中国矿业大学计算机学院

**2022级本科生课程设计报告**

实验名称 实验一：云计算实现大数据

报告时间 2025年6月

学生姓名 杨晓琦

学 号 08222213

专 业 计算机科学与技术

任课教师 徐东红

目 录

[1 准备华为云环境 3](#_Toc201480914)

[1.1购买华为云ECS 3](#_Toc201480915)

[1.2购买OBS 6](#_Toc201480916)

[2 搭建Hadoop集群 10](#_Toc201480917)

[2.1Hadoop集群搭建 10](#_Toc201480918)

[**2.1.1配置ECS** 10](#_Toc201480919)

[**2.1.2安装OpenJDK** 22](#_Toc201480920)

[2.2搭建Hadoop集群 28](#_Toc201480921)

[**2.2.1搭建Hadoop集群** 28](#_Toc201480922)

[**2.2.2 测试与OBS互联** 35](#_Toc201480923)

[3 MapReduce 36](#_Toc201480924)

[3.1 MapReduce存算分离 36](#_Toc201480925)

[**3.1.1 测试Hadoop集群功能** 36](#_Toc201480926)

[4 释放云服务资源 37](#_Toc201480927)

[4.1 释放弹性云服务器ECS 37](#_Toc201480928)

[4.2删除桶OBS 38](#_Toc201480929)

实验一：云主机实现大数据

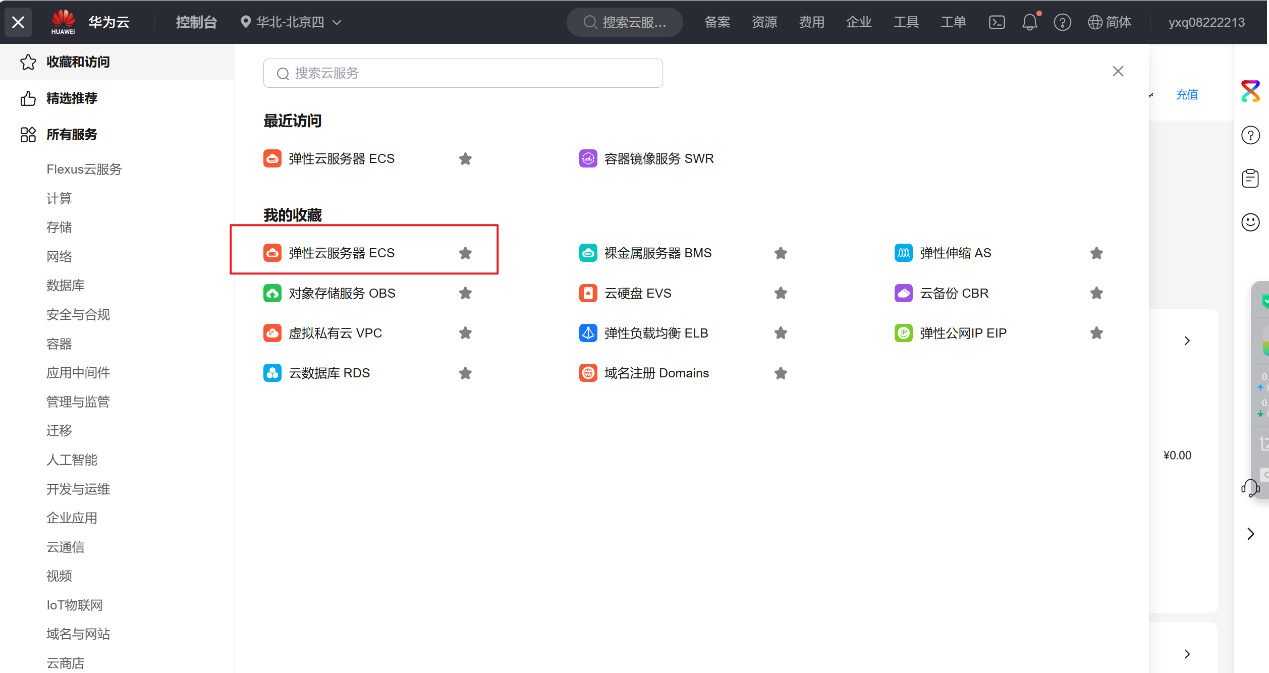
1 准备华为云环境

1.1购买华为云ECS

步骤1 登录华为云

打开华为云地址：https://www.huaweicloud.com/，点击“登录”，输入用户名、密码

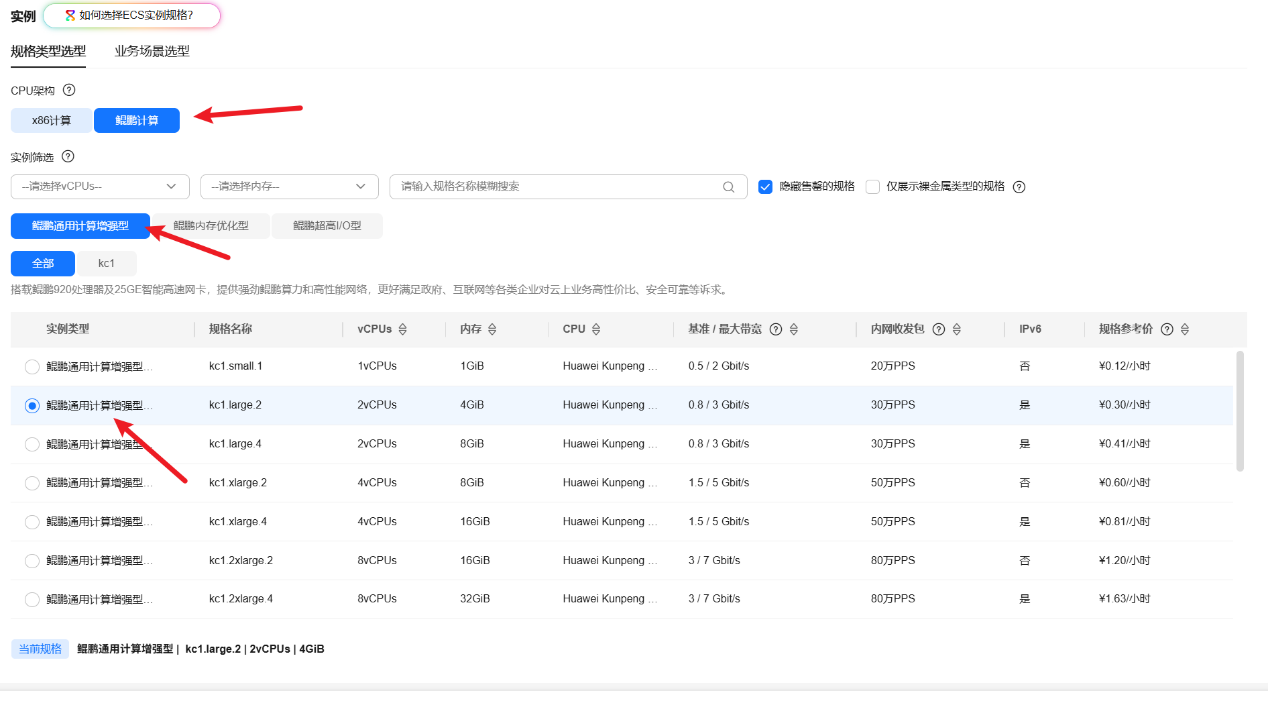
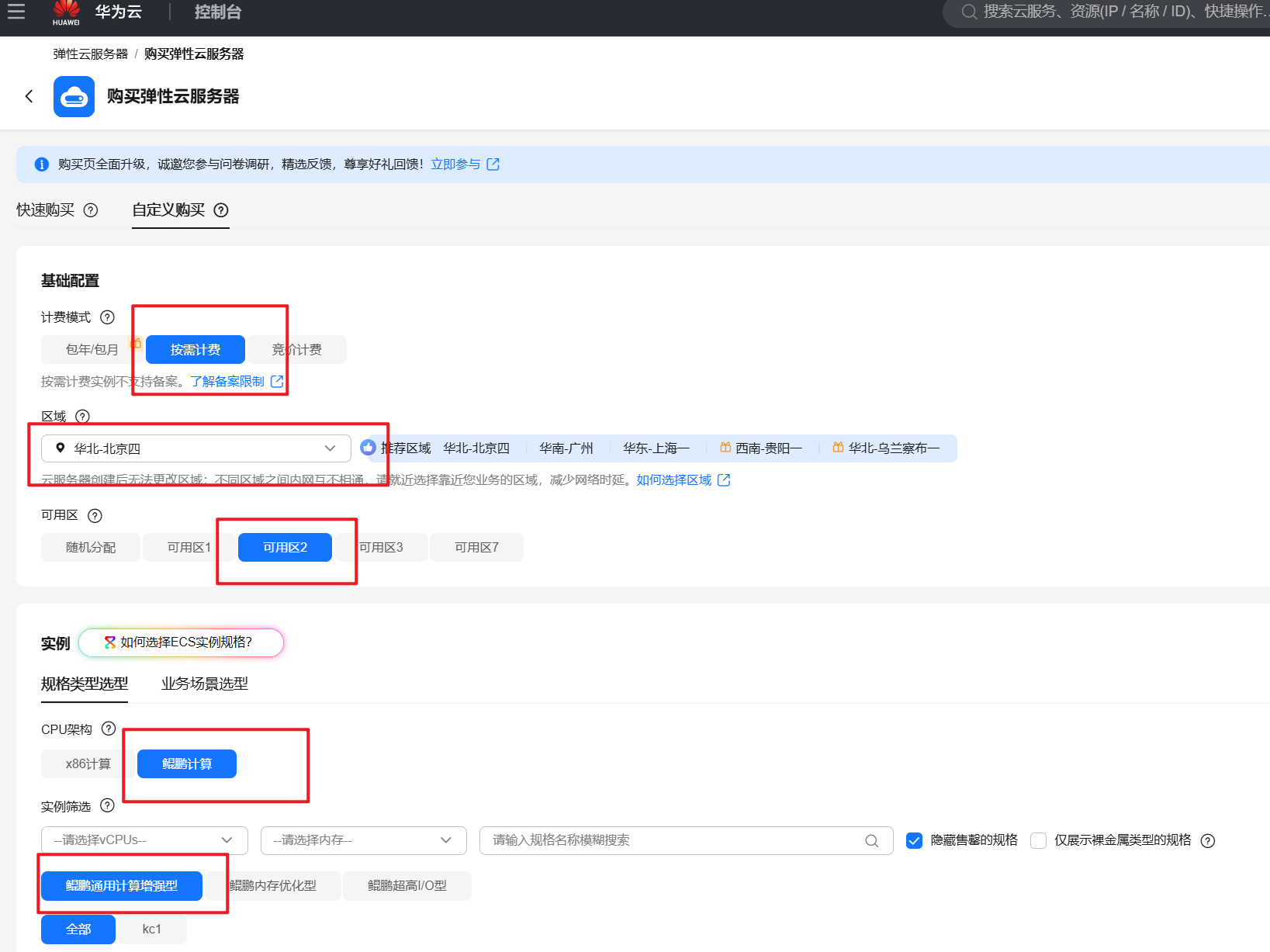
步骤2 点击“控制台”， 选择“弹性云服务器ECS”



步骤3 选择“买弹性云服务器ECS”

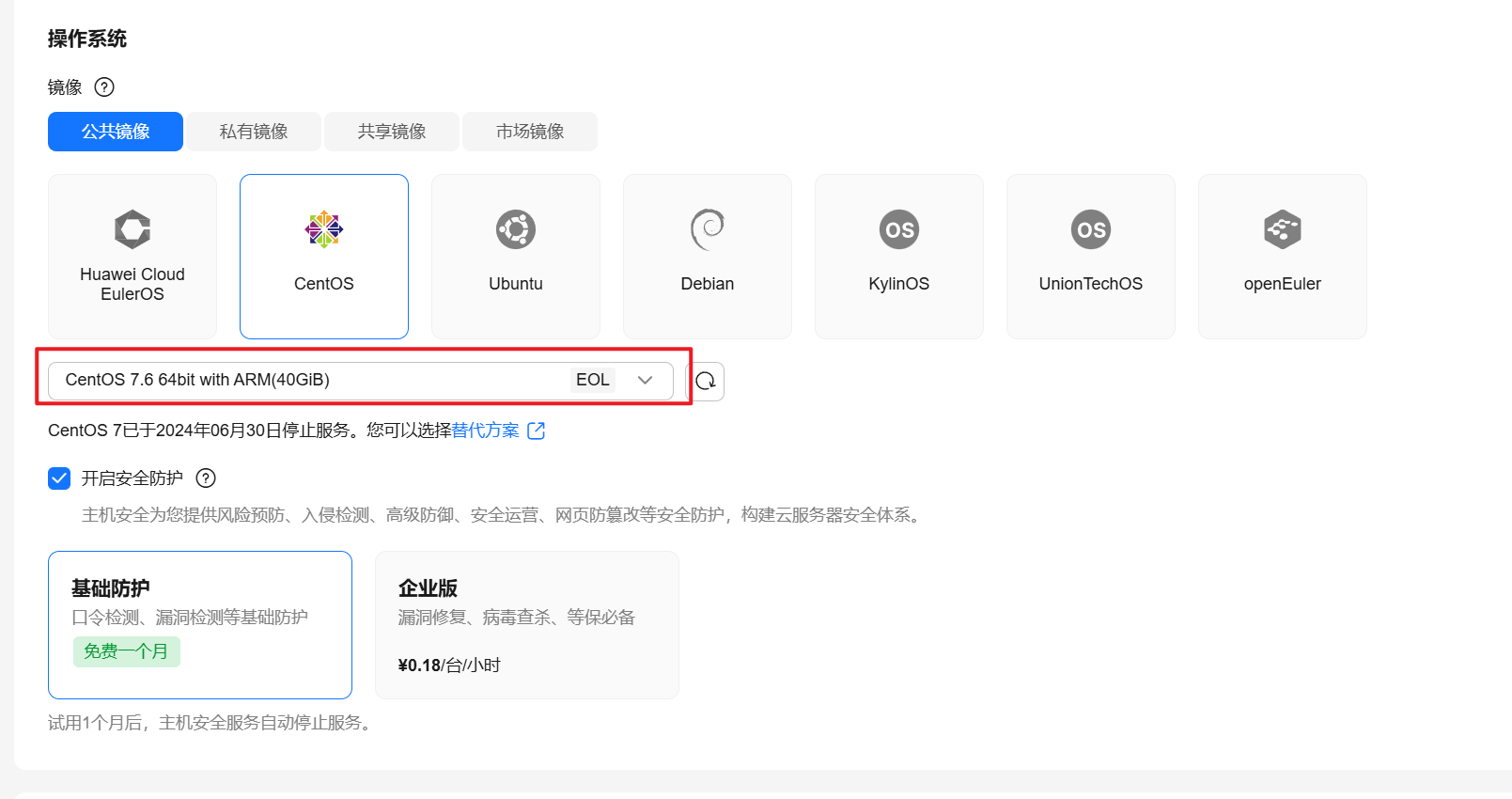


选择“按需计费”，“可用区2”，CPU架构“鲲鹏计算”，选择“鲲鹏通用计算增强型”，2vCPUs|4GB



步骤4 配置操作系统和磁盘

选择“公共镜像”，CentOS7.6，系统盘建议配置40GB，购买数量4台，点击“网络配置”



配置网络，网络选择“vpc-default”，安全组选择“Sys-default”，“现在购买”，选择“全动态BGP”，“按流量计算”，“5M”，点击“高级配置”

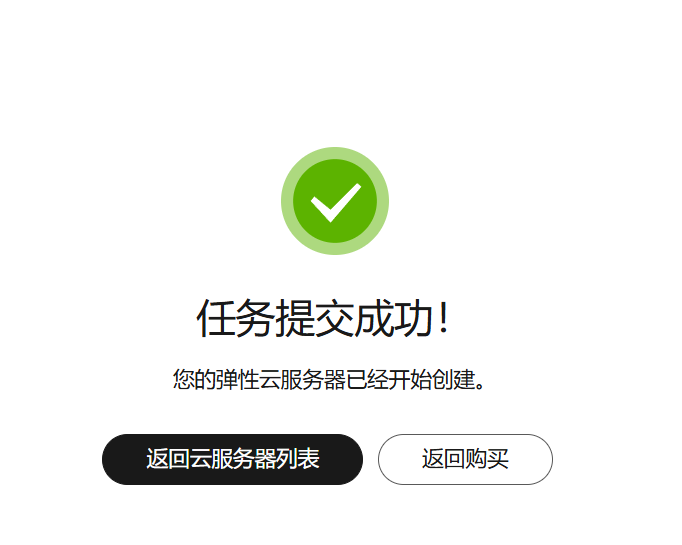


步骤6 配置密码

自定义云服务器名称，自行设置root登录密码，云备份选择“暂不购买”，点击“确认配置”，点中“我已经阅读并同意”，点击“立即购买”



步骤七 弹性云服务器开始创建



1.2购买OBS

步骤1 进入控制台，选择“对象存储服务”



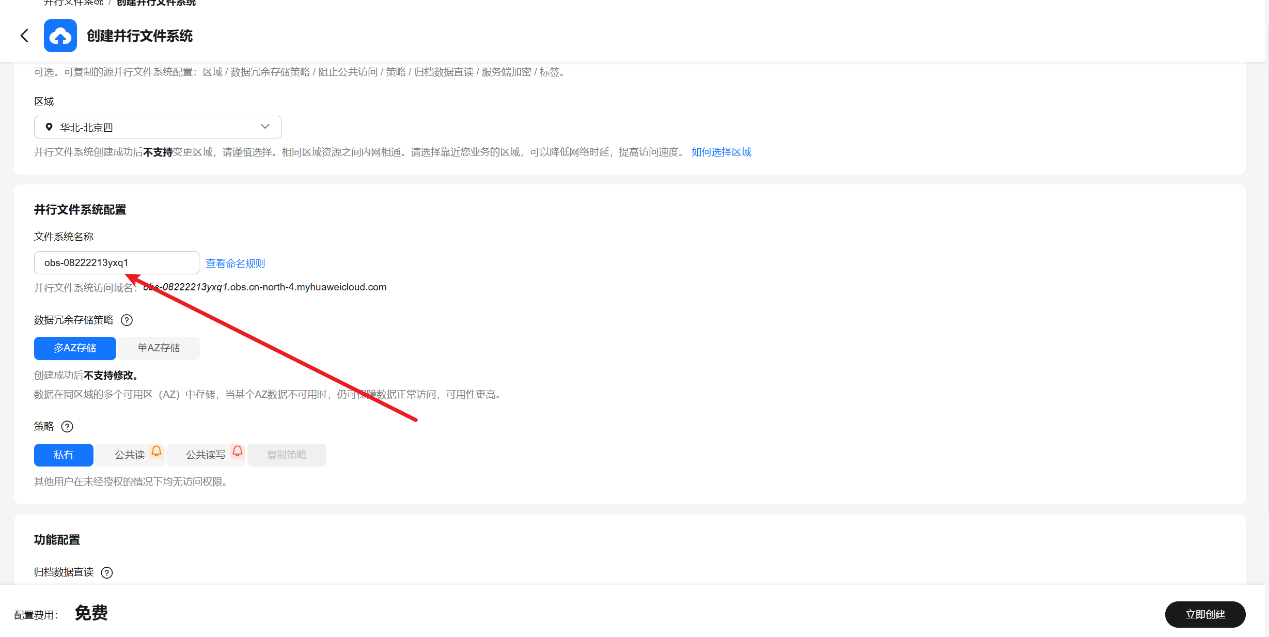
步骤2 选择“创建桶”



步骤3 自定义桶名称，选择“标准存储”，点击“立即创建”

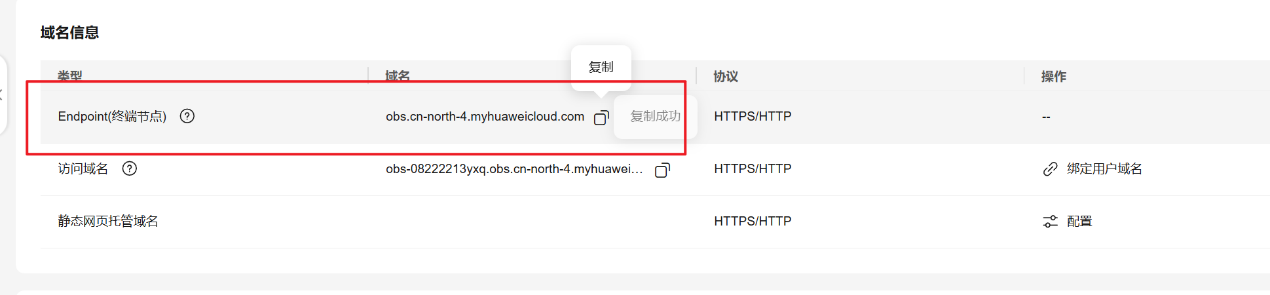


步骤4 选择“创建并行文件系统”，自定义文件系统名称，点击“自定义立即创建”



