实践报告：Function Gragh

学号：2212534

姓名：魏思诚

日期：2024/12/7

南开大学 软件学院

目 录

[1 实验介绍 2](#_Toc7jmh2f)

[1)实验环境1](#_Toclypsj7)

[2)实验流程 n](#_Toczzwyqs)

[2 实验目的 5](#_Tocts8p9h)

[1)学习了解华为云主机的使用方法](#_Tock523rk)

[2)学习了解MapReduce实现存算分离](#_Tocpmdw9e)

[3 实验流程 7](#_Toc5ihwrt)

[1)配置实验环境](#_Toc8tbxzn)

[2)编写代码](#_Toc5489qp)

[4 实验结果或结论 33](#_Toc2fe5jy)

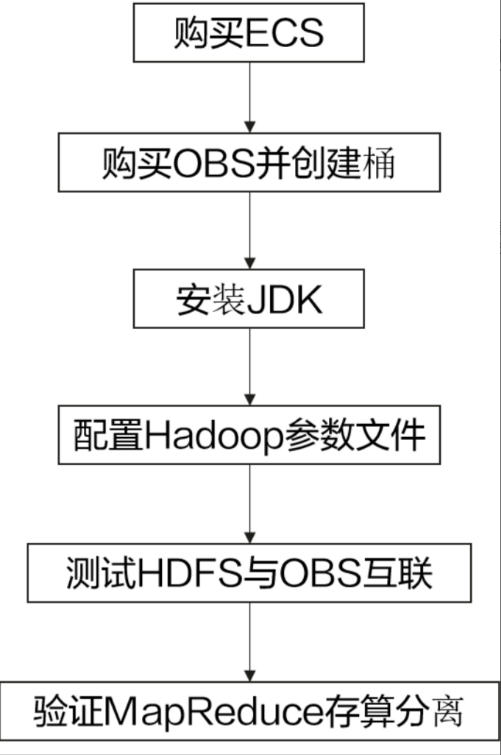
[1)实验结果1](#_Tocmqn8po)

[2)实验结论n](#_Tocaflm23)

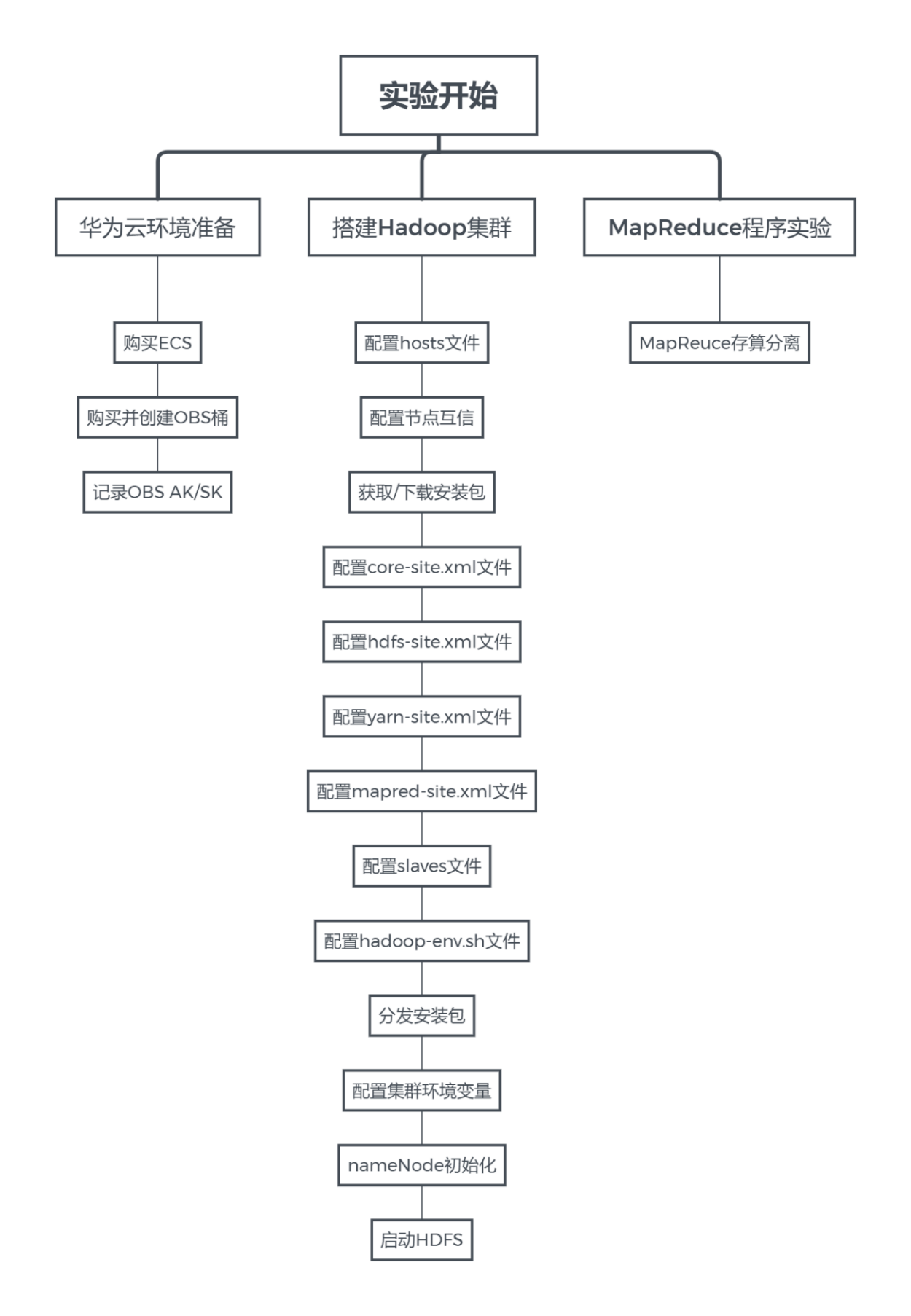
# 实验介绍

## 实验环境

* 本实验通过购买华为云ECS和在华为云上选取对象存储服务OBS服务，为后续实验提供搭建环境基础。本实验的基本步骤包含：购买并配置ECS；购买OBS并获取AK、SK信息；搭建Hadoop集群；搭建MapReduce集群并验证存算分离。



