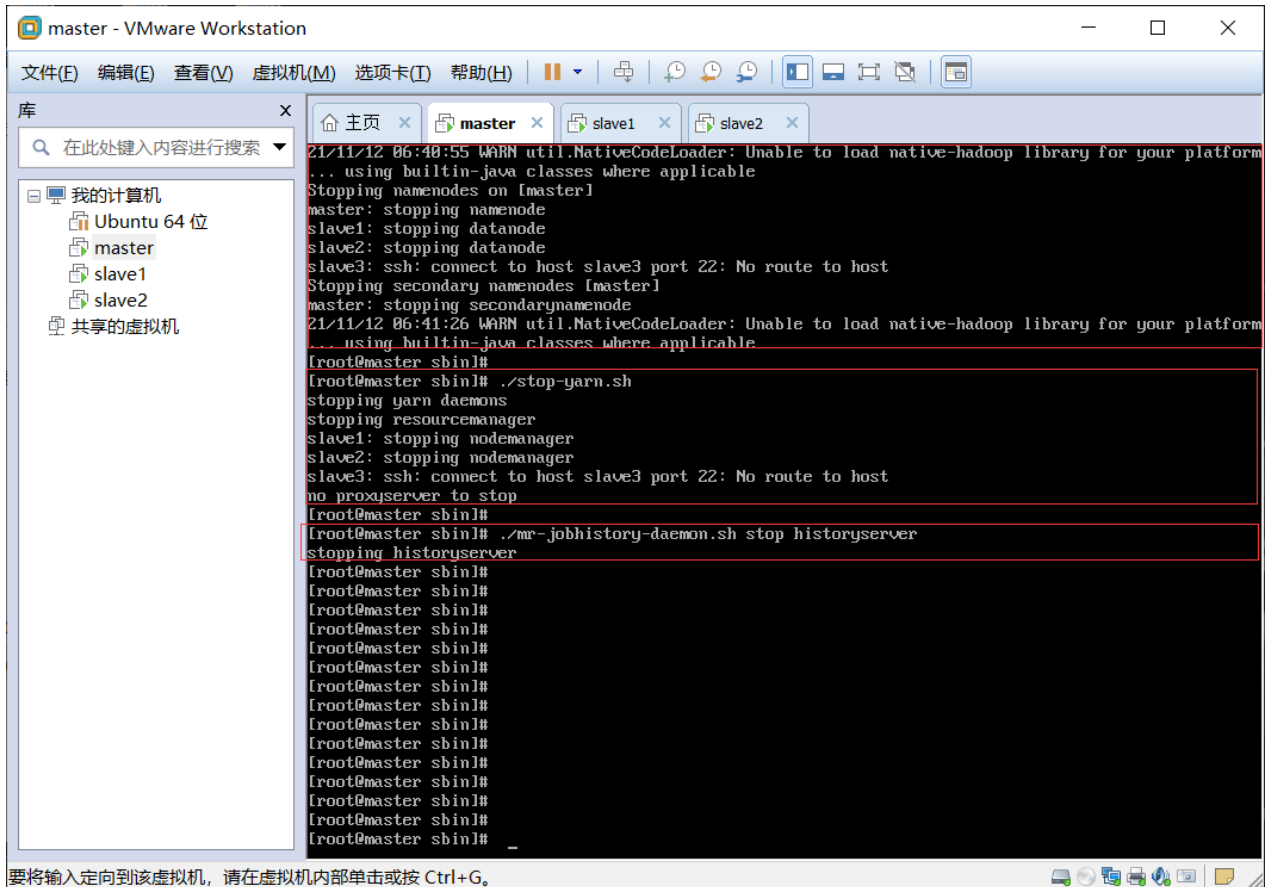
•启动集群

-在主节点执行命令【cd /usr/local/hadoop-2.6.5/sbin】进入目录, 执行命令【./start-dfs.sh】、【./start-yarn.sh】、【./mr-jobhistory-daemon.sh start historyserver】启动集群



•关闭

-在主节点执行命令【cd /usr/local/hadoop-2.6.5/sbin】进入目录，执行命令【./stop-dfs.sh】、【./stop-yarn.sh】、【./mr-jobhistory-daemon.sh stop historyserver】关闭集群



•查看节点进程

-在 3 个节点分别执行命令【jps】查看节点进程

发现在开启HDFS和YARN的情况下有三个节点，关闭服务之后只留下一个节点（本服务器）