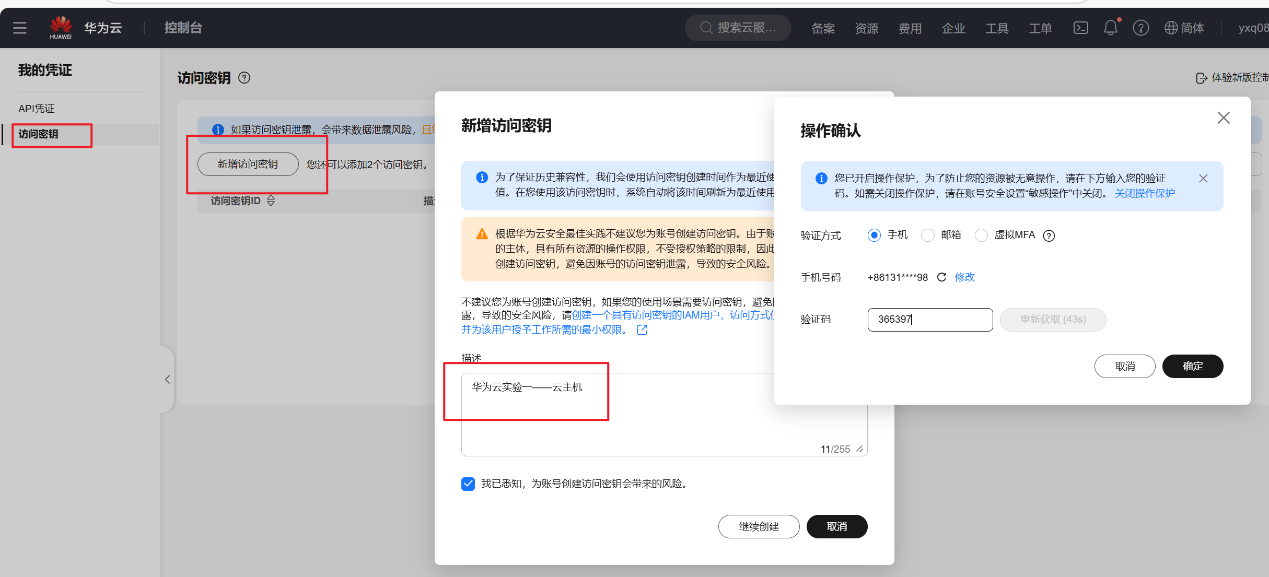
步骤5 进入创建的OBS桶

步骤6 复制该参数，保存到本地文档

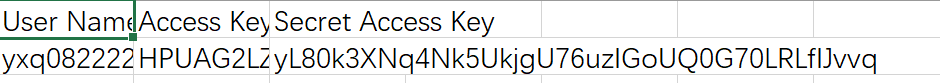


步骤7 获取AK/SK，点击“我的凭证”，选择“访问秘钥”

步骤8 通过手机号码，接收短信验证码，点击“确定”



操作完成后，得到文件，打开即可得到AK/SK

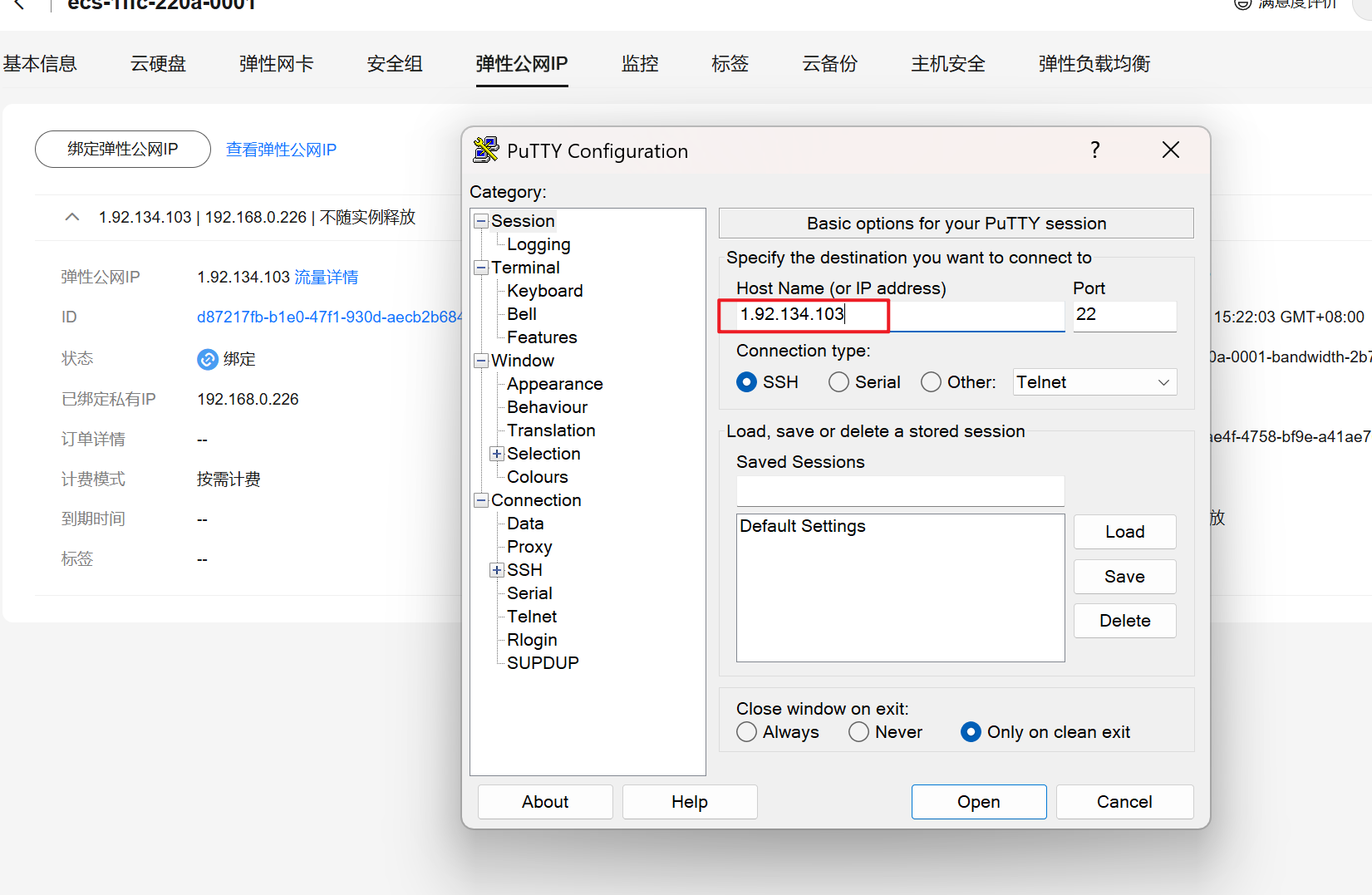


2 搭建Hadoop集群

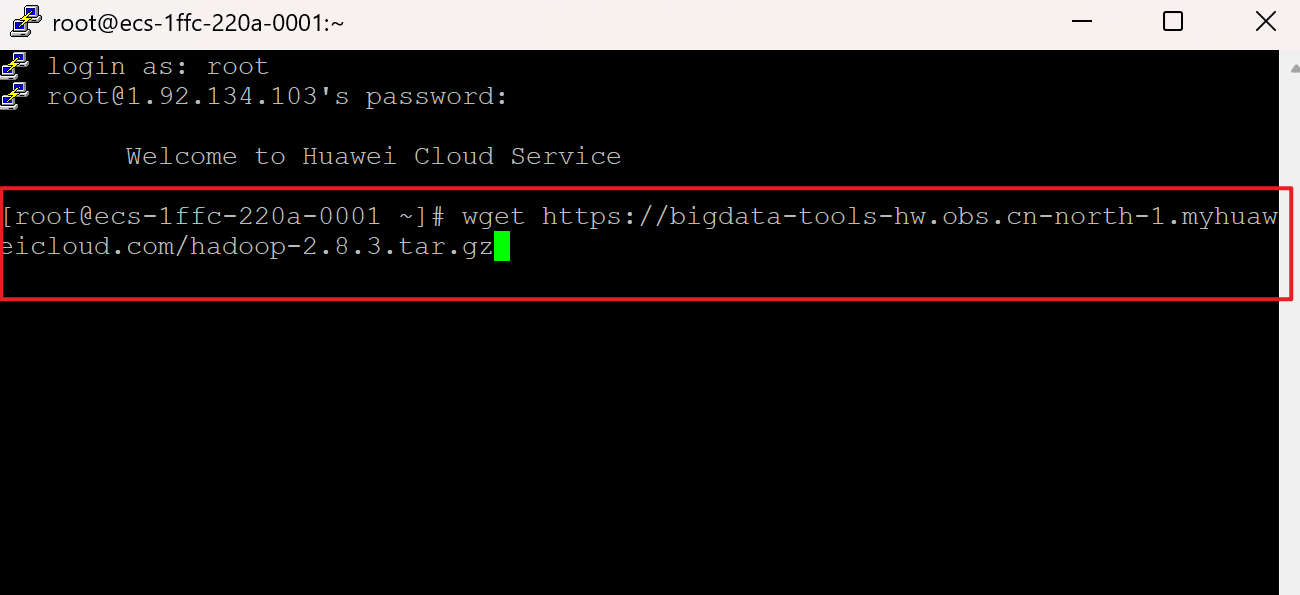
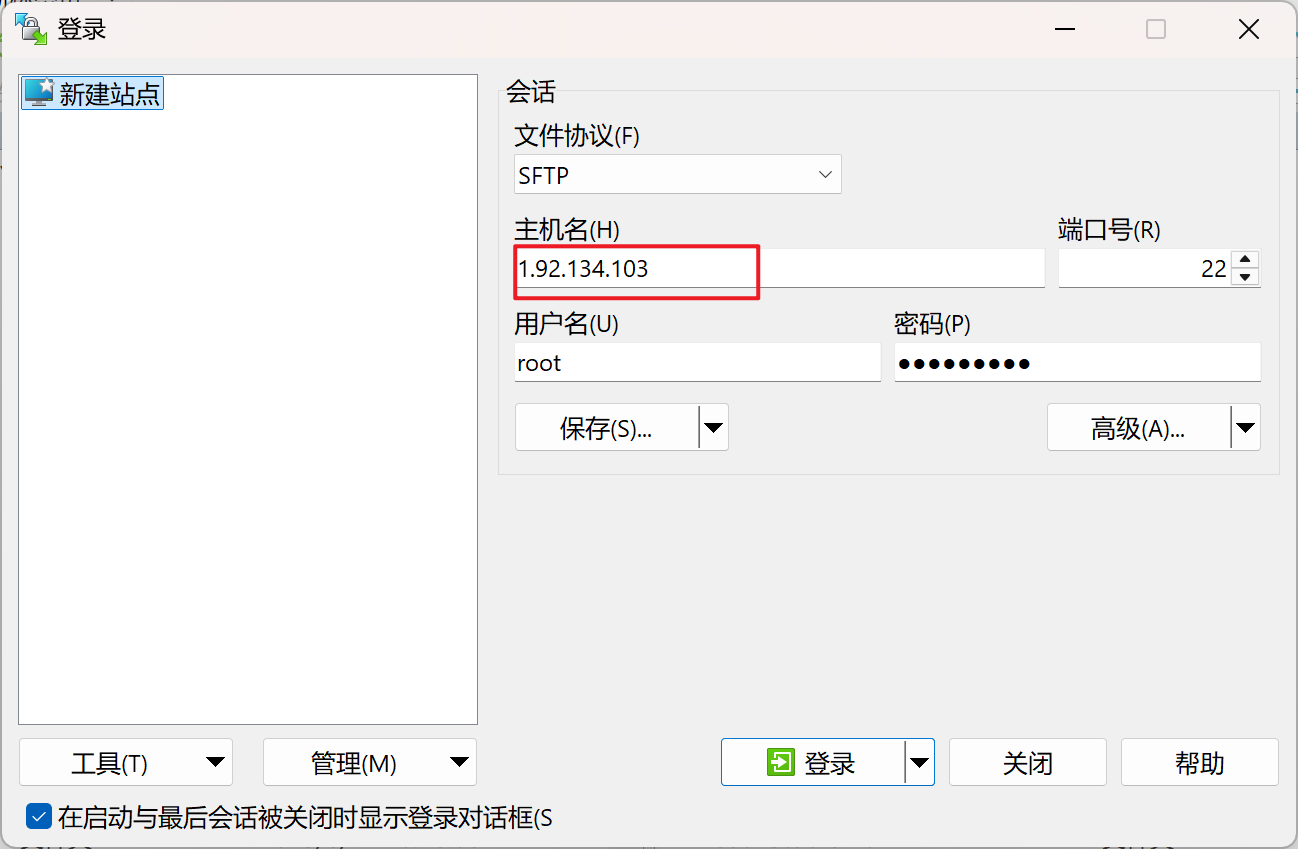
2.1Hadoop集群搭建

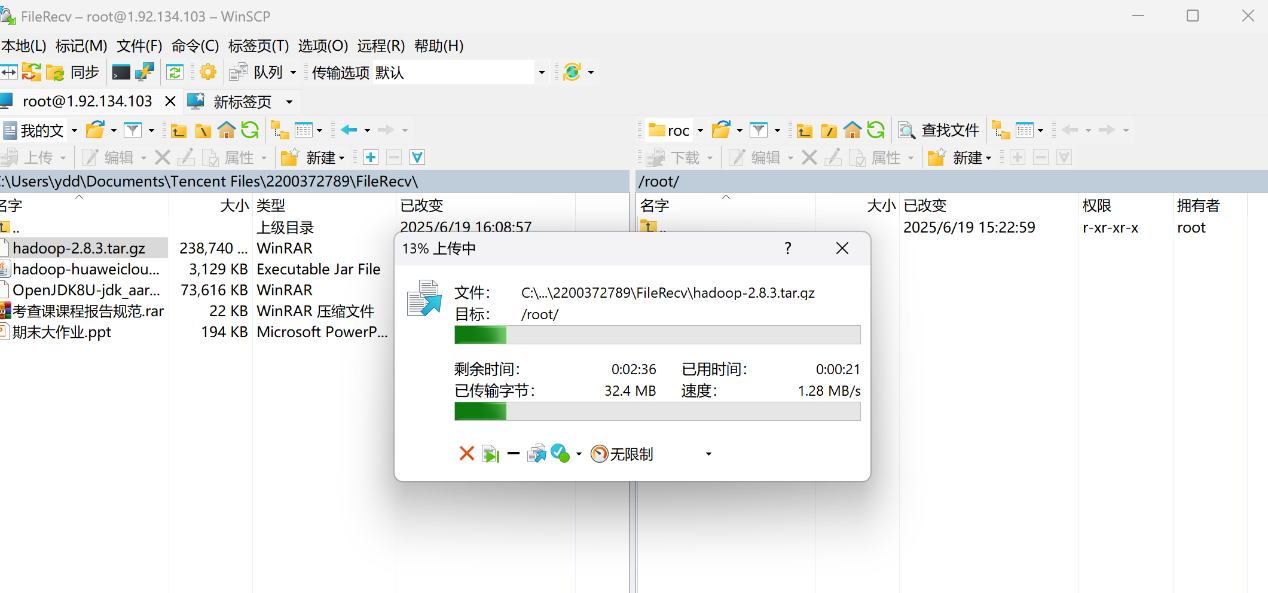
**2.1.1配置ECS**

步骤1 使用putty登录ECS下载Putty工具访问网址https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html，选择putty.exe