## 实验流程

* 

为了满足基于ECS实验需要，建议每位学员提前注册华为云账号并充值，确保通过PUTTY访问外网畅通。

检查设备

* 实验开始之前请每位学员检查自己的实验设备是否齐全，实验清单如下。
* 

# 实验目的

## 学习了解华为云主机的基本使用方法

* 了解华为云主机的部署以及使用,华为CodeArts等编译器以及API调用的方法
* 熟悉华为CodeArts工具链的功能，包括代码开发、编译、构建、测试和部署的全生命周期管理。
* 学习如何配置和使用CodeArts中集成的编译器支持各种语言（如C/C++、Java、Python等），并在云端或本地环境中运行代码。
* 掌握如何使用CodeArts提供的CI/CD流水线，实现代码的持续集成与部署，确保软件开发流程的高效性和规范性。
* 探索CodeArts的协作功能，如代码托管、代码审查、任务分配和项目进度管理，提高团队协作效率。
* 学习华为云API的基础知识，包括RESTful API的使用规范、请求方法（GET、POST、PUT、DELETE）以及如何通过HTTP协议与云服务进行交互。
* 掌握API调用的认证方式，特别是如何生成和使用访问密钥（AK/SK）进行安全认证，以及通过IAM进行权限管理。
* 使用工具（如Postman）进行API接口调试，并通过脚本语言（如Python或Java）实现对云服务的程序化控制。
* 熟悉常用的API调用场景，例如资源管理（启动、停止、删除云主机）、数据交互（上传、下载对象存储中的文件）以及日志分析（调用日志、服务监控）。