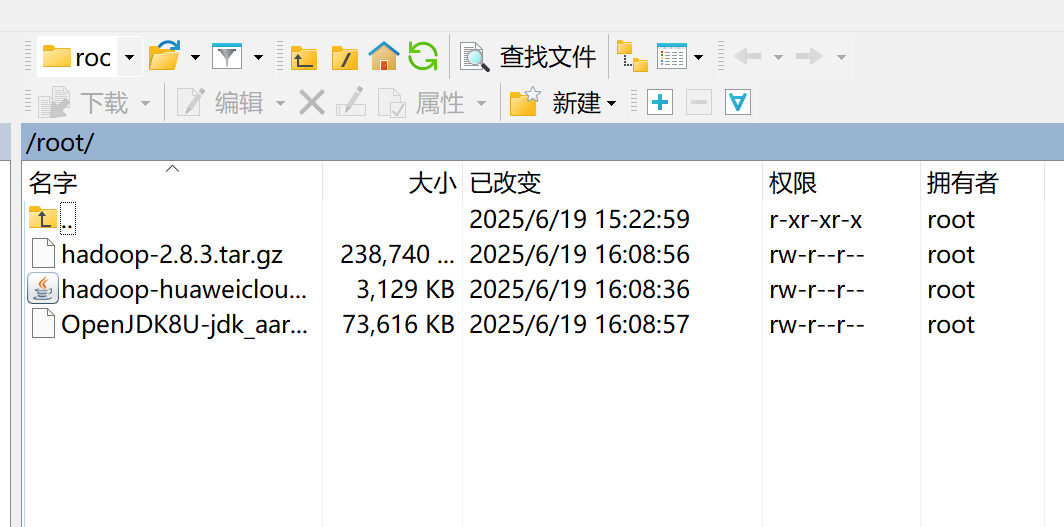
步骤2 node1节点登录成功后执行命令在线下载安装包



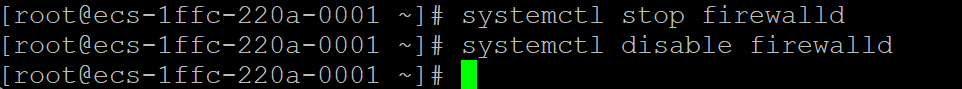


步骤3 下载OBSFileSystem

步骤4 下载OpenJDK

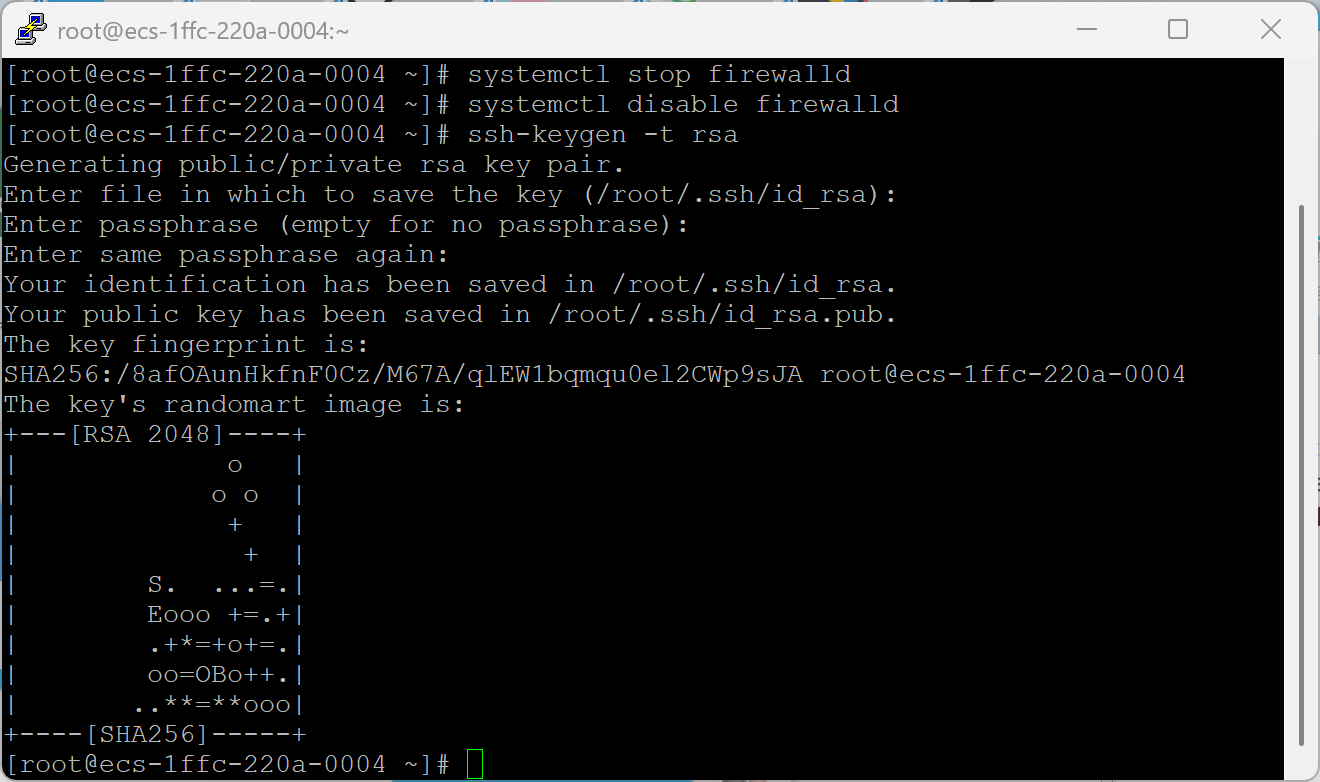
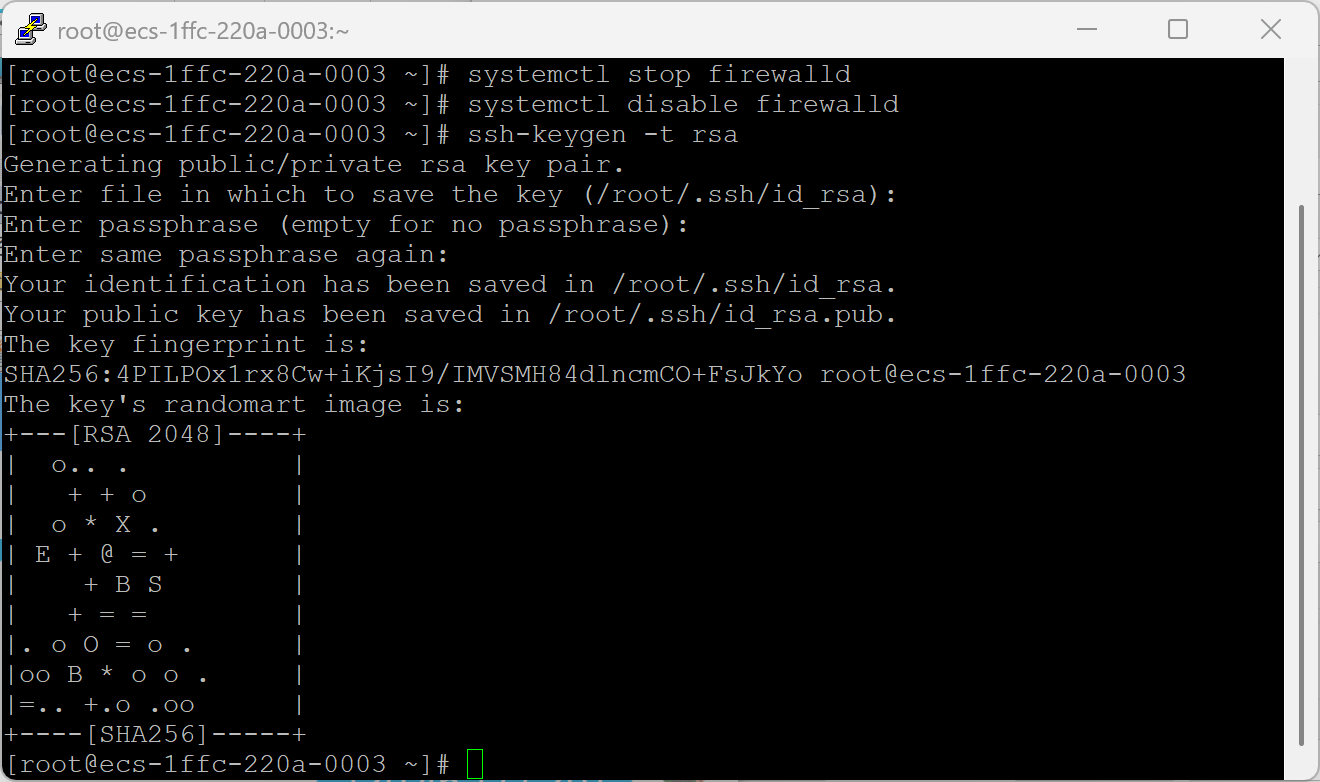
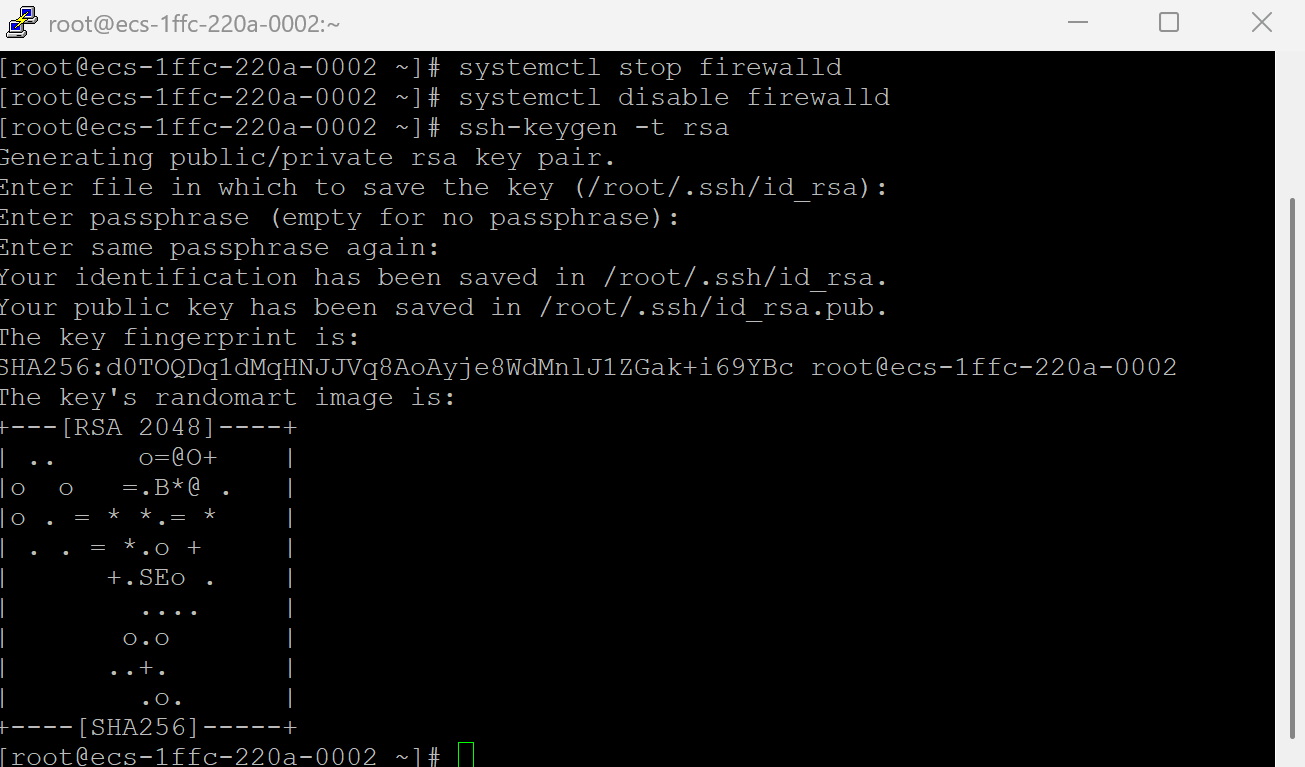
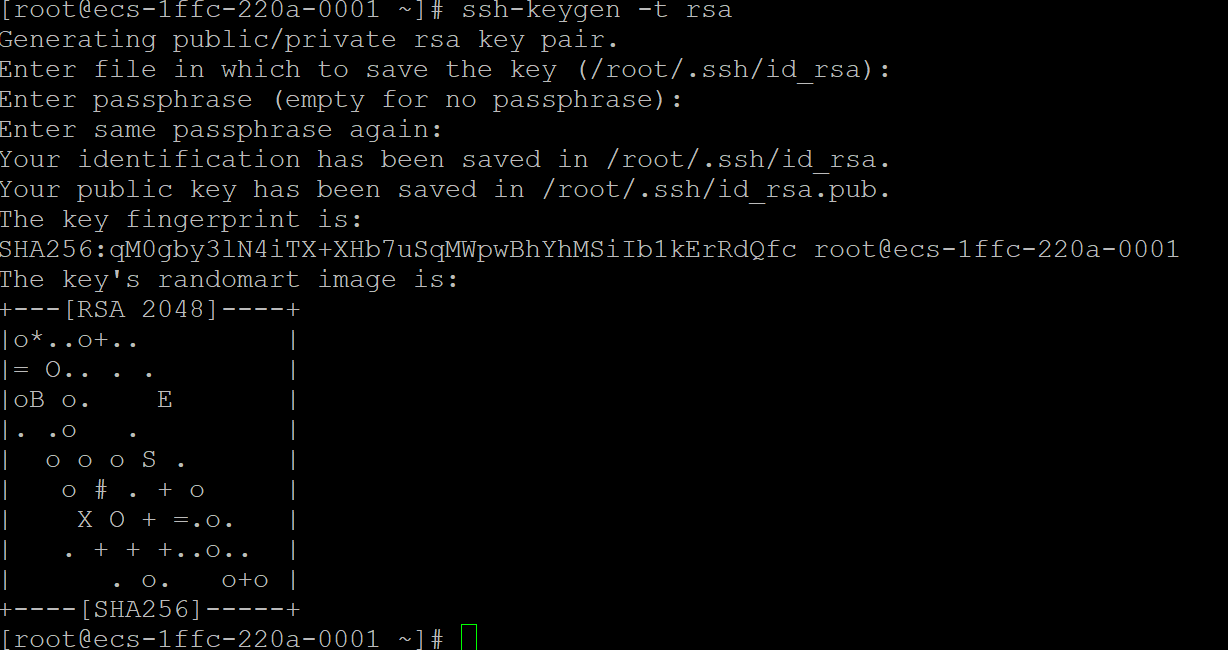


步骤5 关闭防火墙

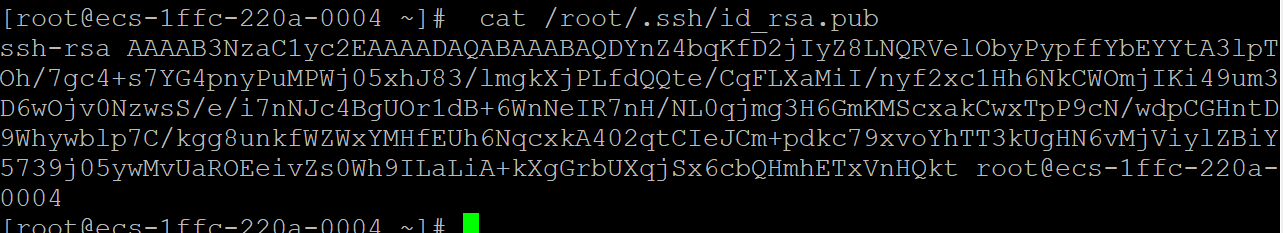
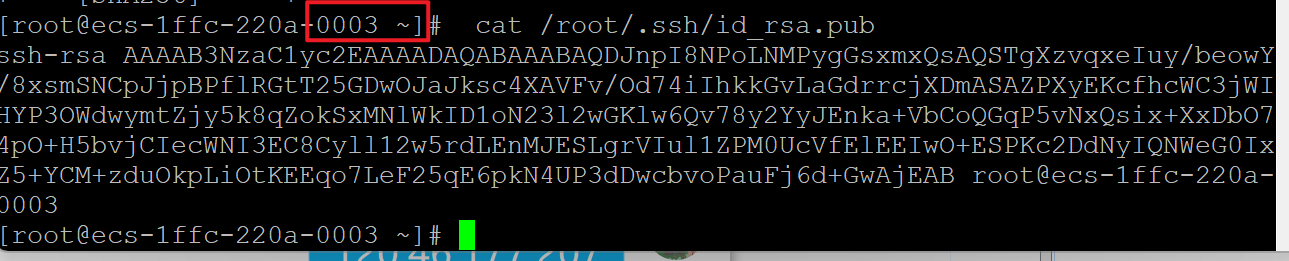
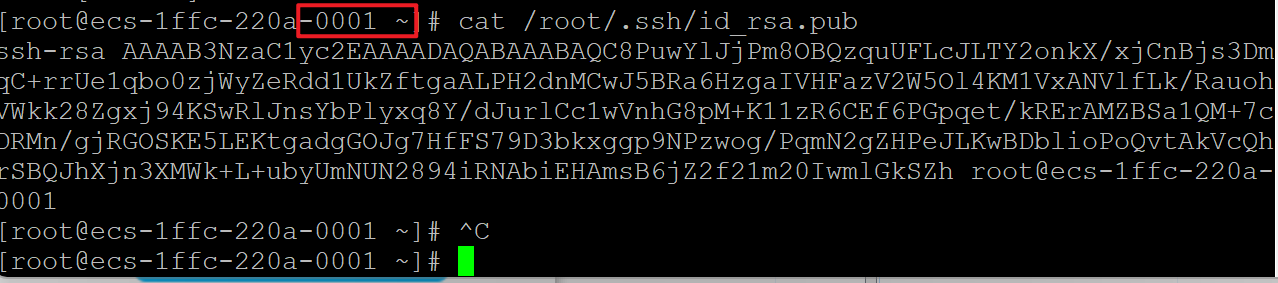


步骤6 配置节点互信

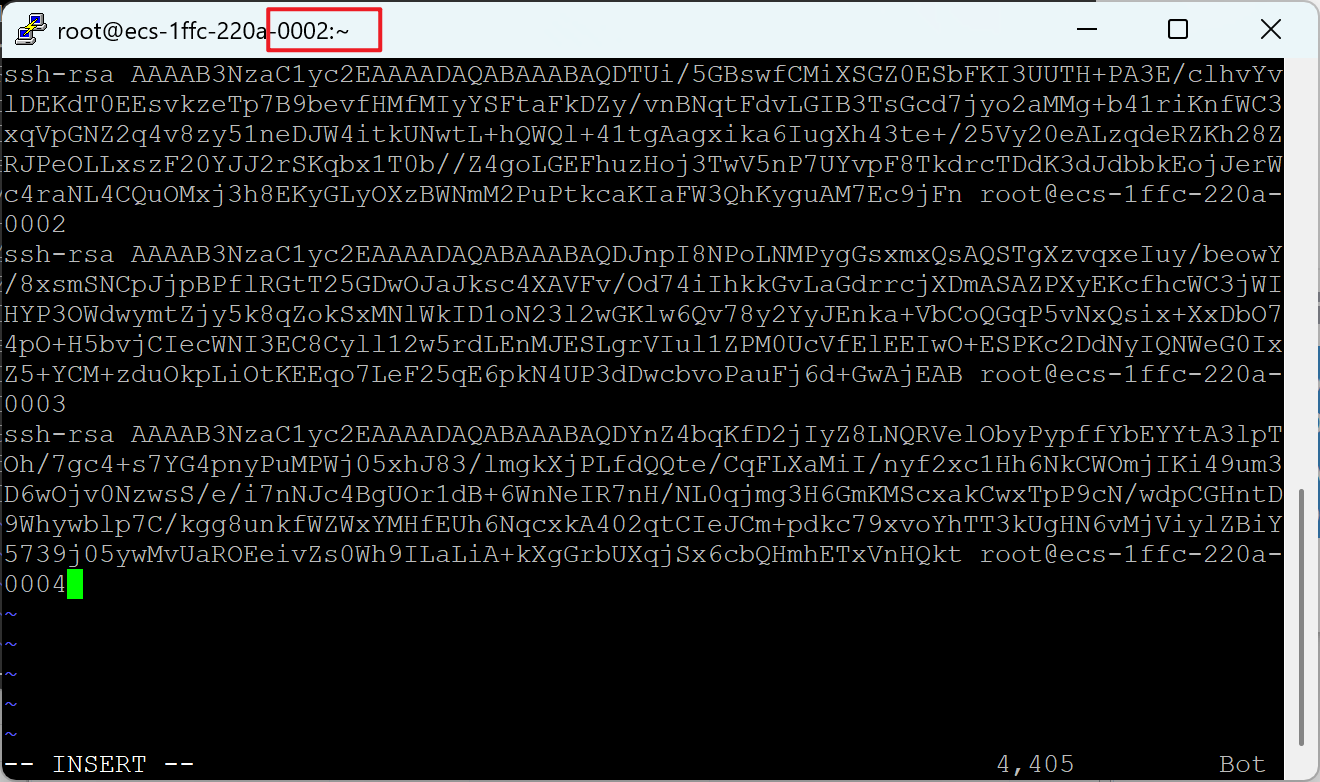
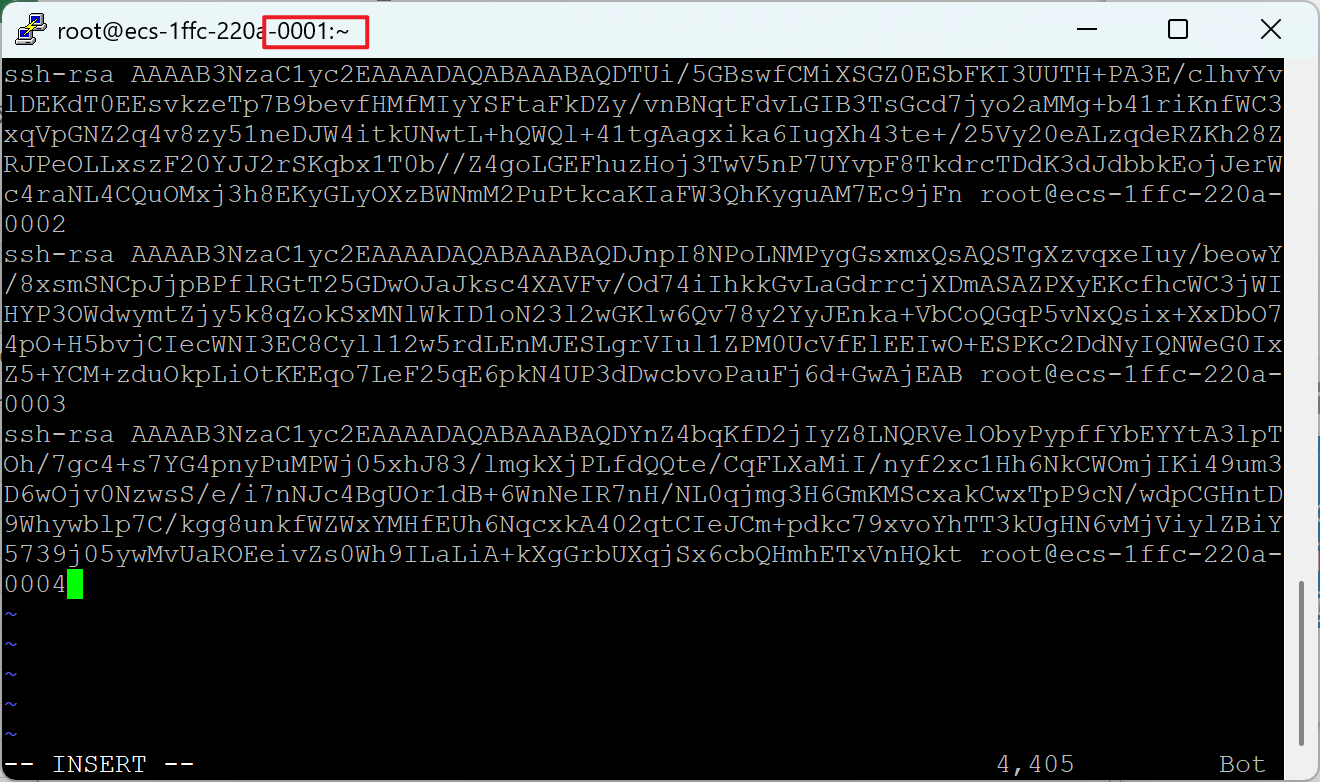
node1~node4节点分别执行如下命令



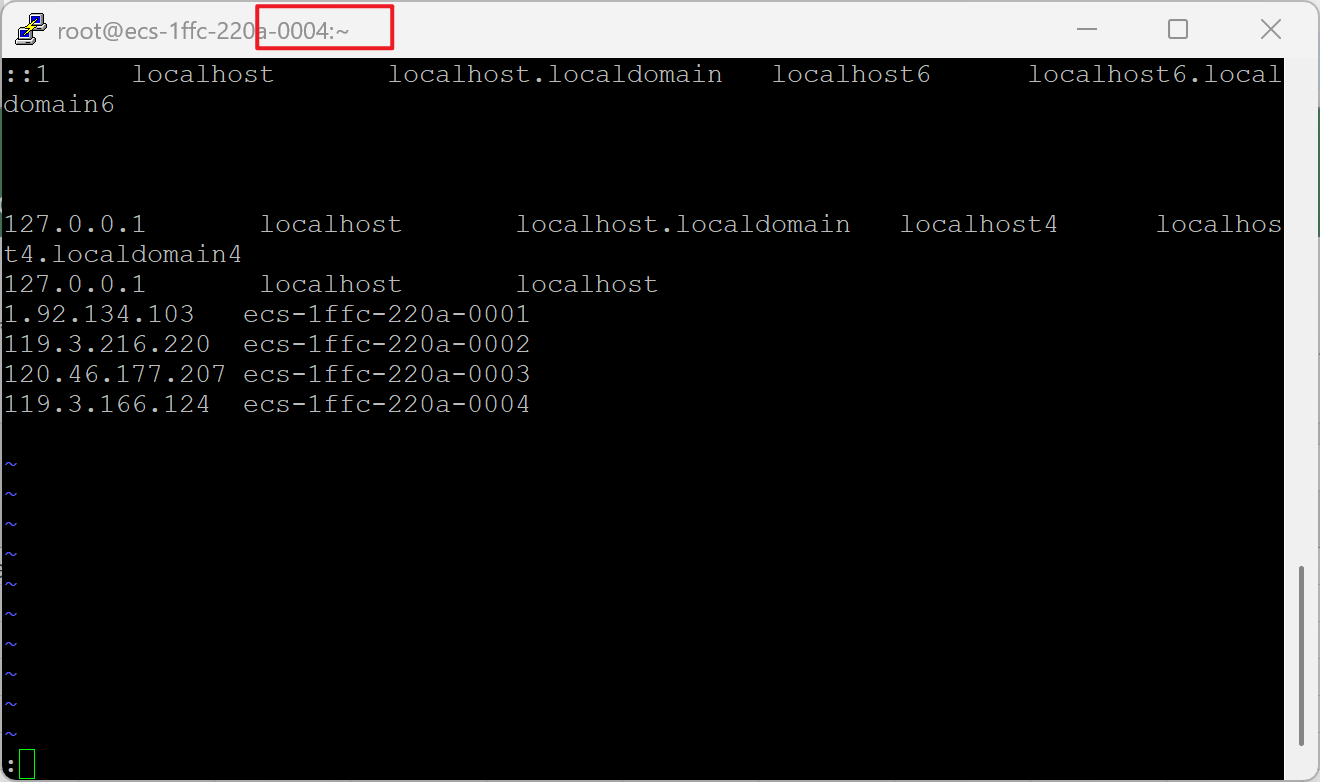
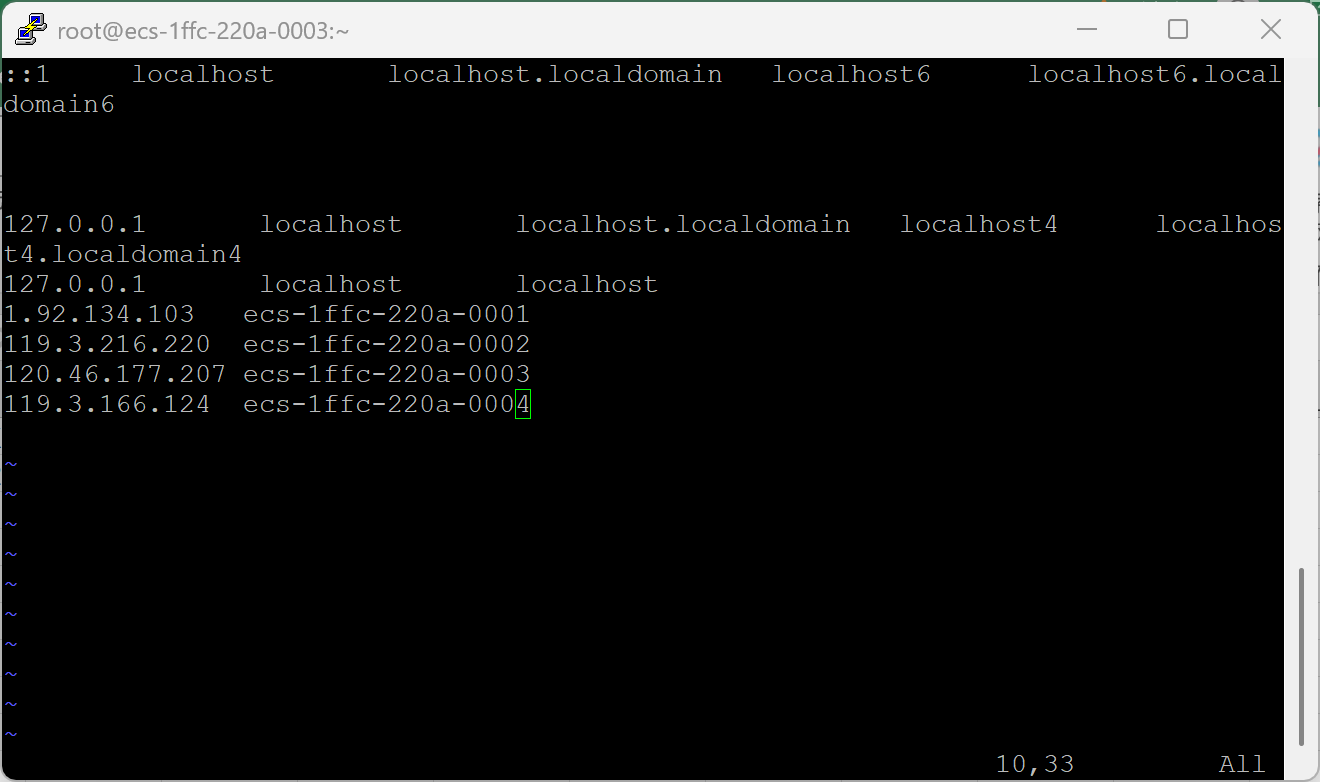
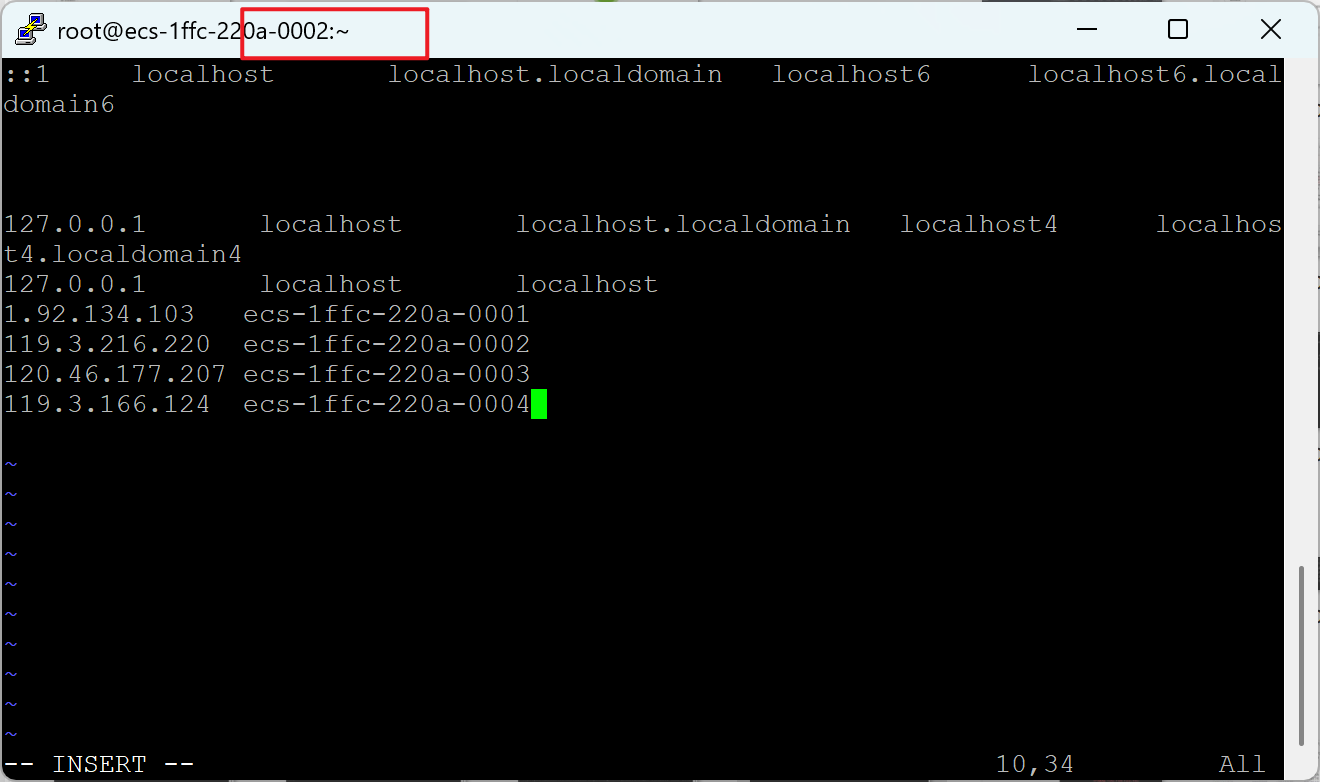
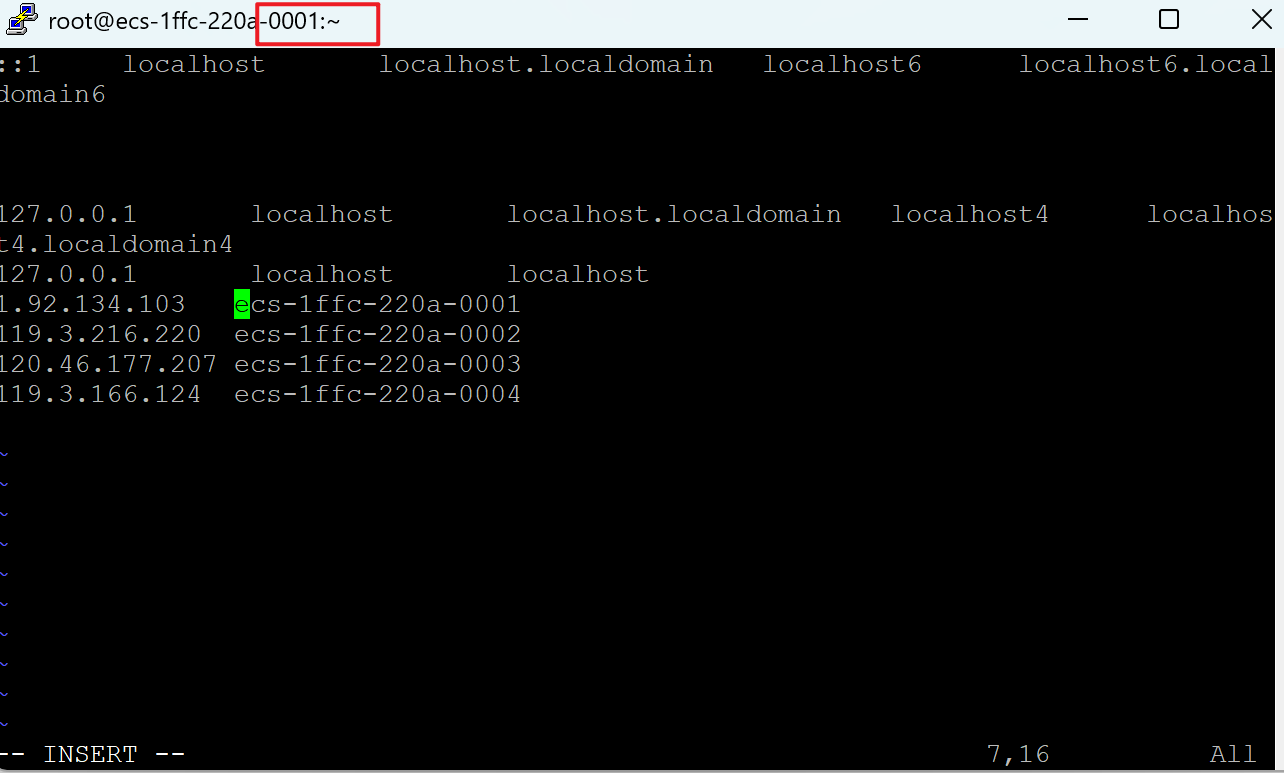
node1~node4节点分别执行命令cat /root/.ssh/id\_rsa.pub命令



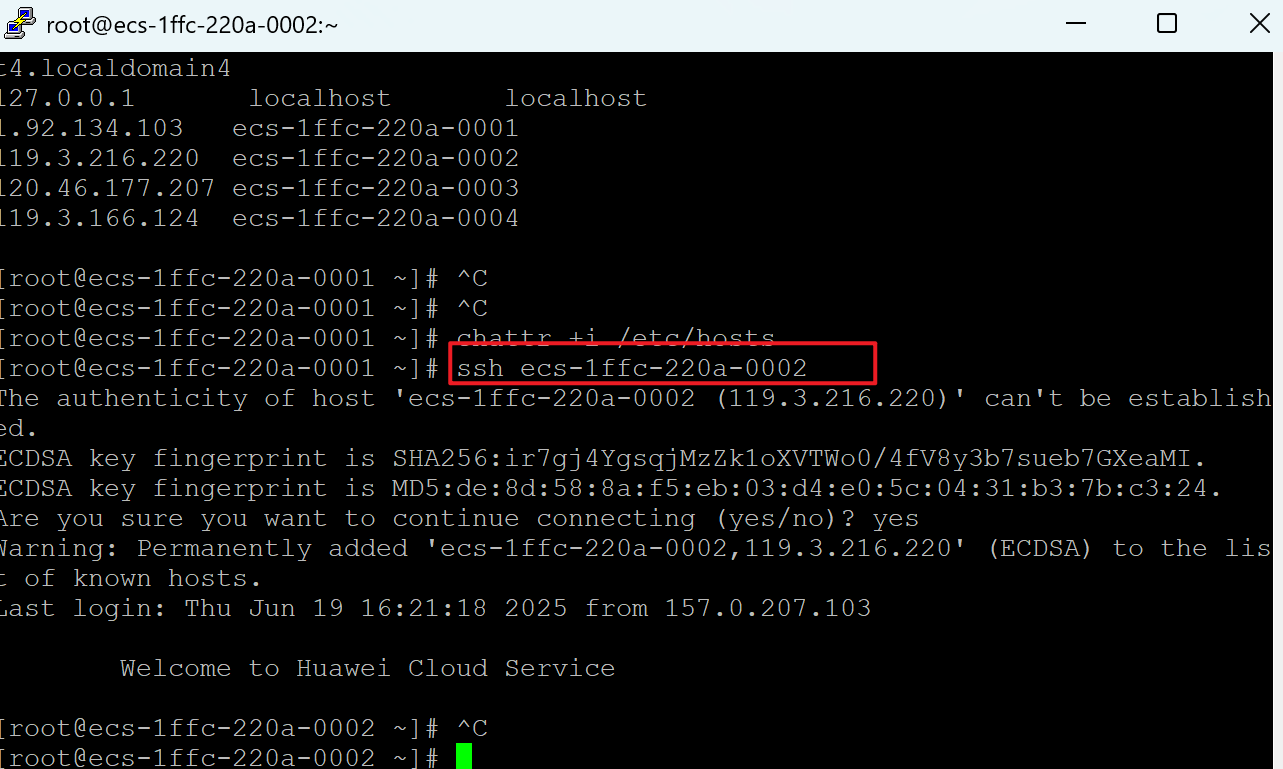
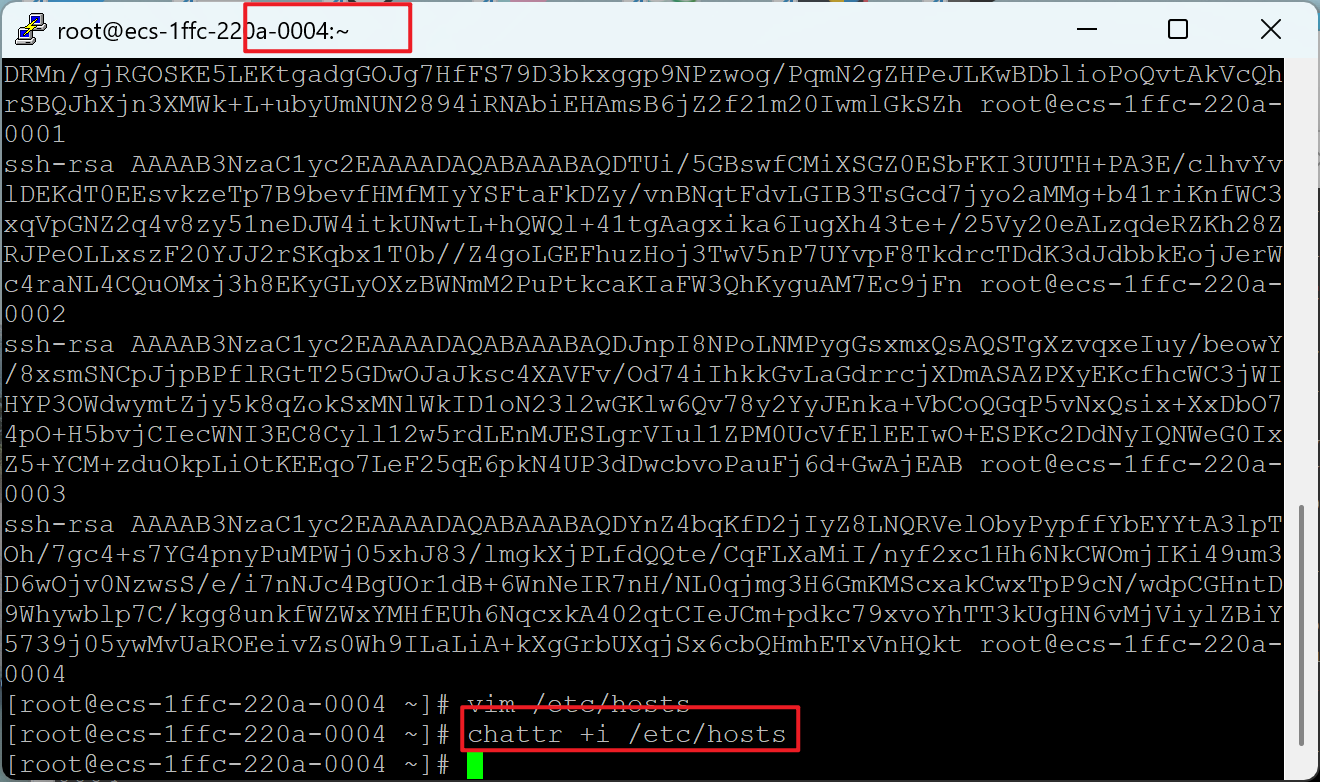
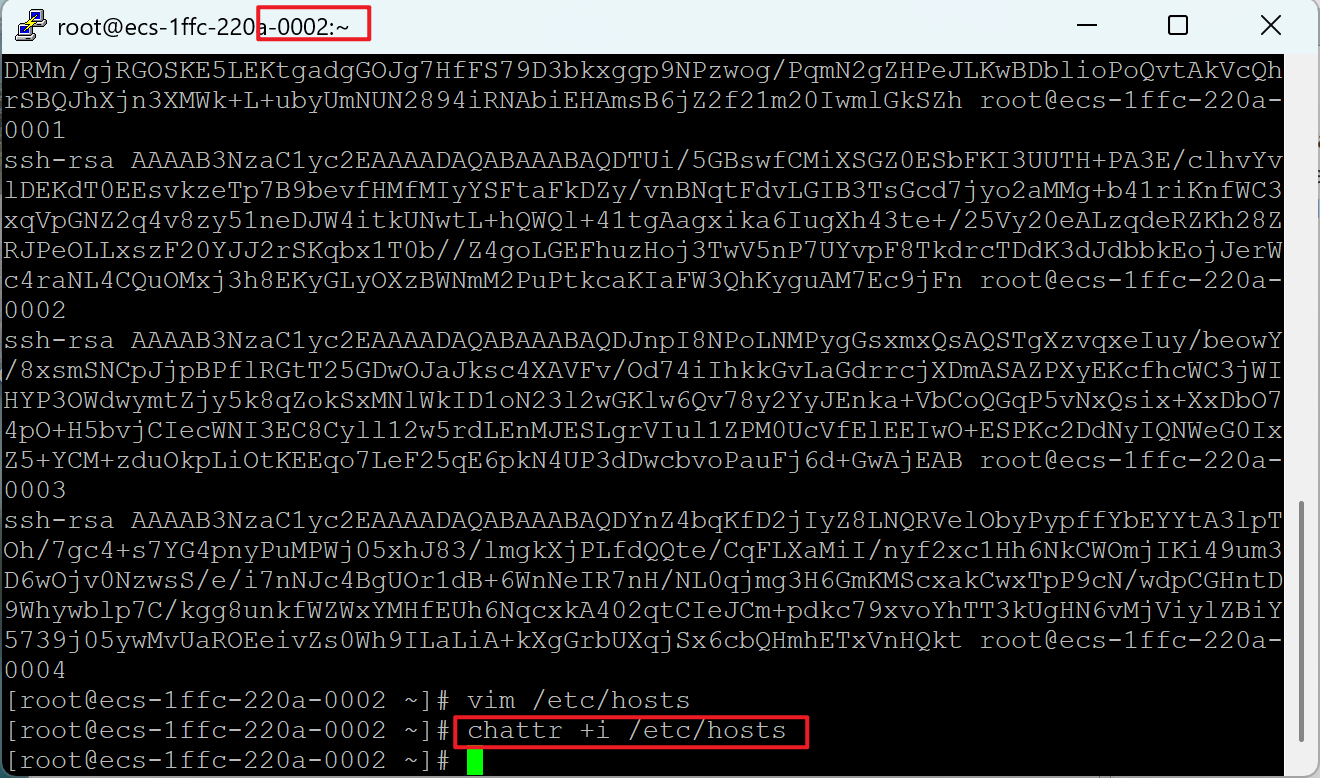
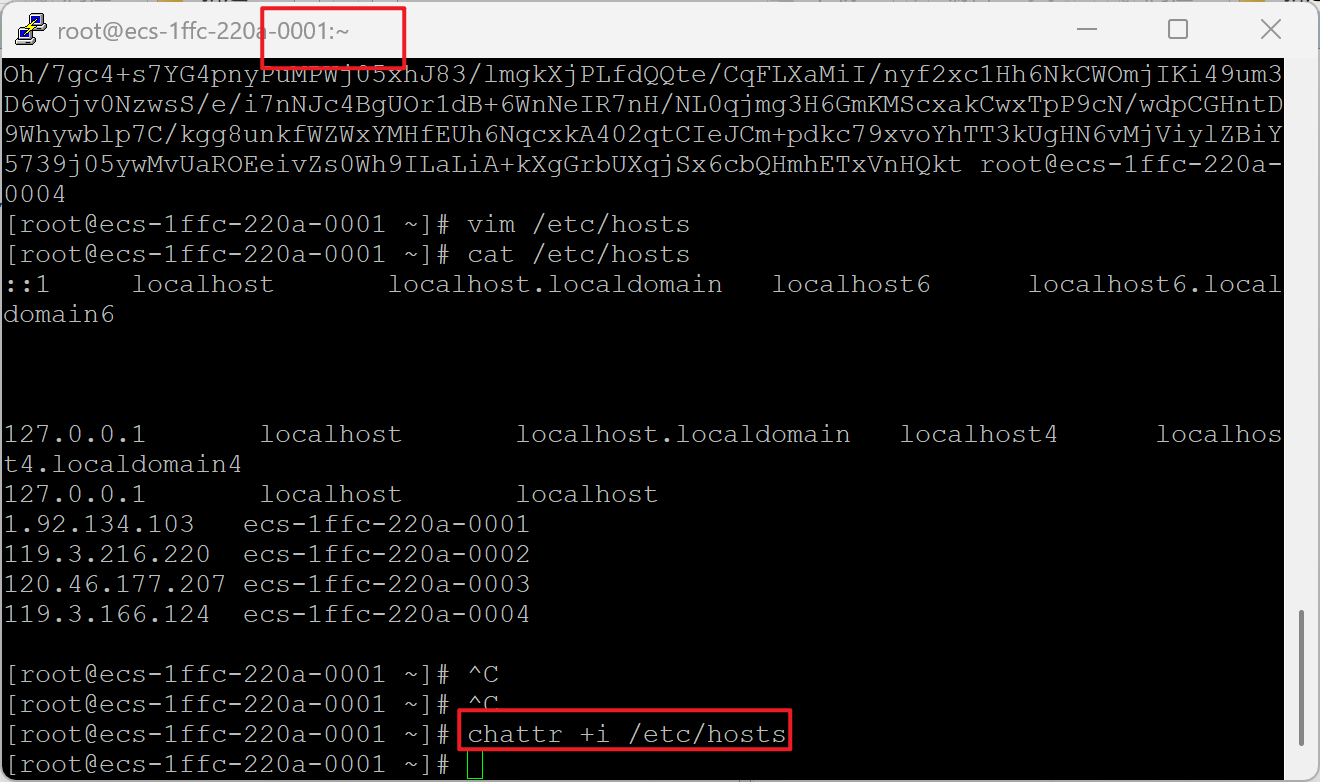
将4个节点的内容拷贝汇总到一个文本中，再将该文本内容拷贝到node1、node2、node3、node4的/root/.ssh/authorized\_keys中