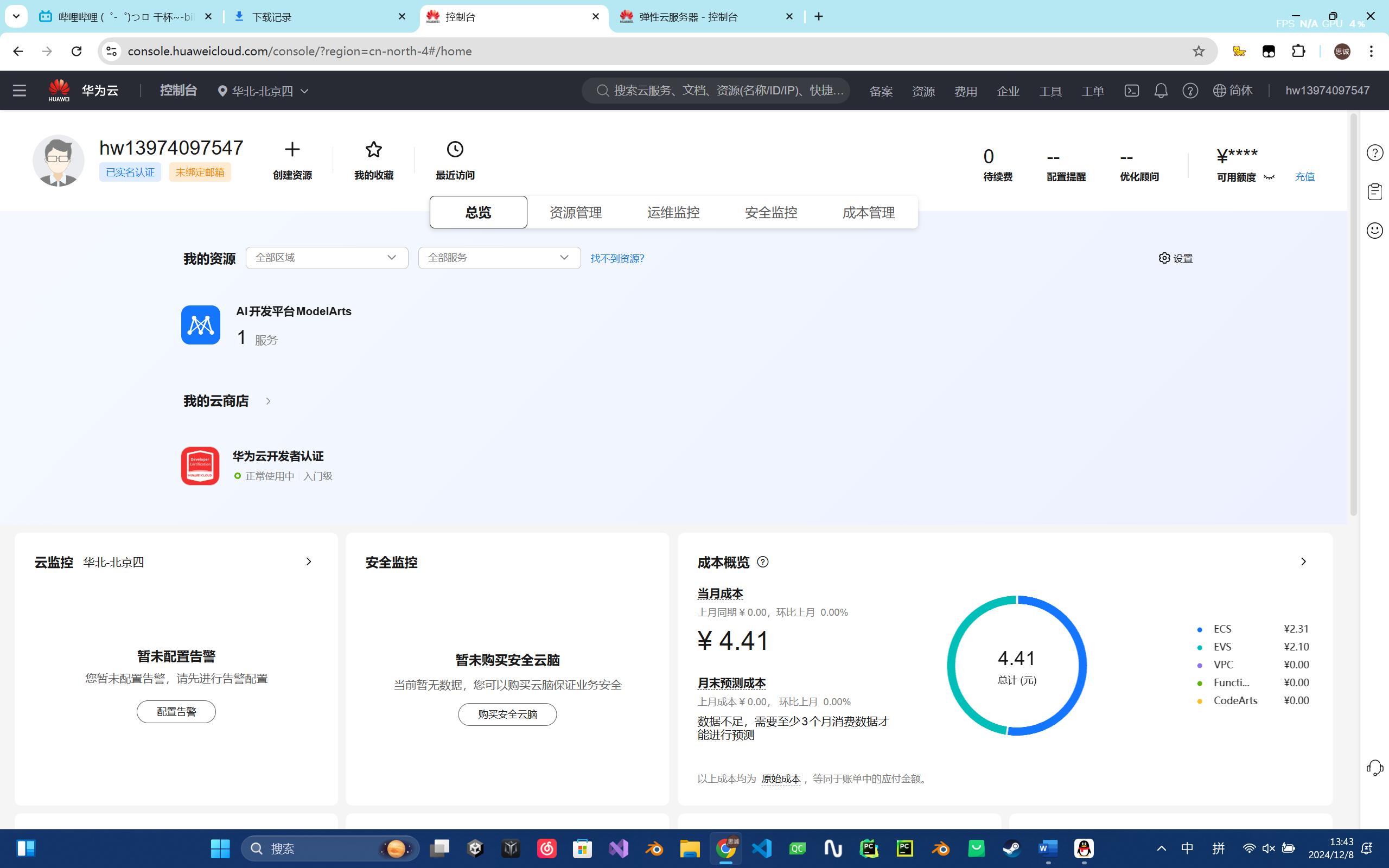
## 掌握MapReduce实现存算分离

* **理解MapReduce编程模型**： 学生将深入学习MapReduce的基本原理，掌握其如何通过Map和Reduce两个阶段完成数据处理任务。通过Map阶段将数据分割并映射成键值对，Reduce阶段对映射的键值对进行合并与聚合。
* **实现存算分离的架构**： 学生将学习如何通过MapReduce实现存算分离，即将数据存储和计算任务分别交由不同的系统和节点来处理。数据存储通常由分布式存储系统（如HDFS）负责，而计算任务则由MapReduce框架执行。这种架构能够提升系统的灵活性和扩展性，使得数据存储和计算可以独立扩展。
* **提升大规模数据处理效率**： 通过实现存算分离，学生将了解如何利用MapReduce框架高效处理大规模数据。数据存储和计算的分离使得可以独立优化存储系统和计算资源，进而提升系统的整体性能。
* **掌握分布式计算的实现**： 学生将学习如何将复杂的计算任务分解为多个子任务，并在分布式系统中并行处理。这不仅包括MapReduce的实现，还涉及到如何在实际应用中协调存储和计算过程，确保计算的正确性和高效性。
* **应用案例分析**： 通过具体的实验案例，学生将学会如何使用MapReduce框架处理大数据任务，如大规模日志分析、数据挖掘等，掌握存算分离的实际应用。

# 实验流程

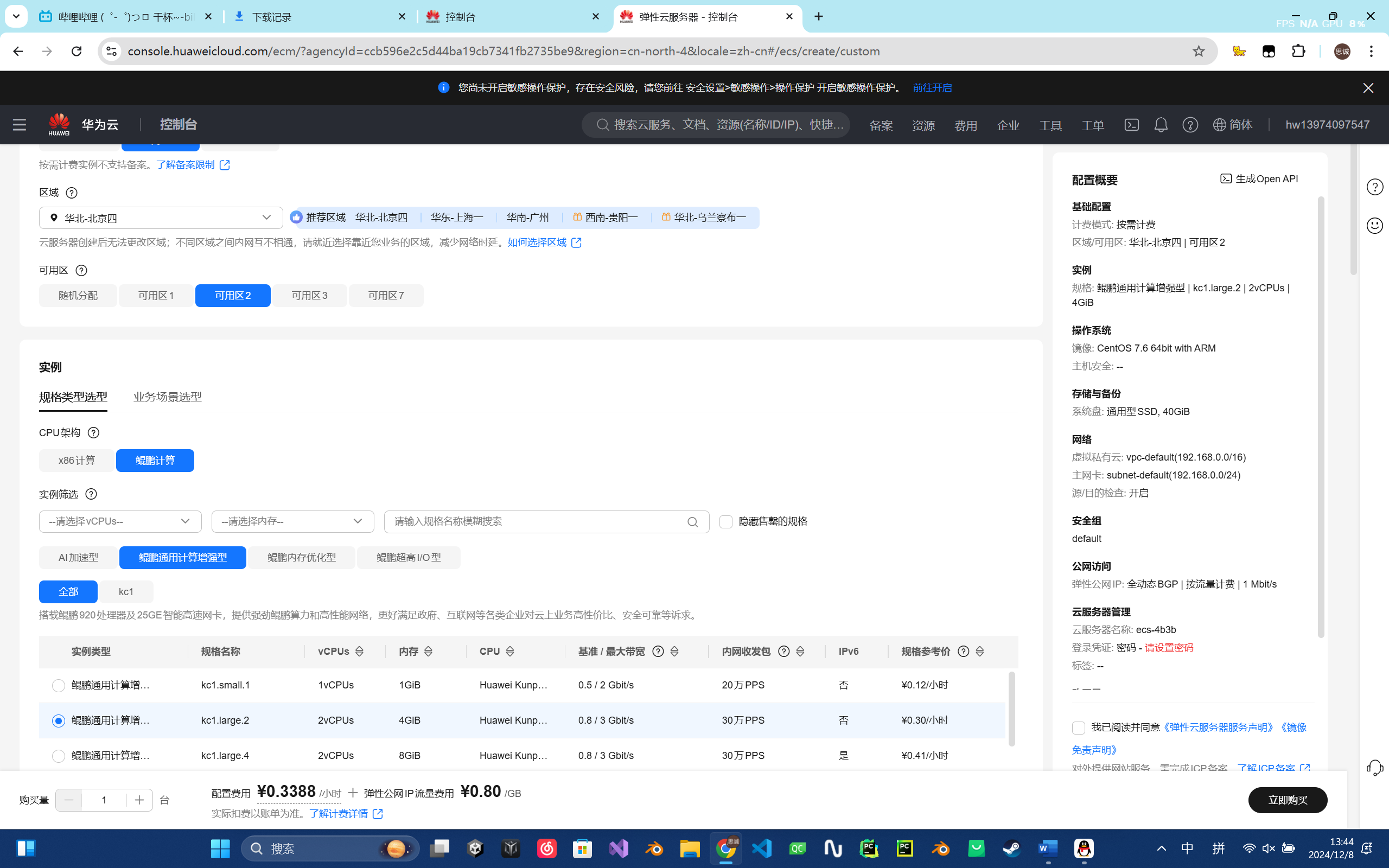
## 配置实验环境

浏览登录华为云

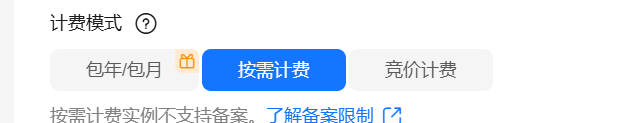


购买服务器

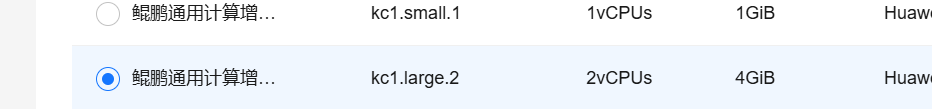
* 选择“购买弹性云服务器ECS”，如下图。



* 选择“按需计费”，“可用区2”，CPU架构 “鲲鹏计算”，如下图



* 选择“鲲鹏通用计算增强型”，2vCPUs|4GB, 配置操作系统和磁盘，如下图



* 选择“公共镜像”，CentOS7.6，系统盘建议配置40GB，购买数量4台，如下图



* 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”，如下图



* 选择“全动态BGP”，“按流量计算”，“5M”，如下图



* 点击“高级配置”, 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”，如下图



* 点击“确认配置”，点中“我已经阅读并同意”，点击“立即购买”, 点击“我已经阅读并同意”，可以点击“返回云服务器列表”，创建过程需要等待几分钟。



购买OBS

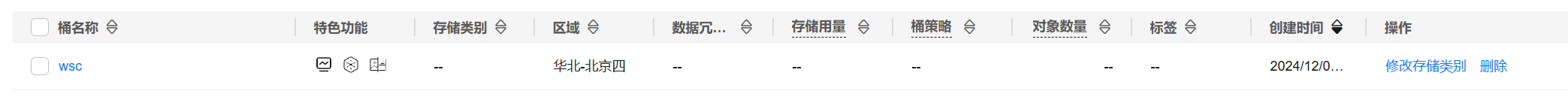
* 进入控制台，选择“对象存储服务”，如下图。



* 选择“创建桶”，如下图。

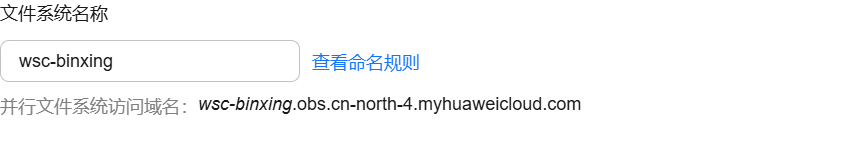


* 自定义桶名称，选择“标准存储”，点击“立即创建”，如下图。



* 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”，如下图。

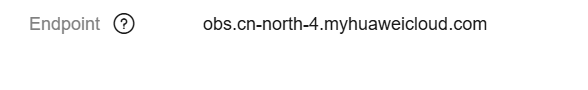




* 进入创建的OBS桶，如下图。



* 复制该参数，保存到本地文档，如下图。

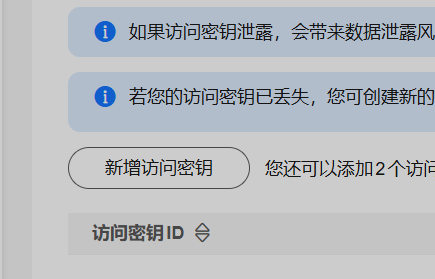


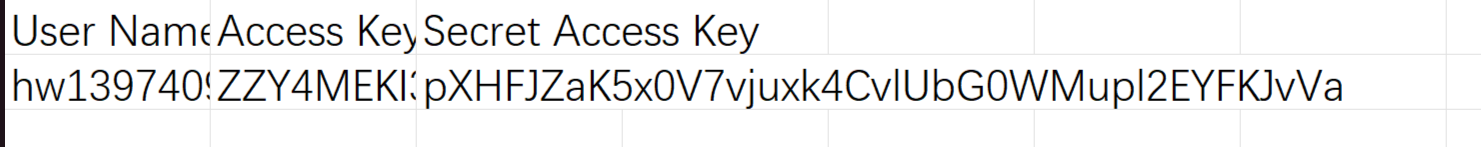
|  |
| --- |
| **obs.cn-north-4.myhuaweicloud.com** |