删除4个节点中的127.0.0.1 ecs-xxx-0001 ecs-xxx-0001一行



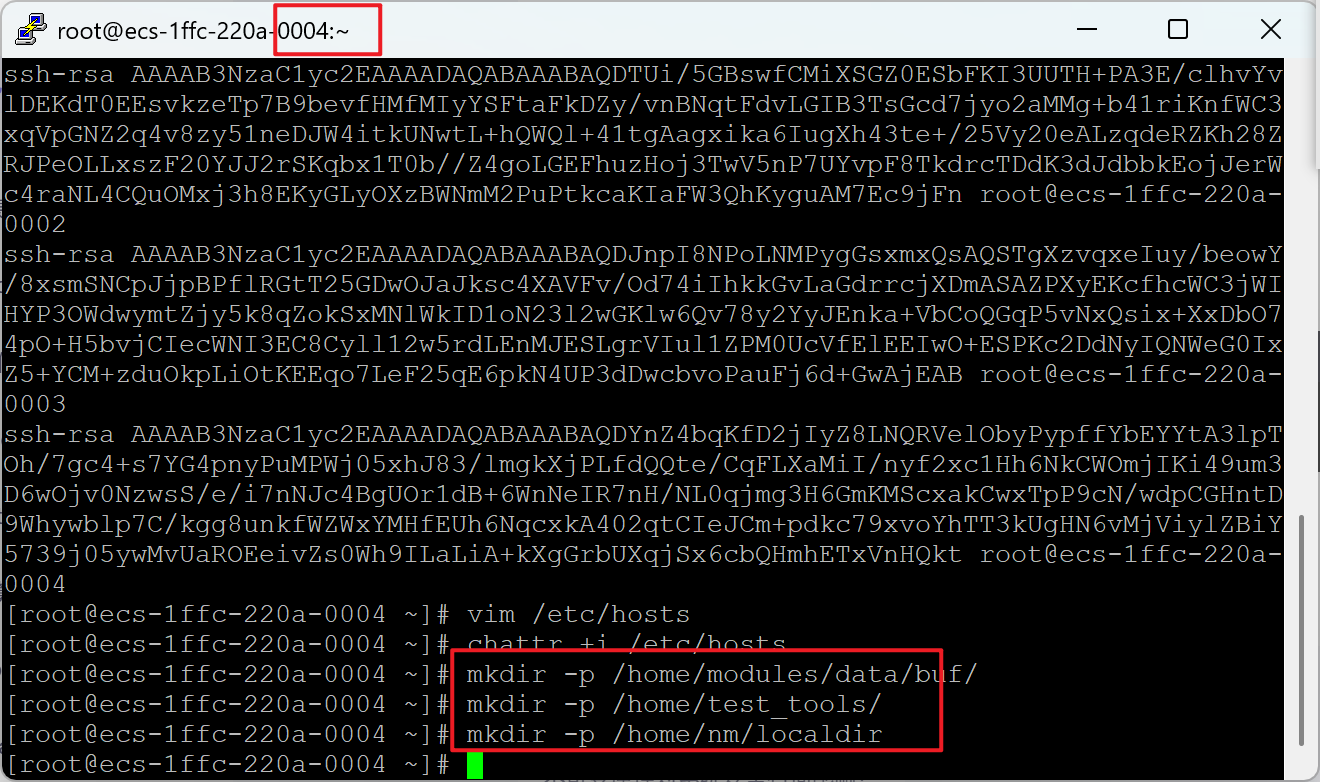
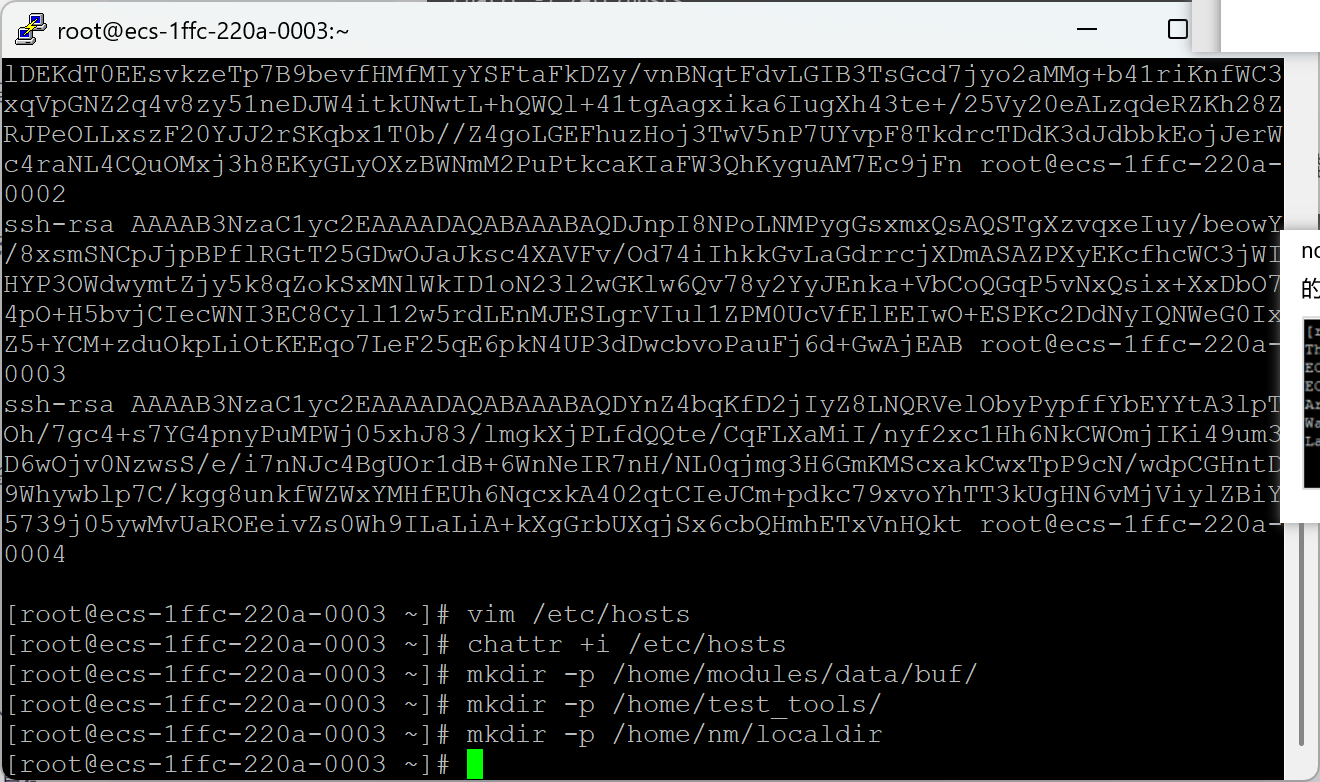
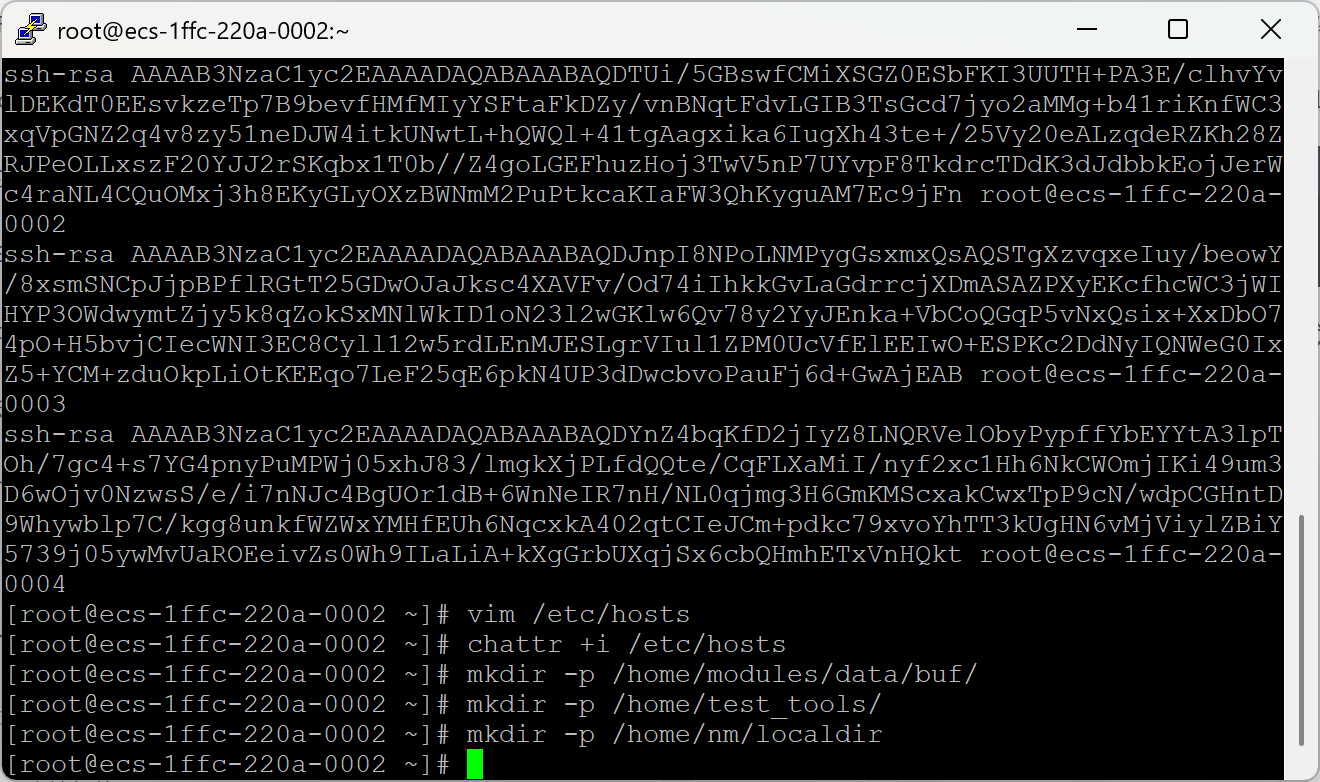
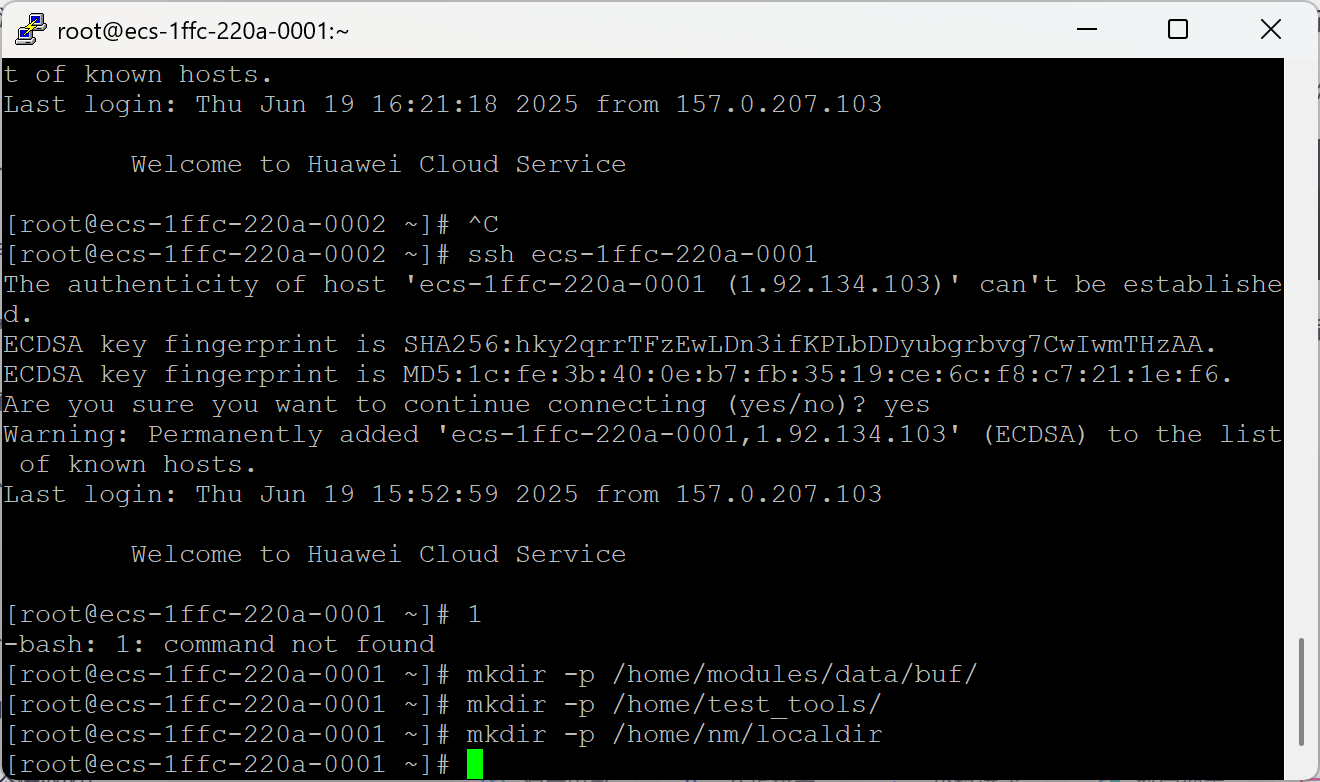
***执行***chattr +i /etc/hosts命令



**2.1.2安装OpenJDK**

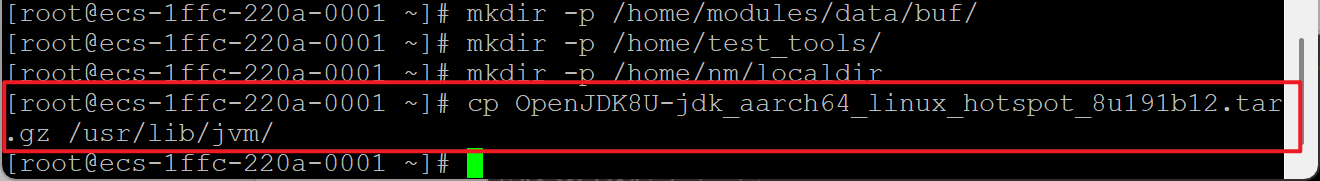
步骤1 创建目录

node1~node4四个节点分别执行下图命令

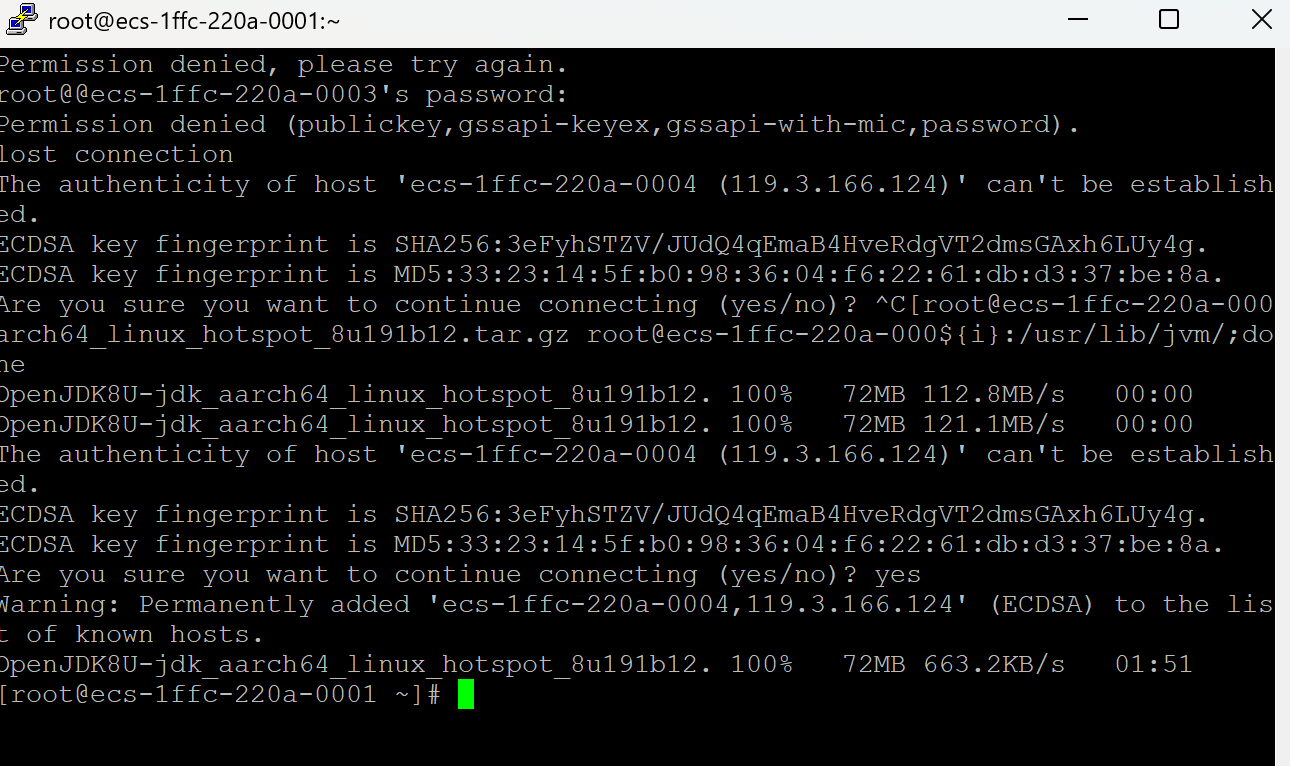


步骤2 安装OpenJDK

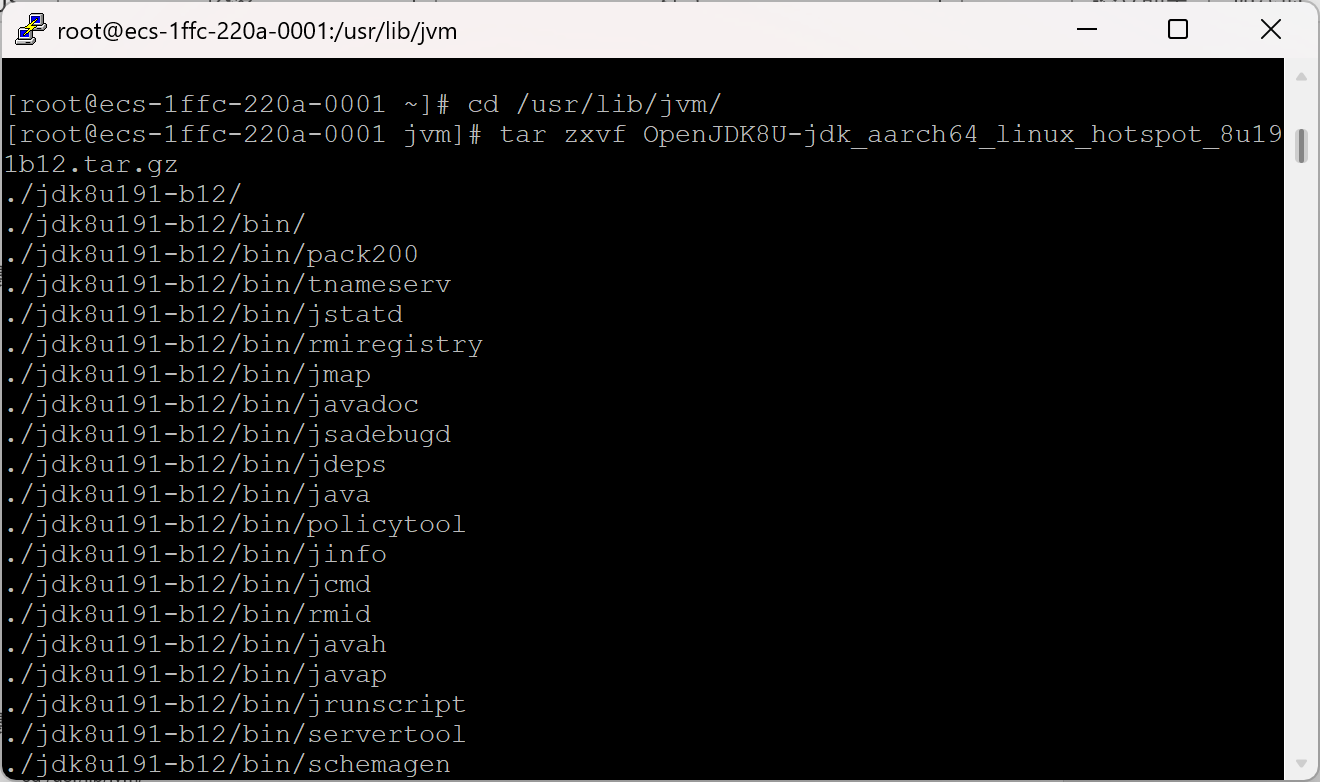
在node1节点，执行如下命令，将jdk安装包拷贝到/usr/lib/jvm目录下