* 获取AK/SK，点击“我的凭证”，选择“访问秘钥”,如下图。

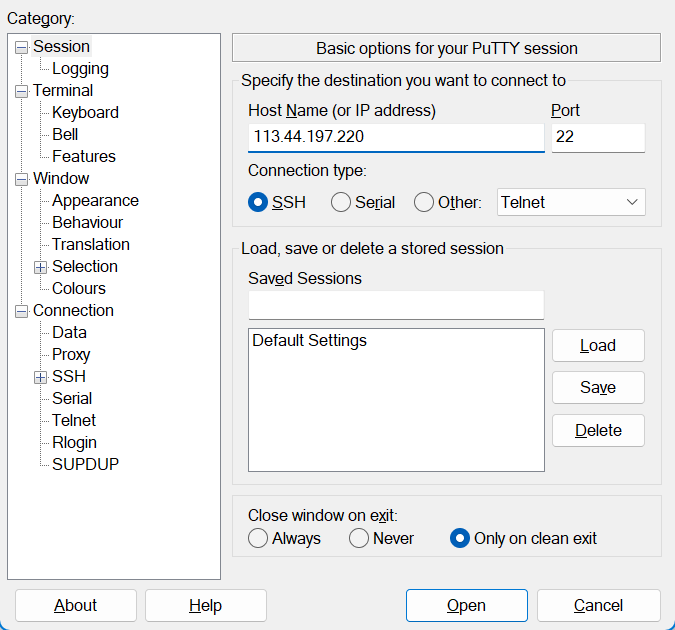


* 点击新增访问秘钥，根据提示进行操作。



* 通过手机号码，接收短信验证码，点击“确定”，如下图。操作完成后，得到文件，打开即可得到AK/SK

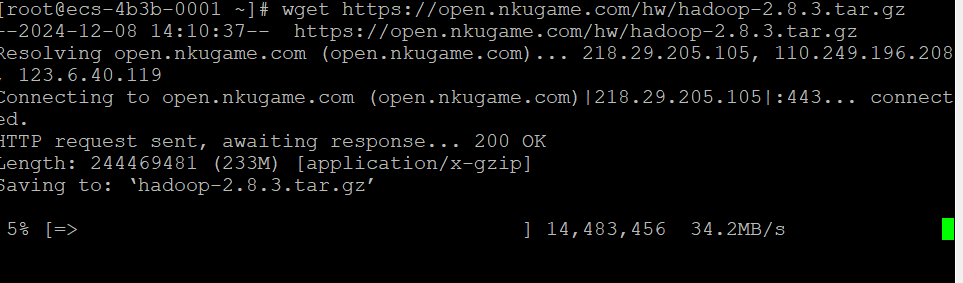
## 搭建Hadoop集群





node1节点登录成功后执行命令在线下载安装包

wget <https://open.nkugame.com/hw/hadoop-2.8.3.tar.gz>



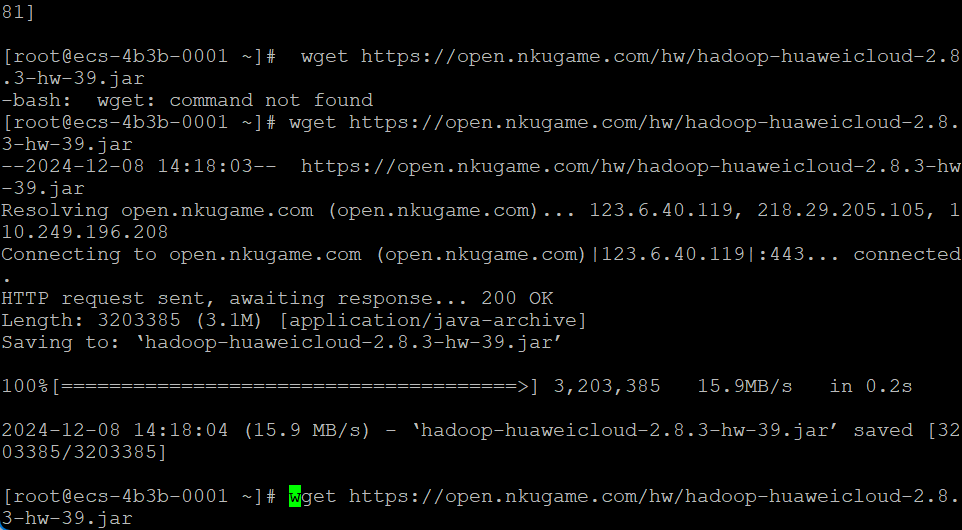
下载OBSFileSystem

wget <https://open.nkugame.com/hw/hadoop-huaweicloud-2.8.3-hw-39.jar>



下载OpenJDK

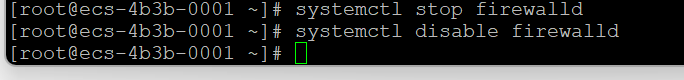
wget <https://open.nkugame.com/hw/hadoop-huaweicloud-2.8.3-hw-39.jar>