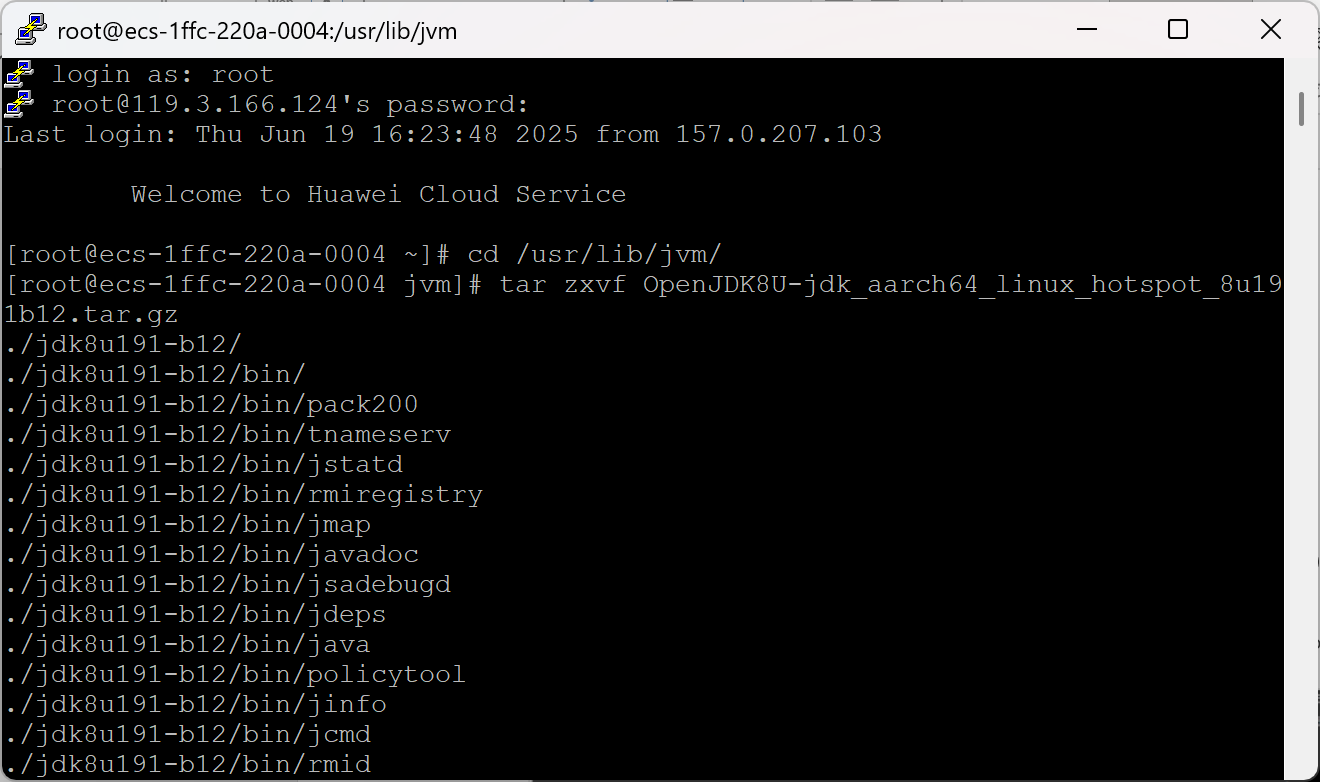
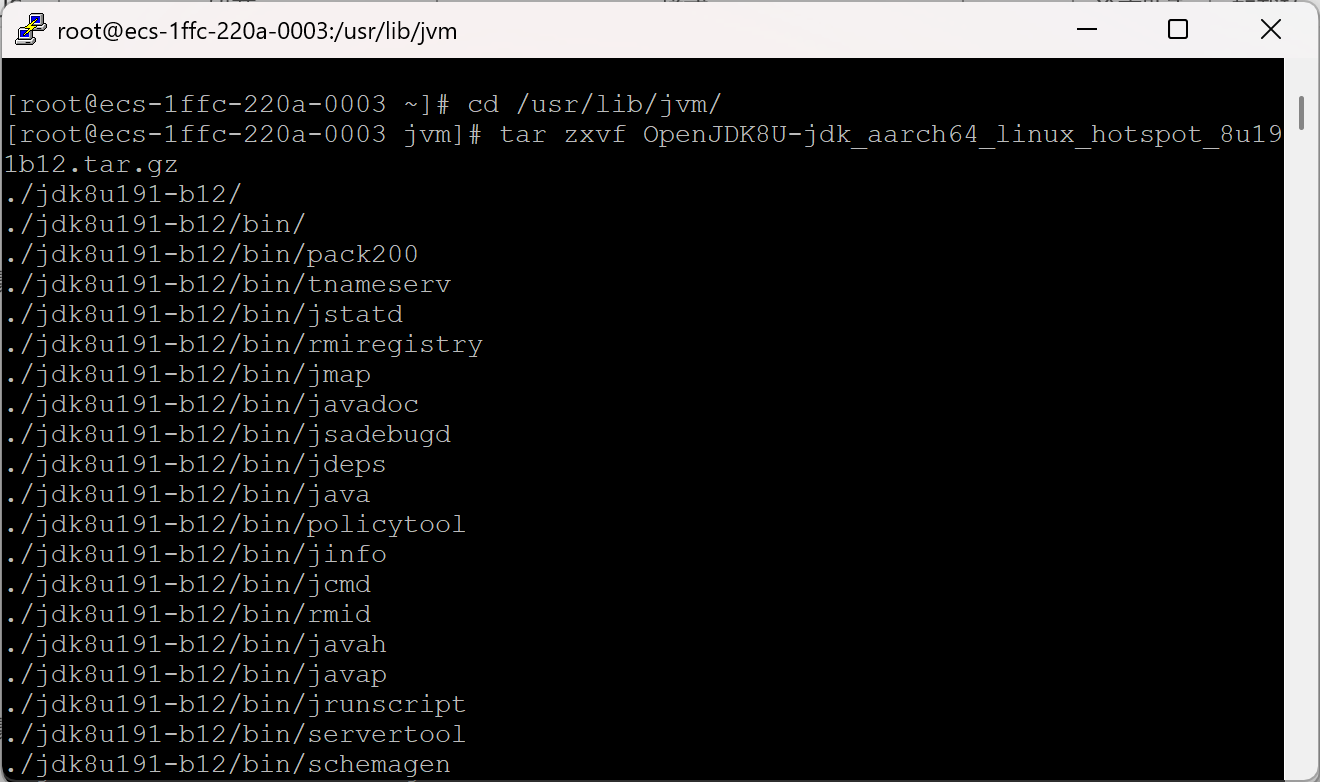
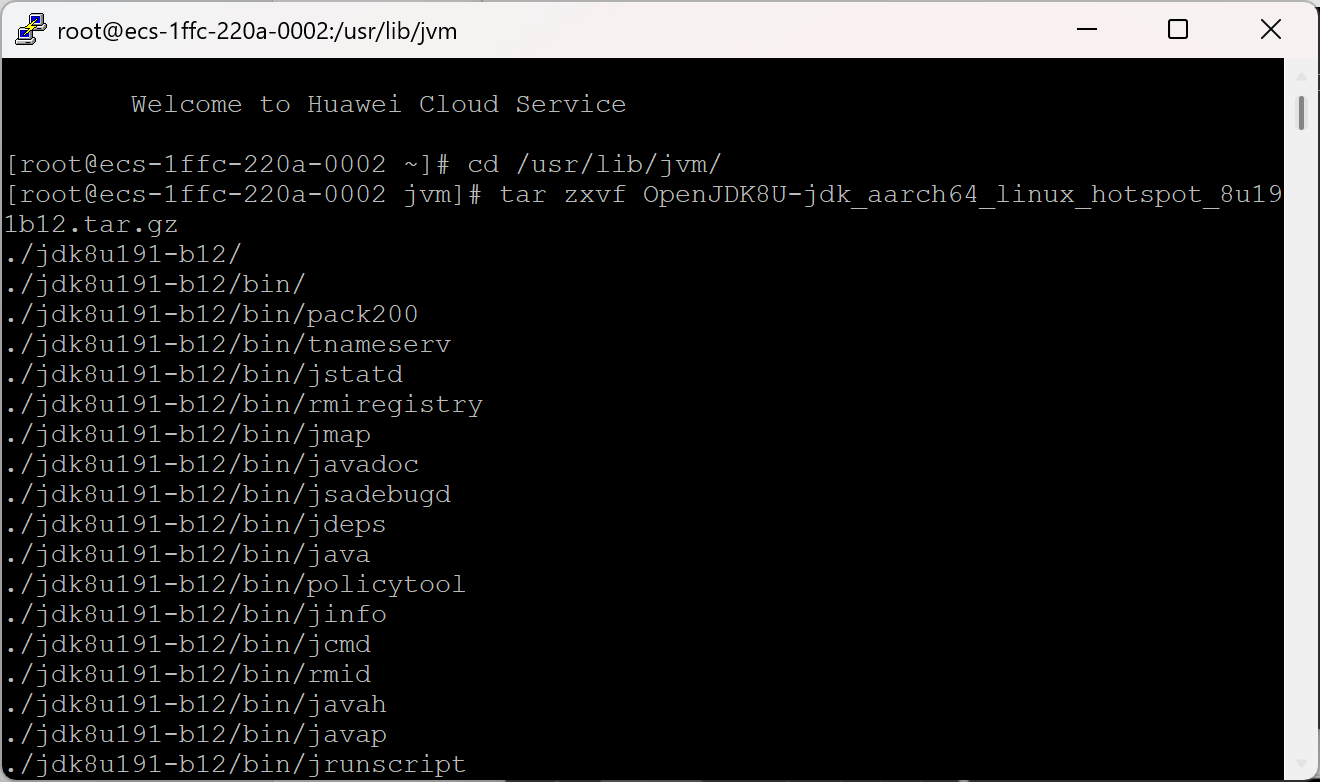