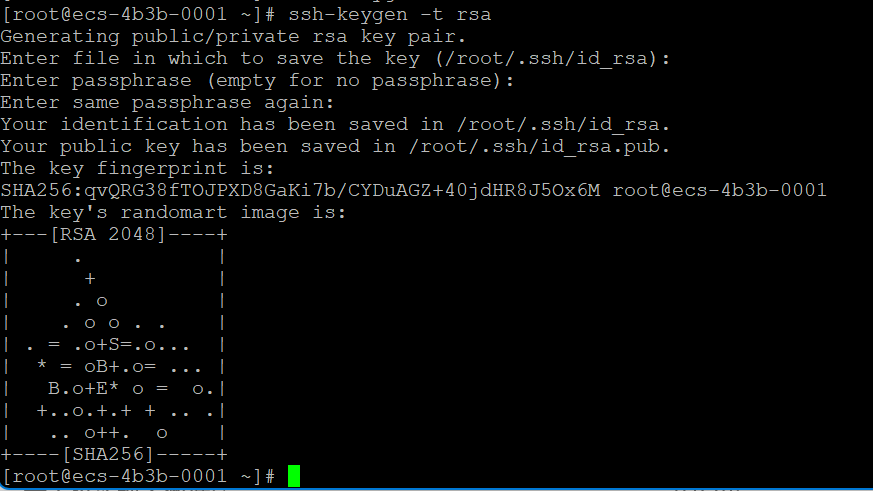
关闭防火墙

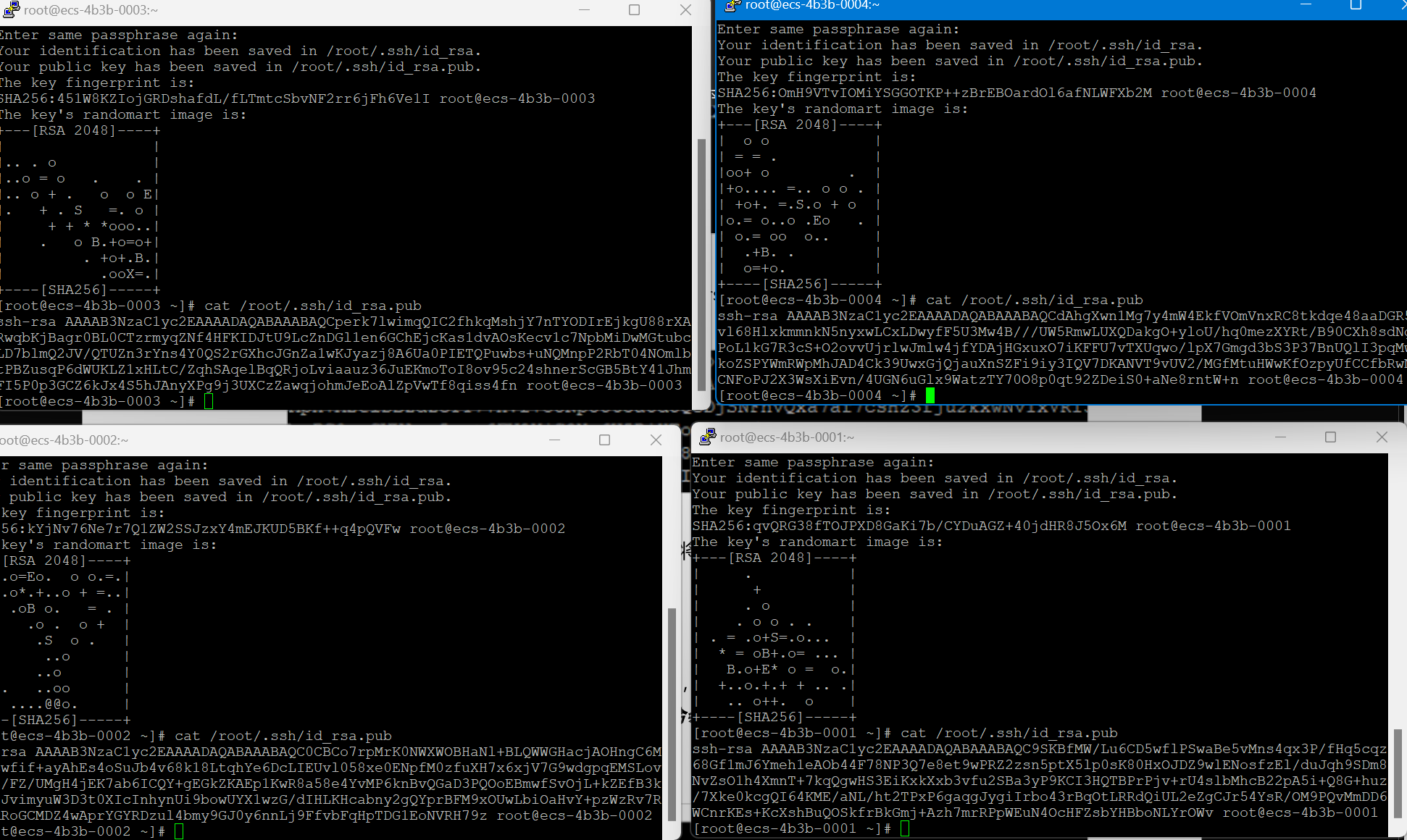
systemctl stop firewalld

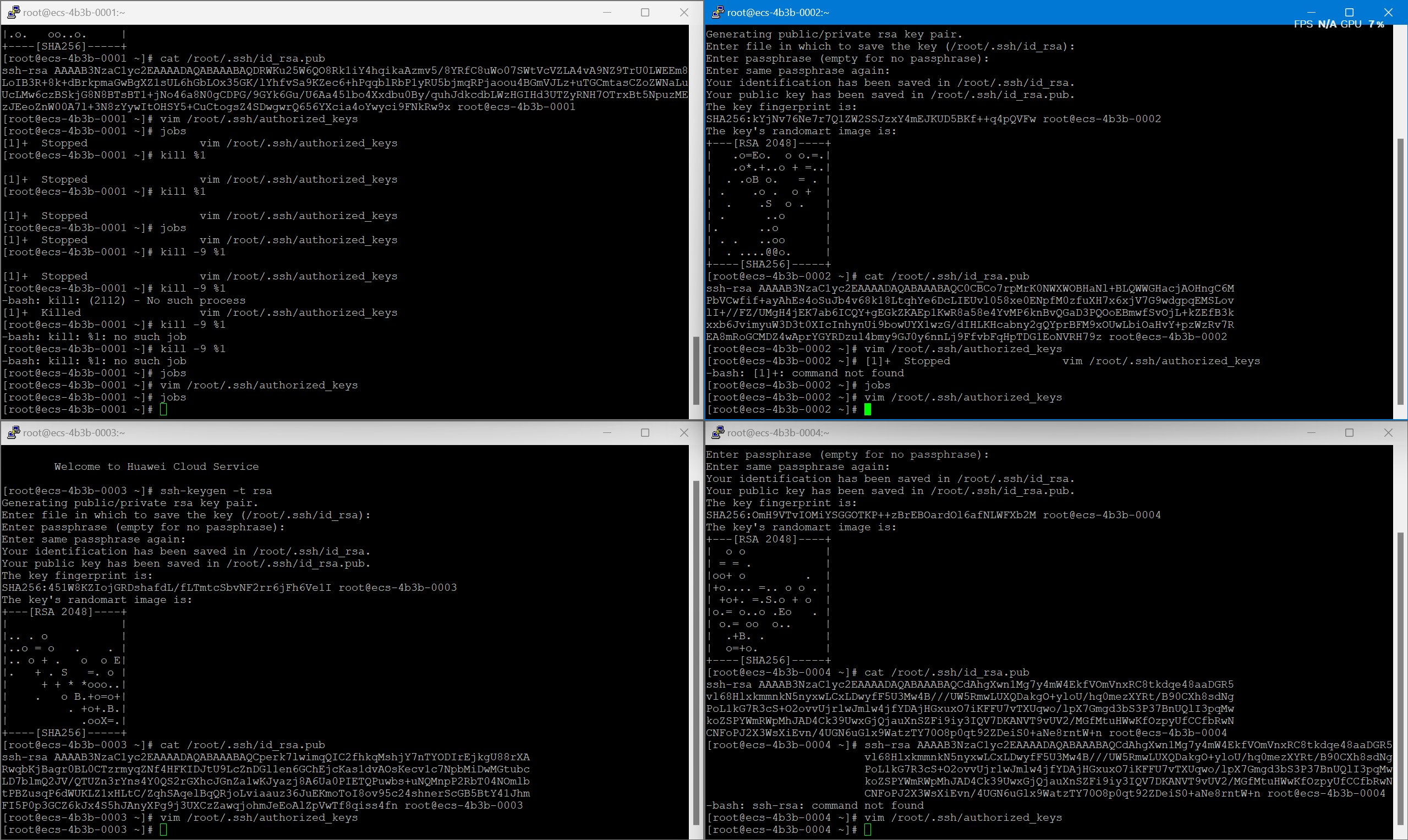