在node1节点，执行如下命令，将jdk安装包拷贝到其他几个节点，其中ecs-000为自己节点前面名称, i为节点数



步骤3 在node1~node4四个节点分别执行命令

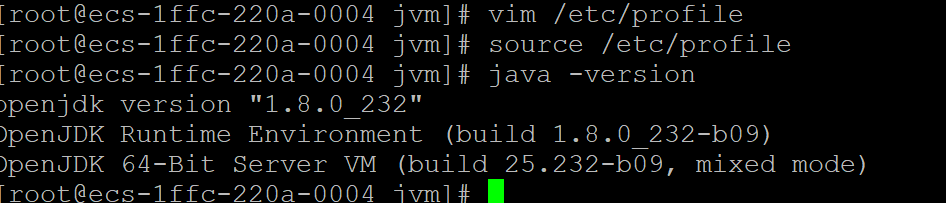




步骤4 在node1~node4四个节点上编辑/etc/profile增加如下的配置



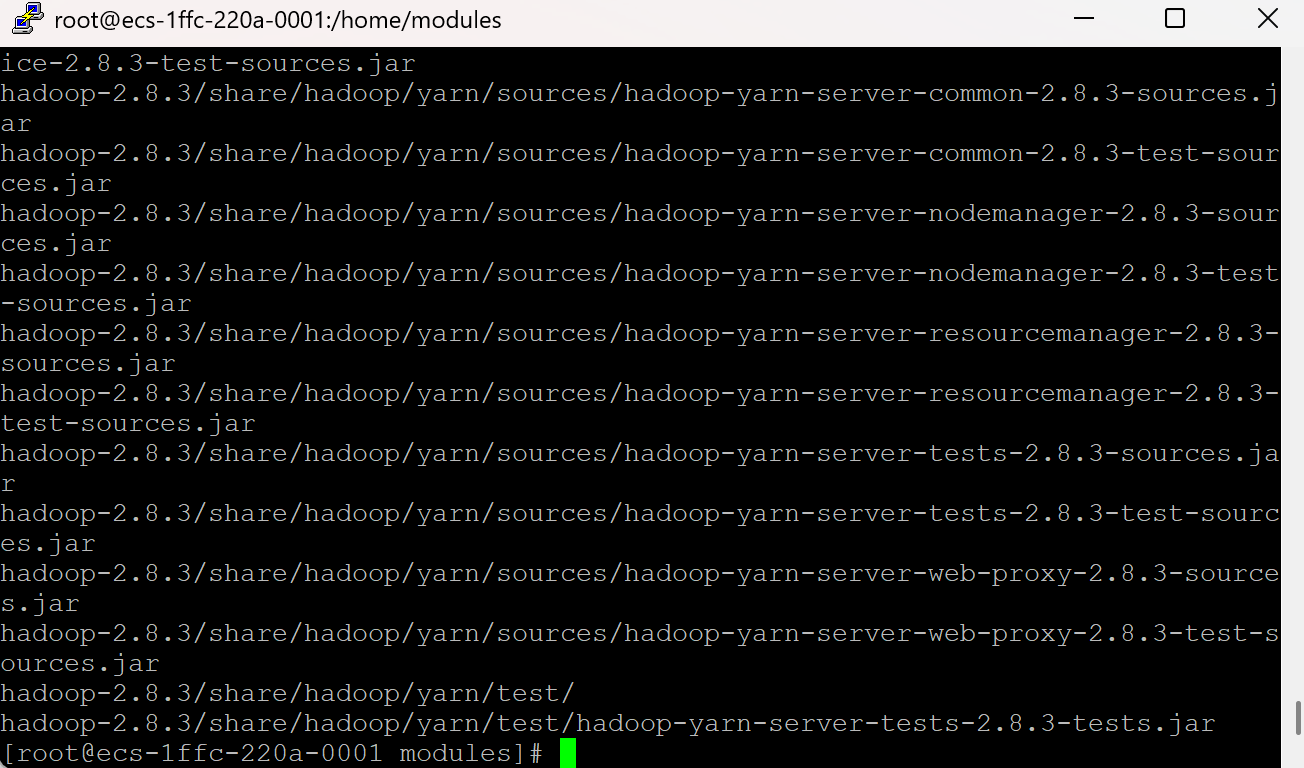
步骤5 确认java版本



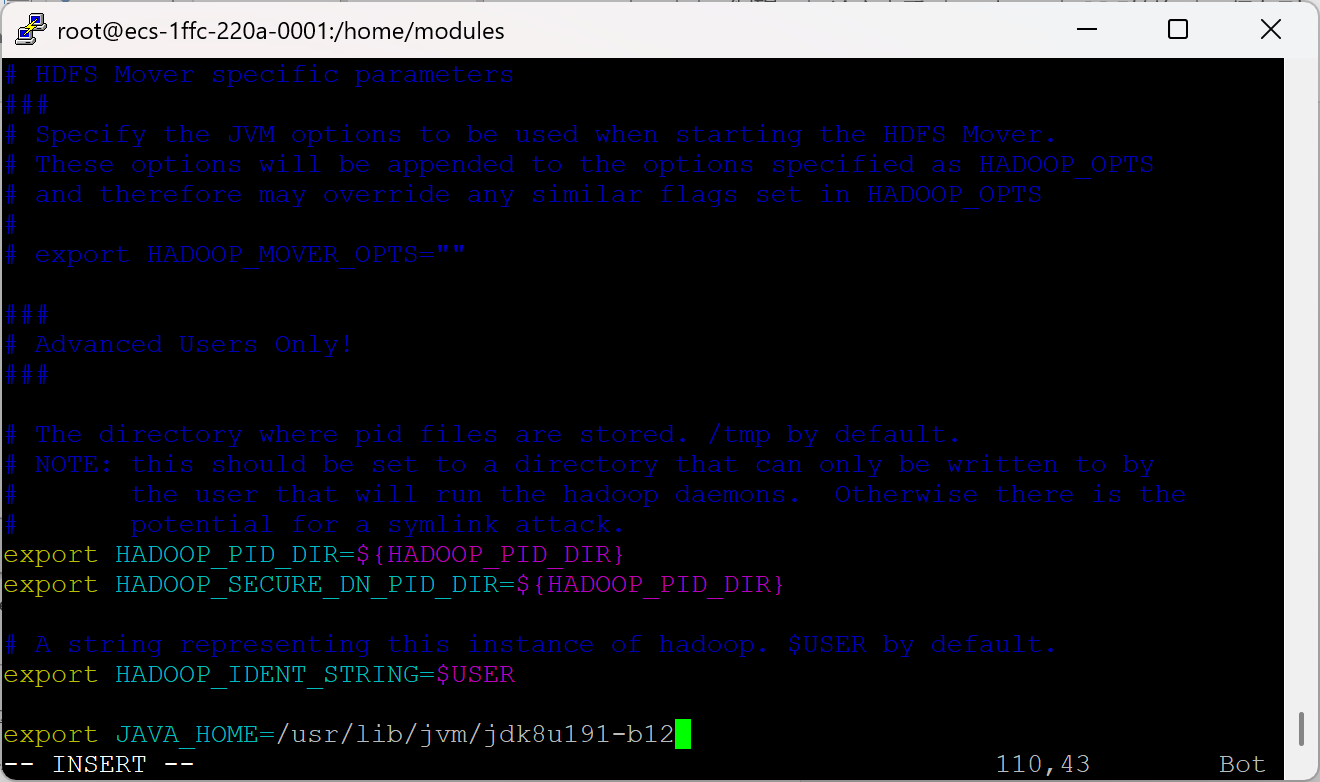
2.2搭建Hadoop集群

**2.2.1搭建Hadoop集群**

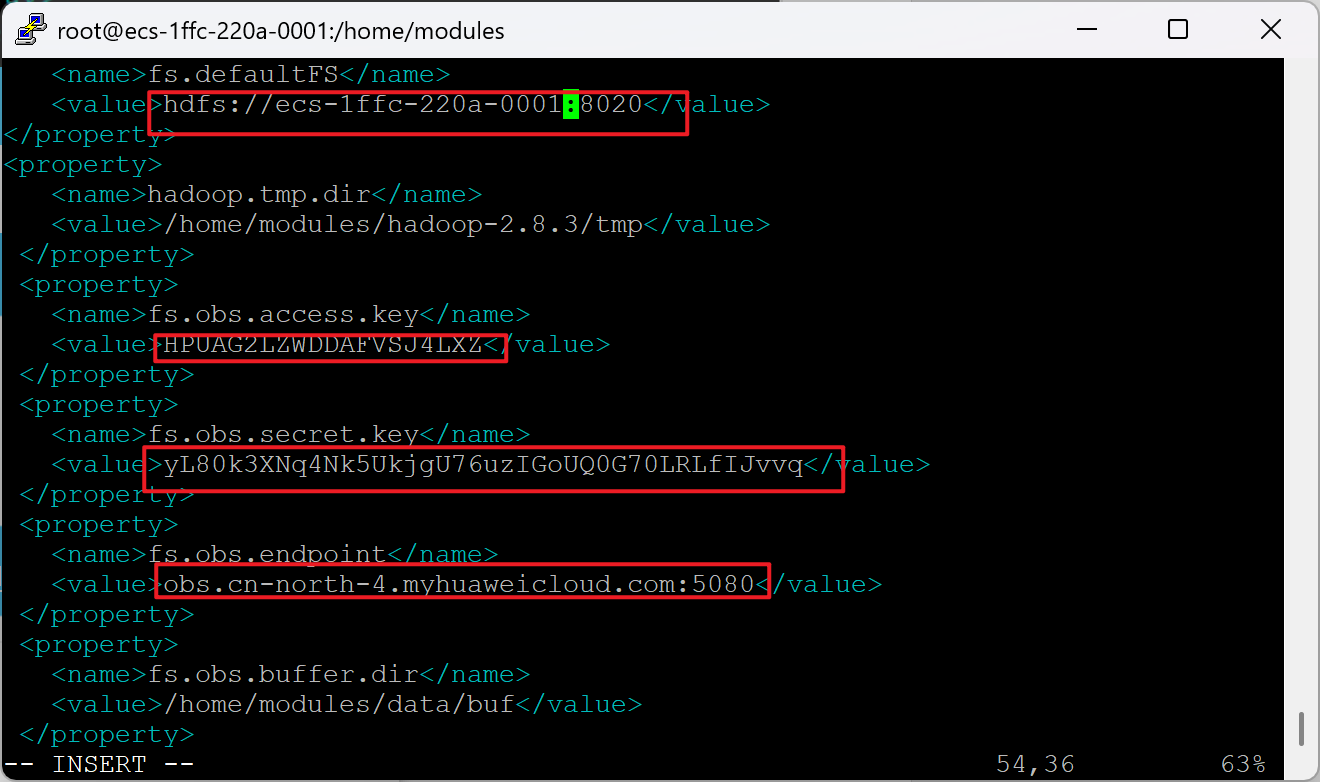
步骤1 登录node1节点，解压hadoop安装包



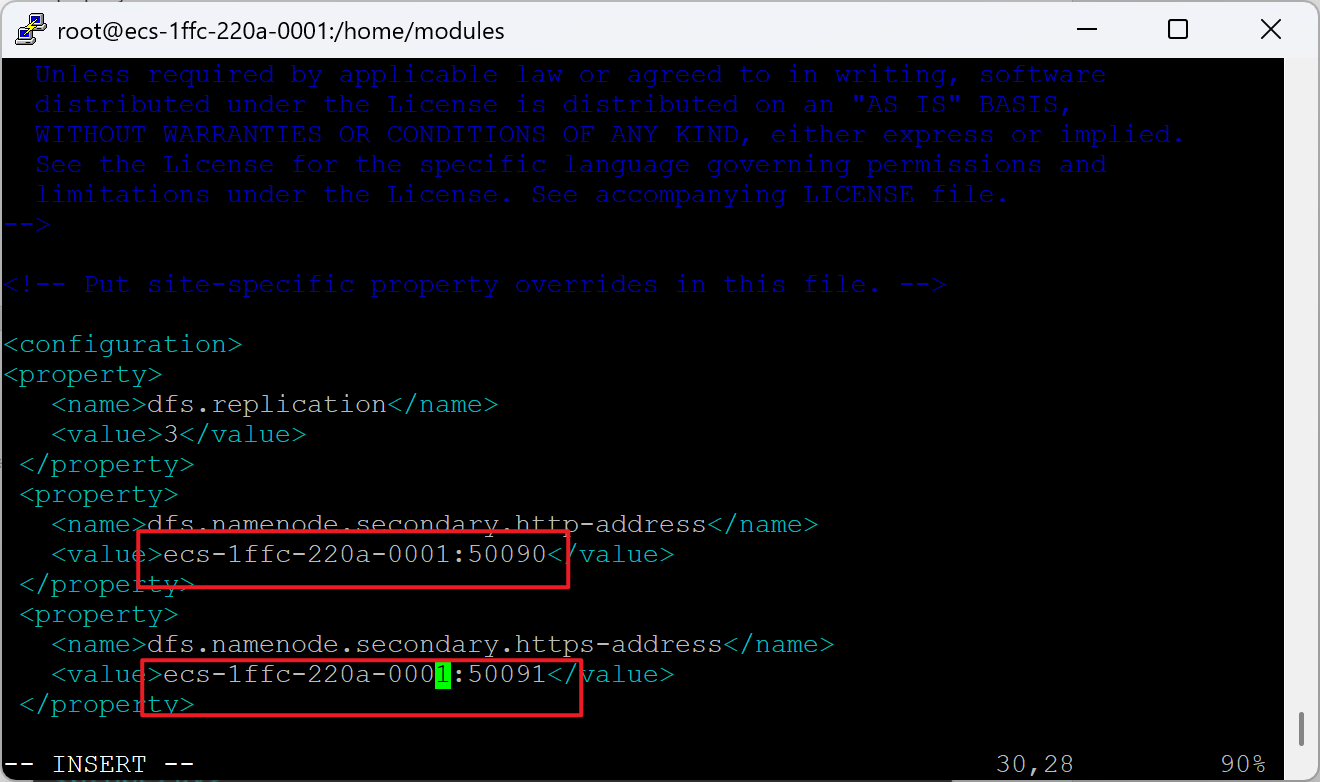
步骤2 配置Hadoop环境变量



步骤3 配置hadoop core-site.xml配置文件



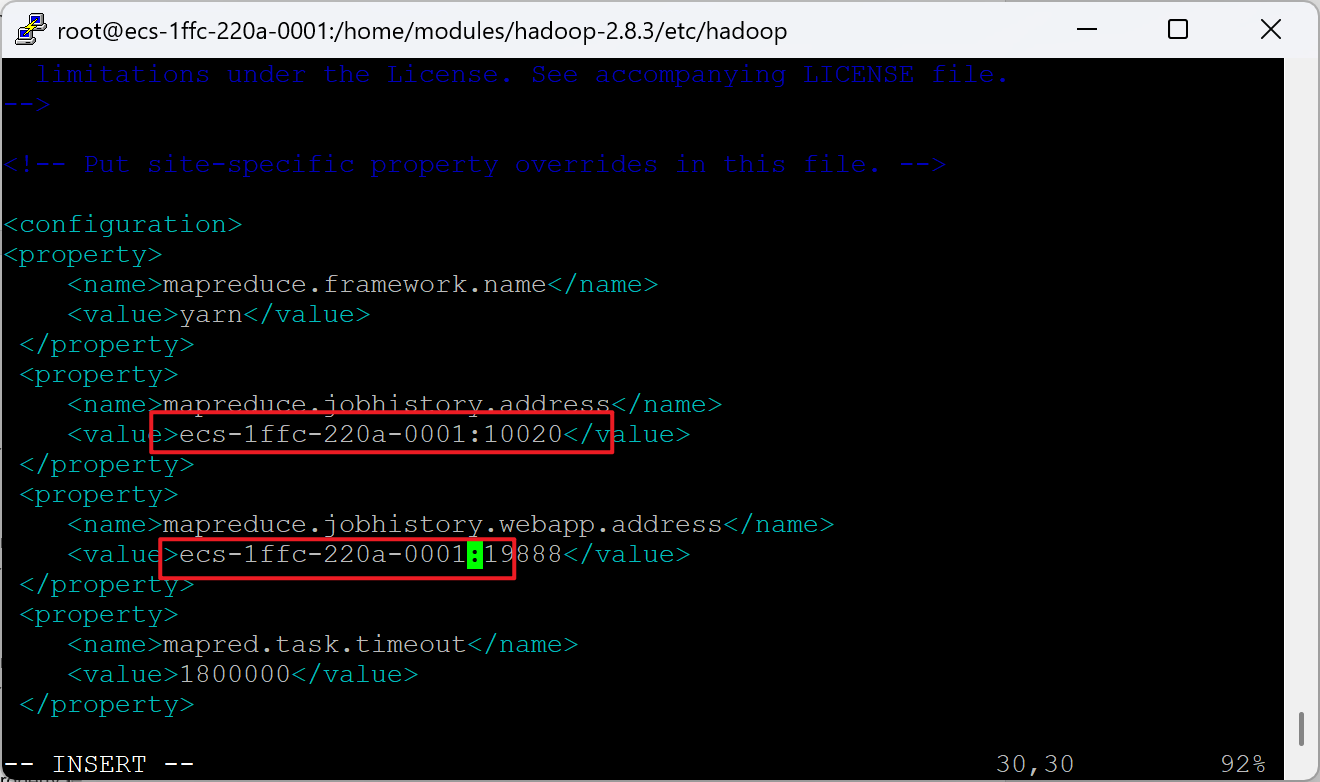
步骤4 配置hdfs-site.xml



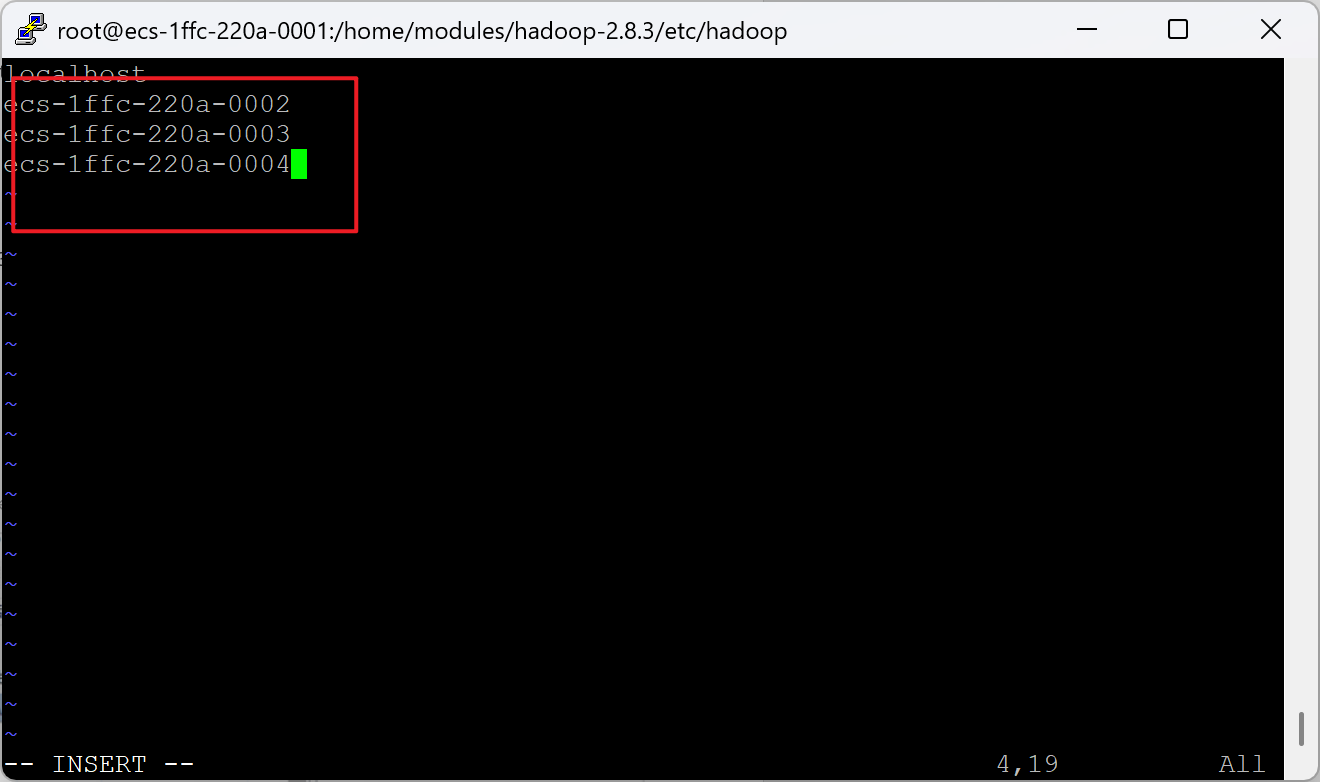
步骤5 配置yarn-site.xml



步骤6 配置mapred-sit.xml

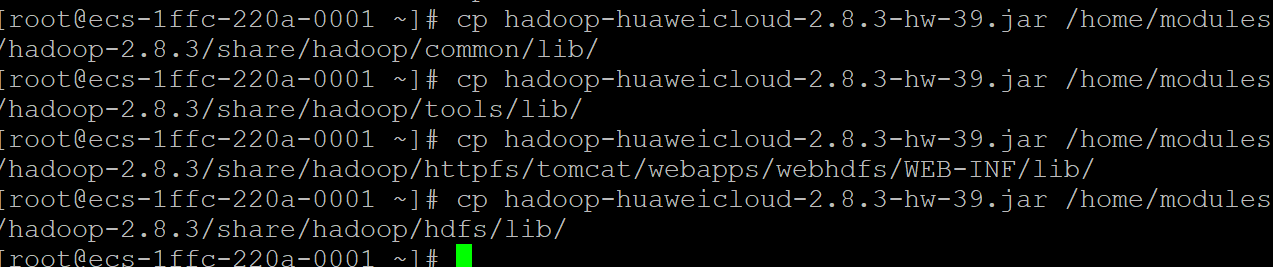
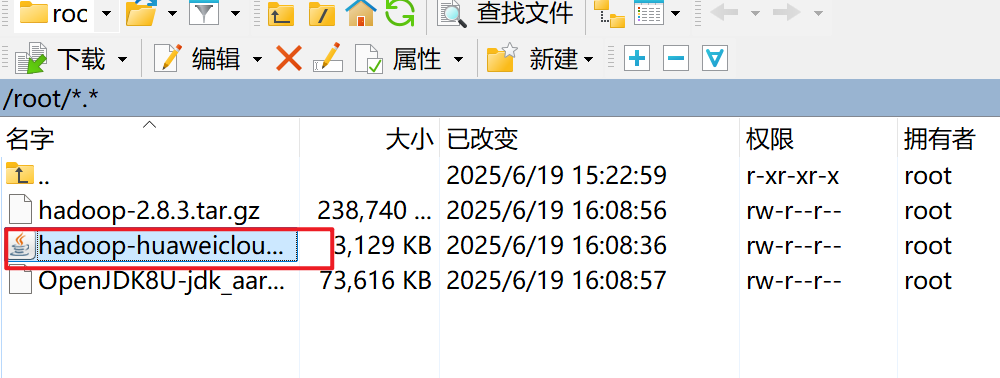


步骤7 配置slaves

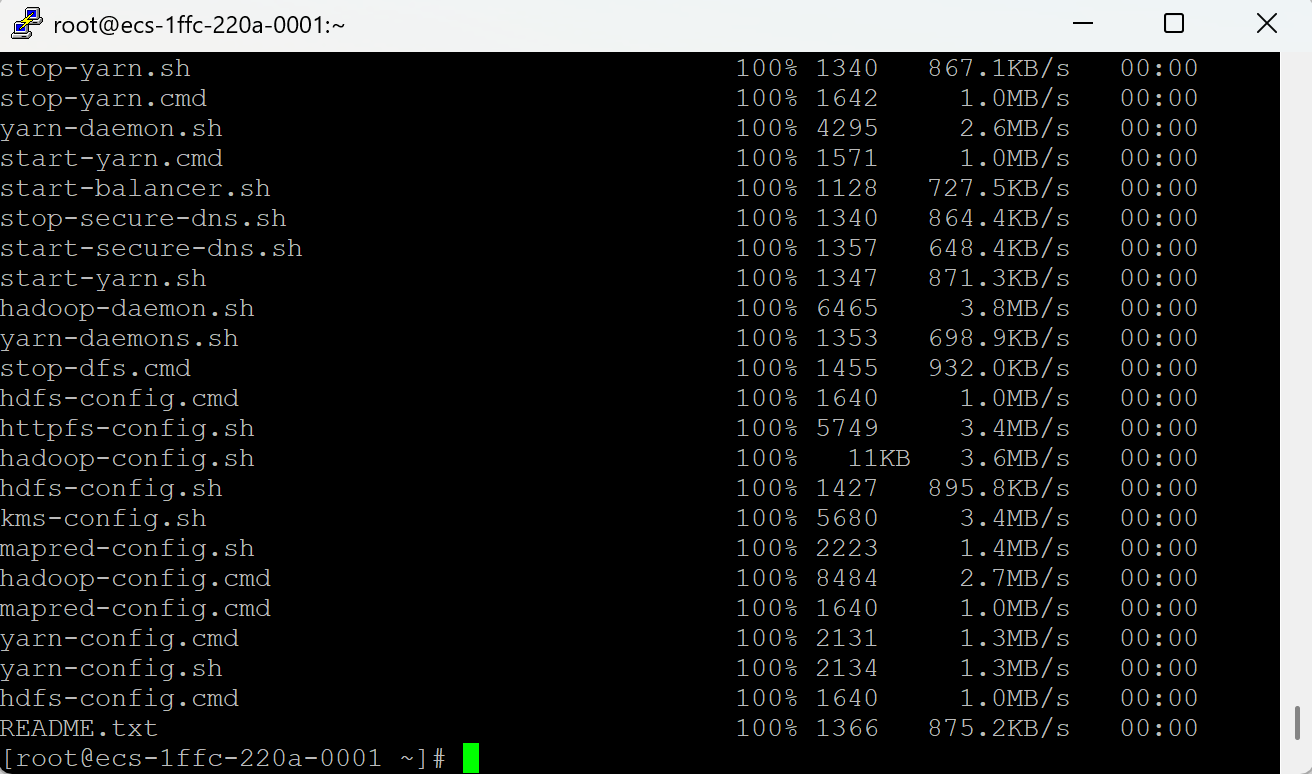


步骤8 配置jar包

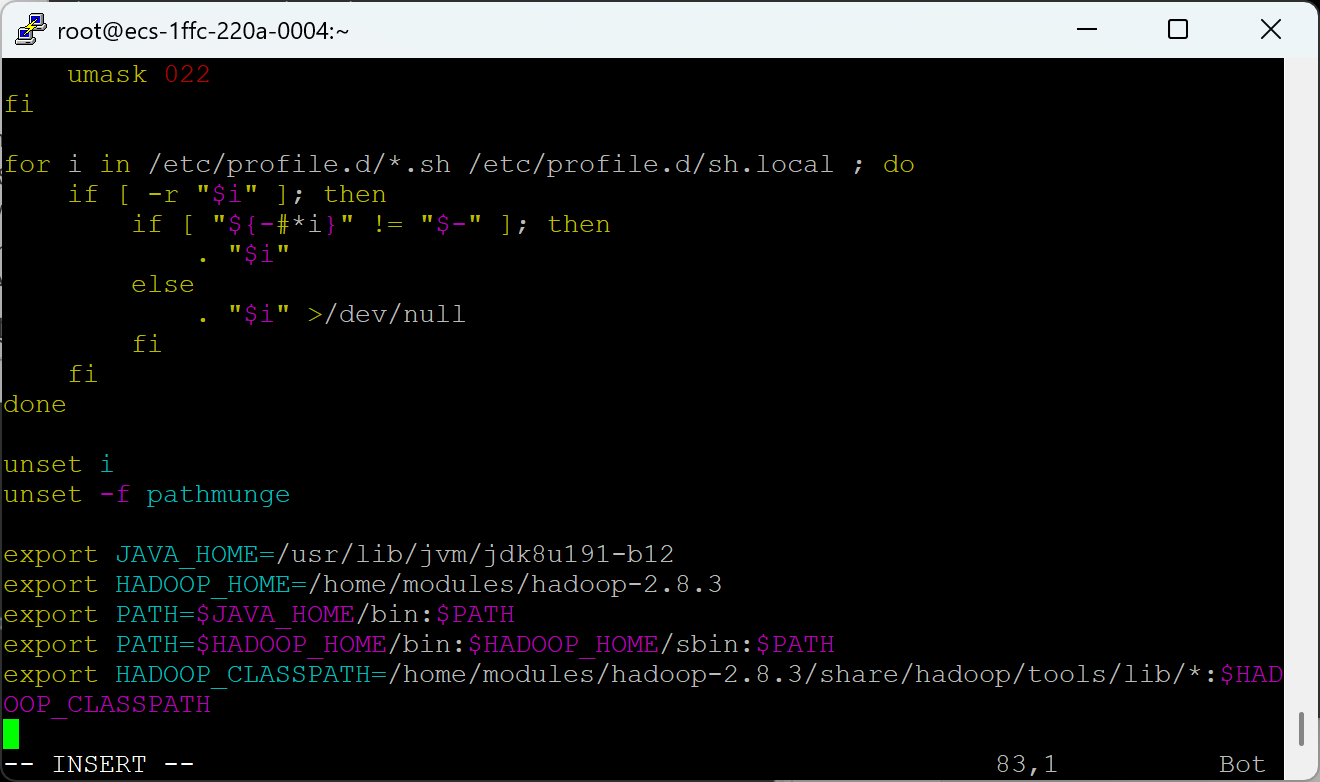
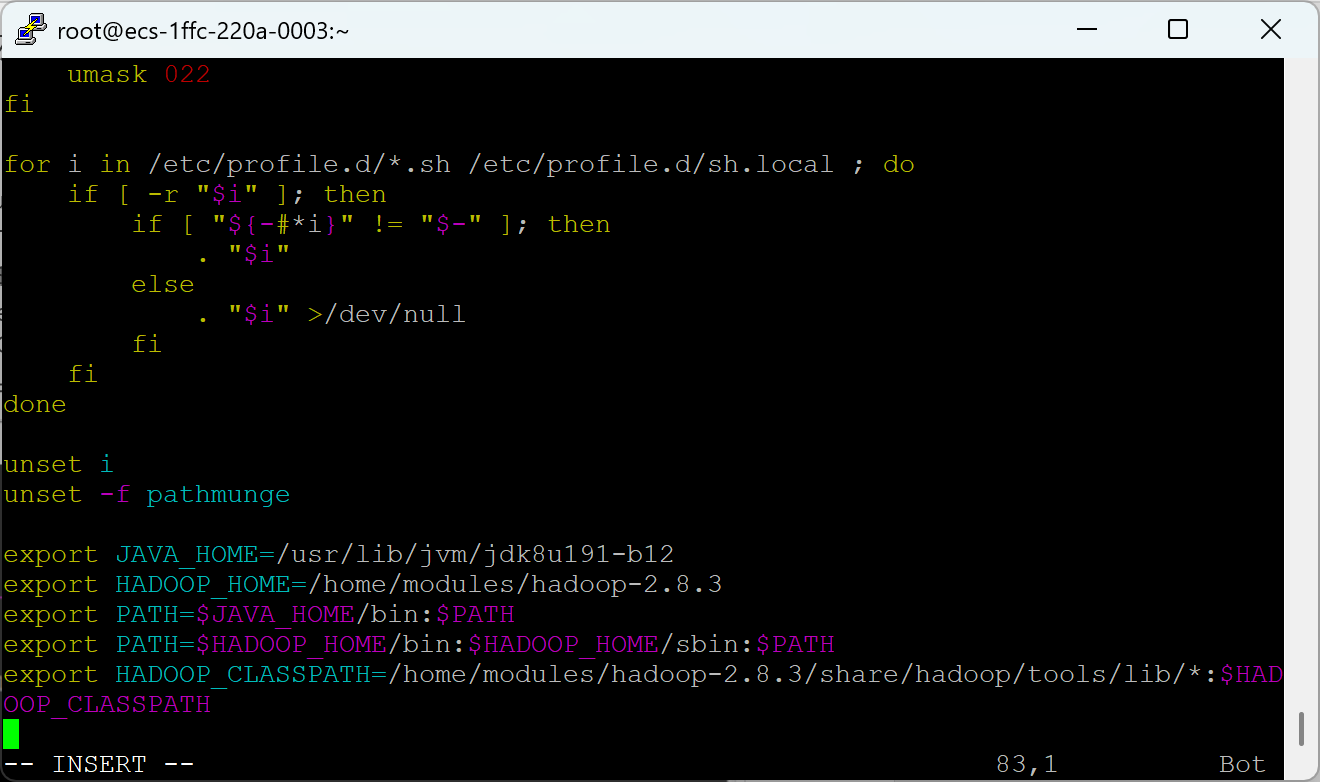
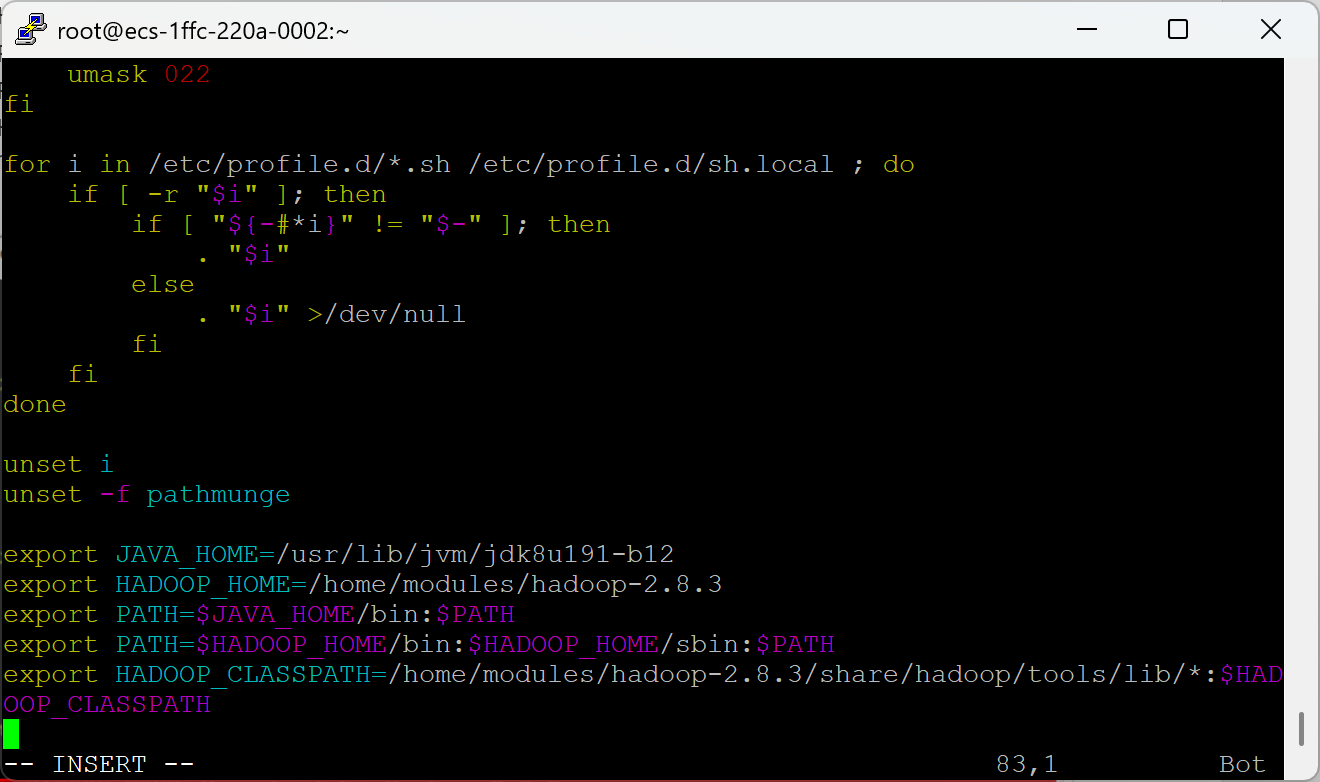
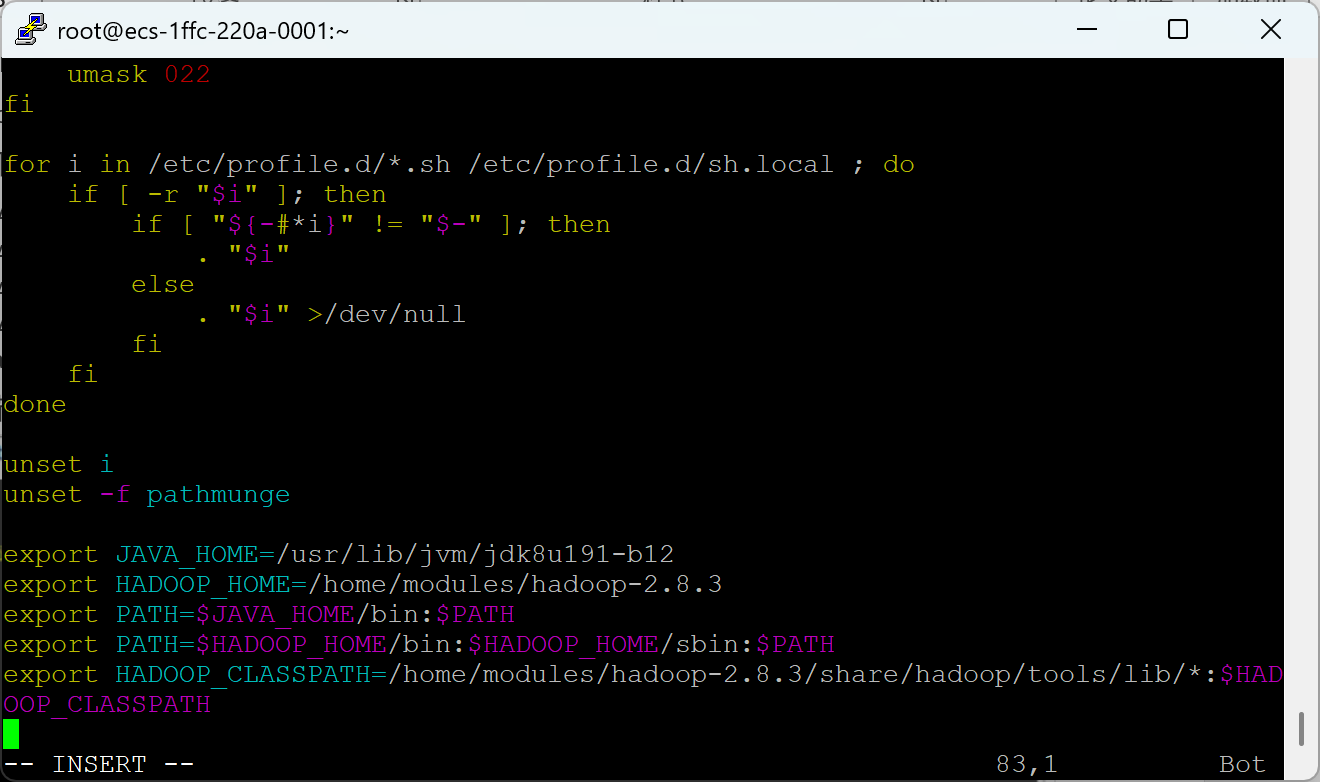
通过WinSCP上传至服务器

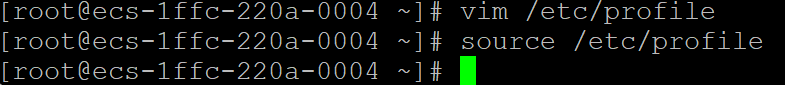
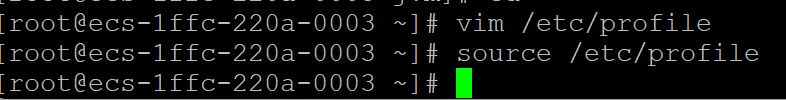
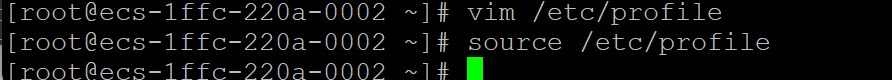
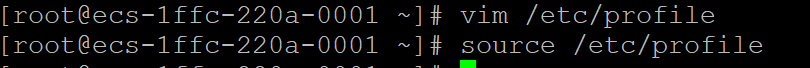


步骤9 分发Hadoop包到各节点

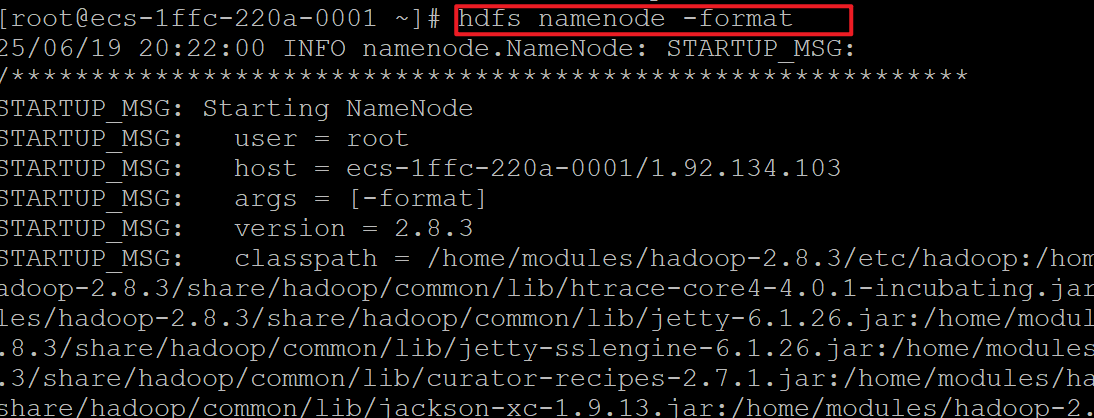


步骤10 配置1-4节点的环境变量

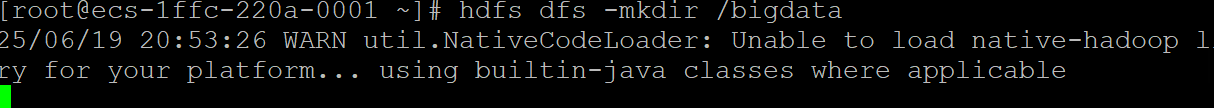




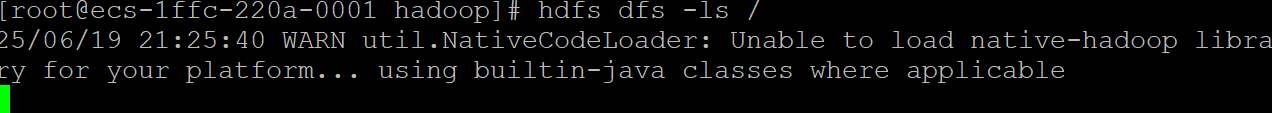
步骤11 namenode初始化



步骤12 启动HDFS

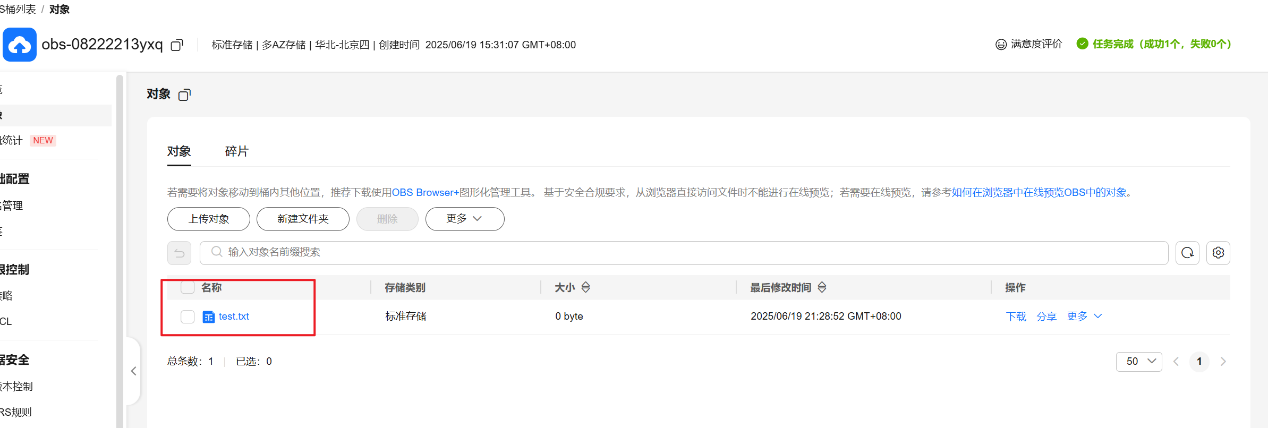


步骤13 执行hdfs命令

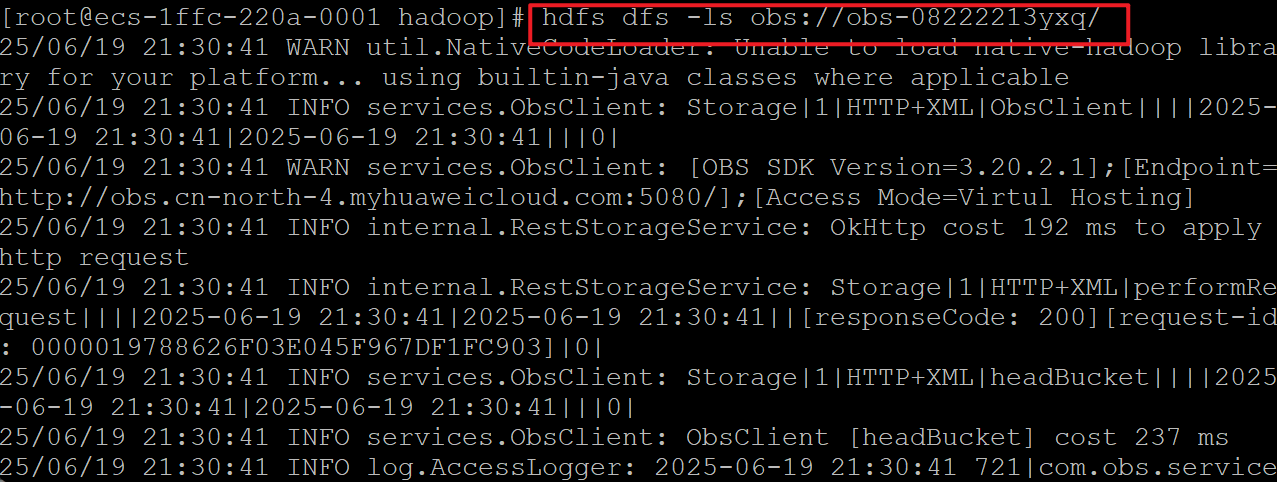


**2.2.2 测试与OBS互联**

步骤1 在OBS上传文件 选择“对象存储服务OBS”，选择“桶名称“



步骤2 Hadoop集群与OBS互联成功

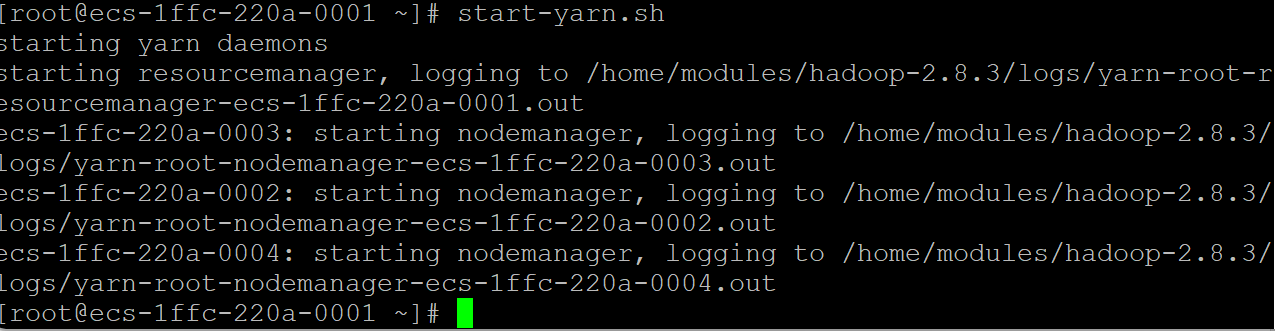


3 MapReduce

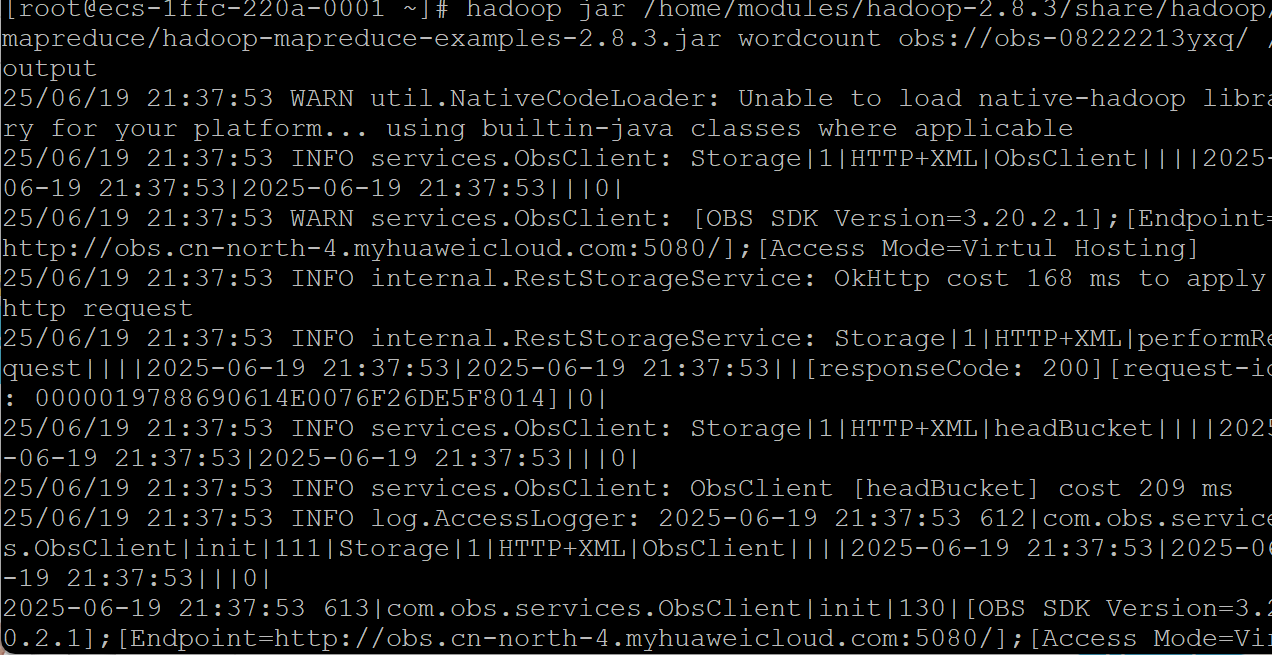
3.1 MapReduce存算分离

**3.1.1 测试Hadoop集群功能**

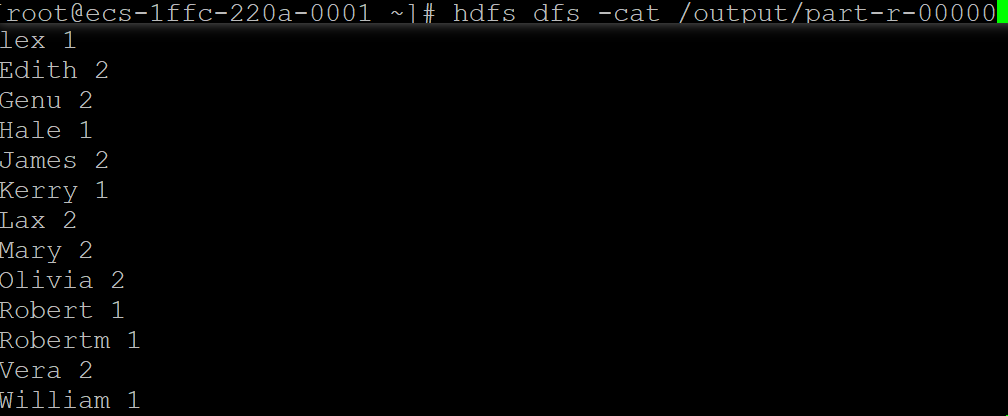
步骤1 node节点启动yarn执行命令



步骤2 执行Hadoop wordcount

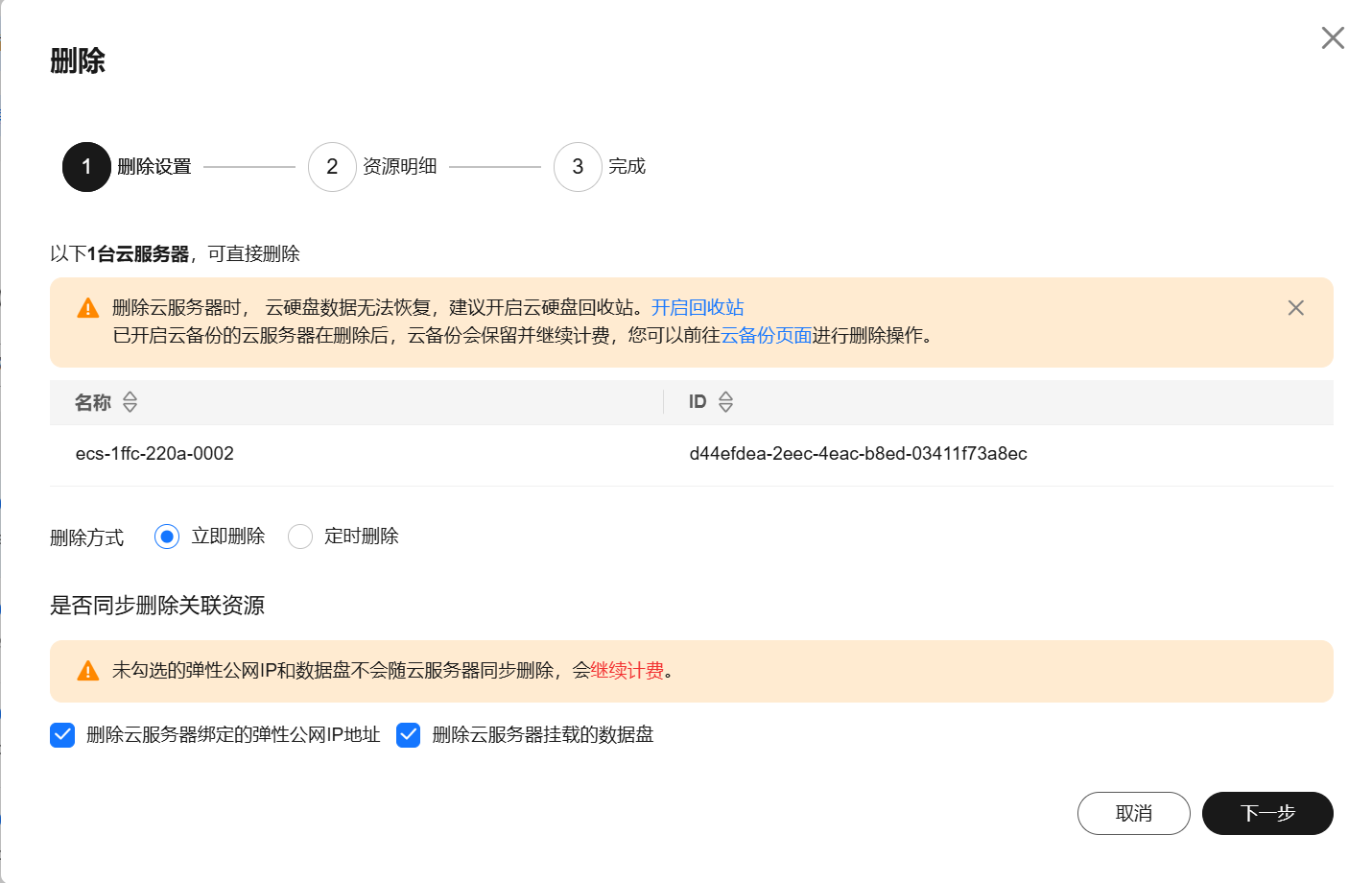
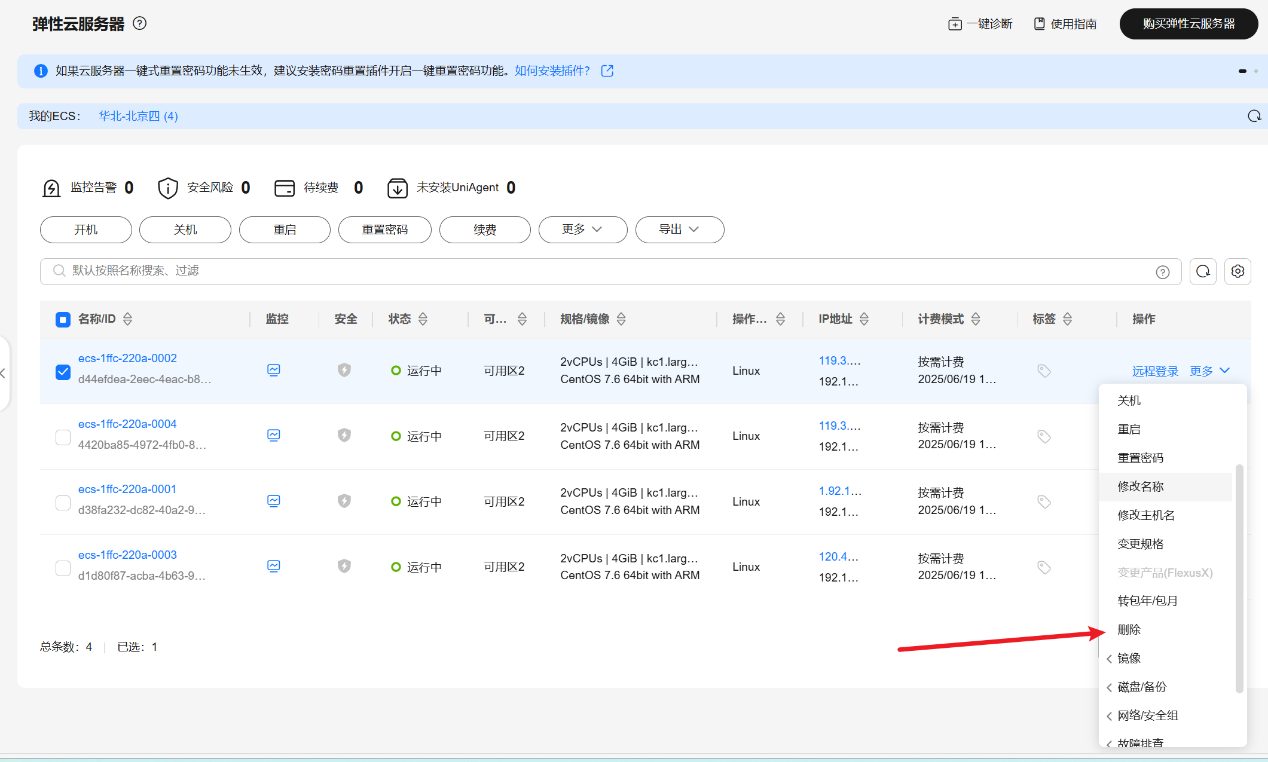


步骤3 查看结果，应用MapReduce实现存算分离计算成功!



4 释放云服务资源

4.1 释放弹性云服务器ECS



4.2删除桶OBS

进入OBS服务，点击“对象存储”，在右下方点击“删除”，删除桶