systemctl disable firewalld

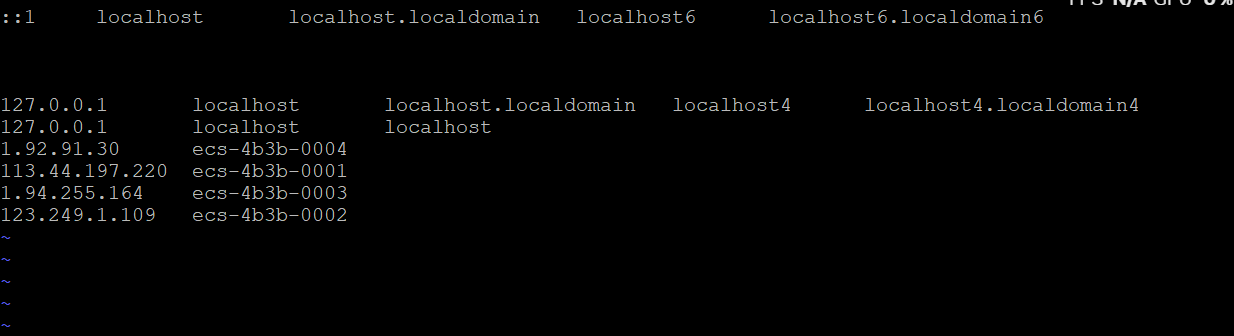


配置节点互信









### 安装OpenJDK

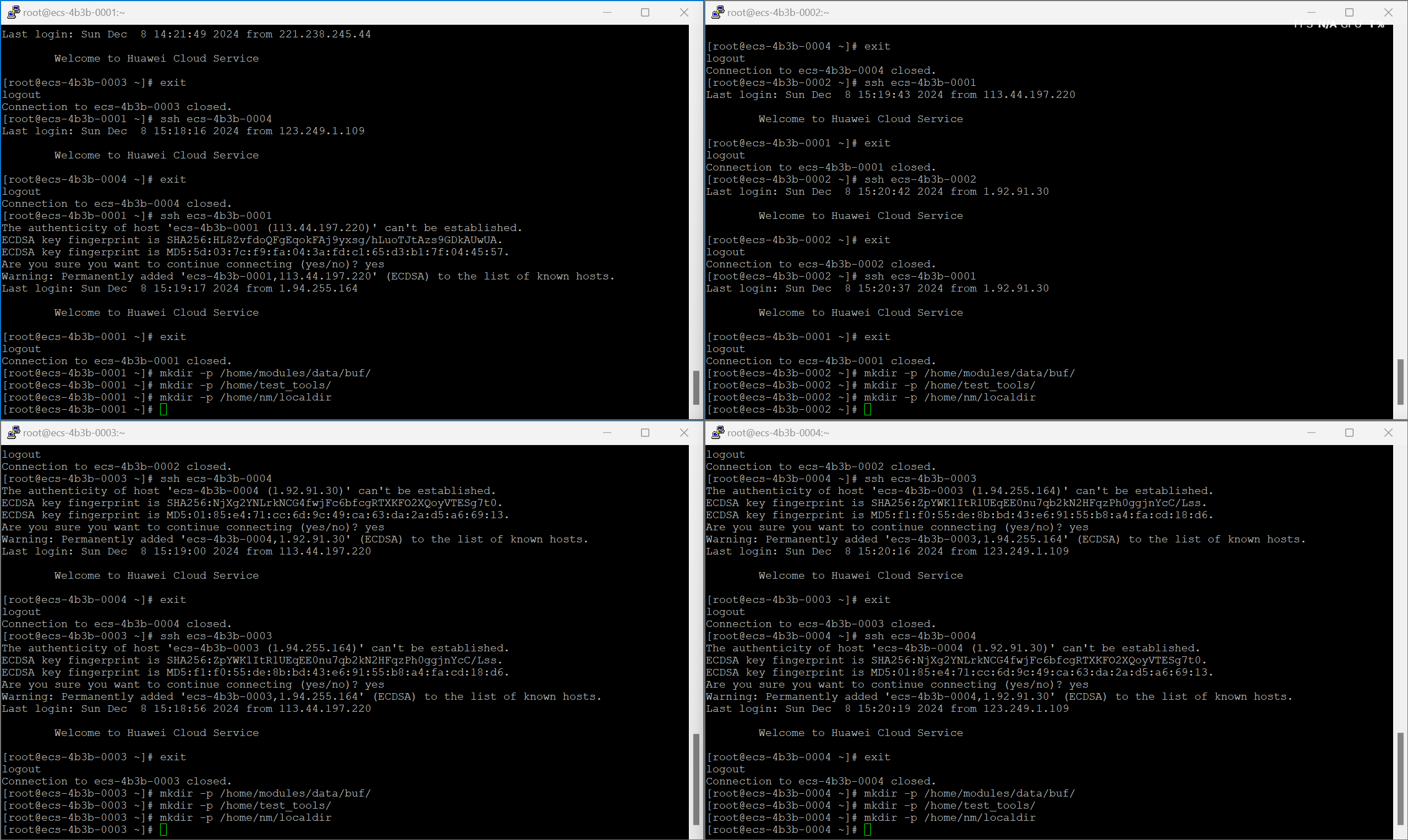
创建目录

node1~node4四个节点分别执行下列命令。

mkdir -p /home/modules/data/buf/

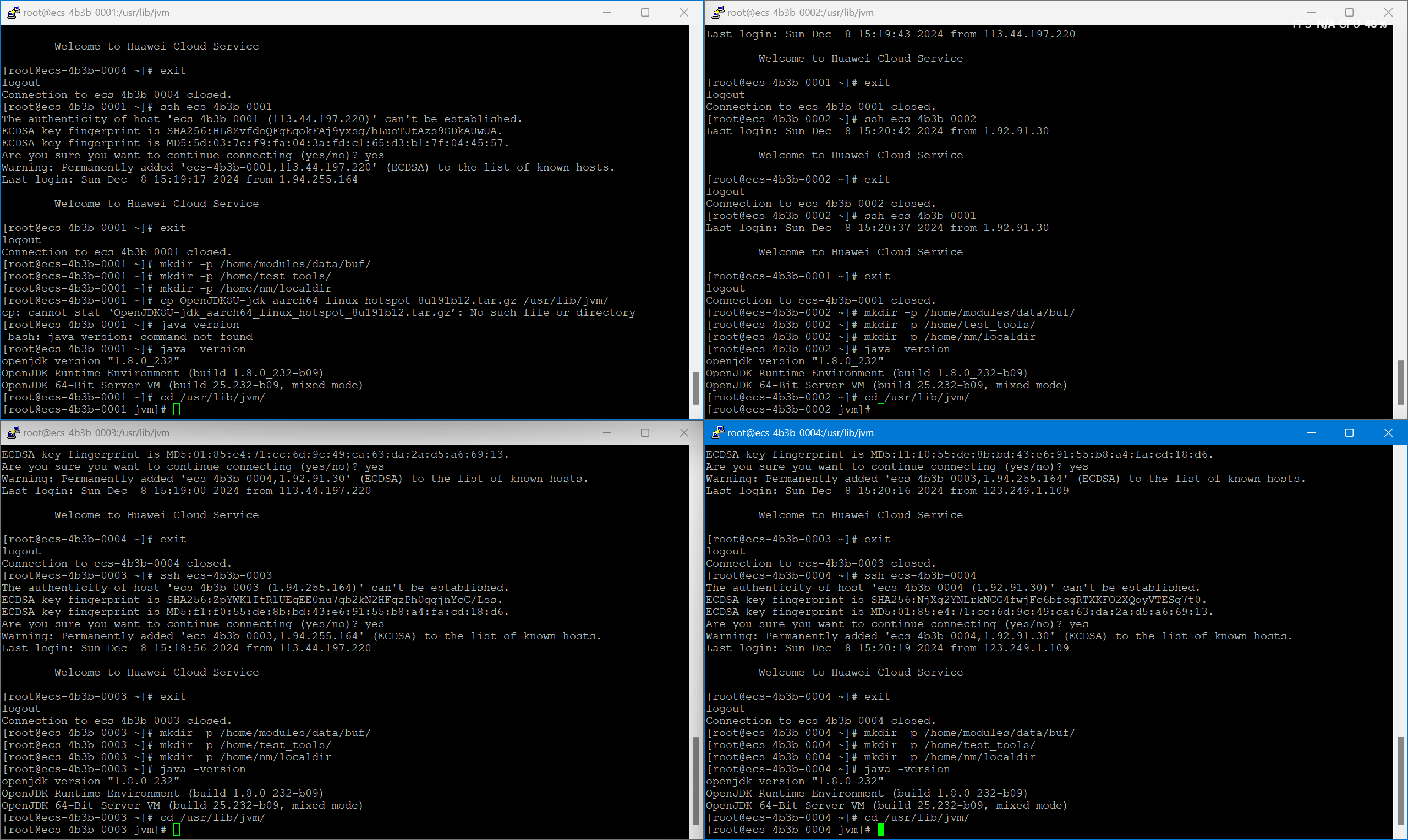
mkdir -p /home/test\_tools/

mkdir -p /home/nm/localdir

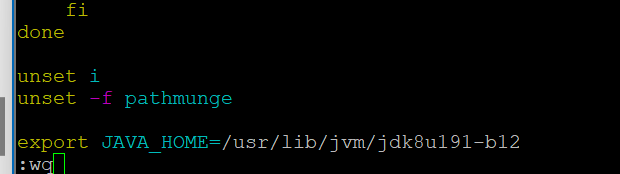


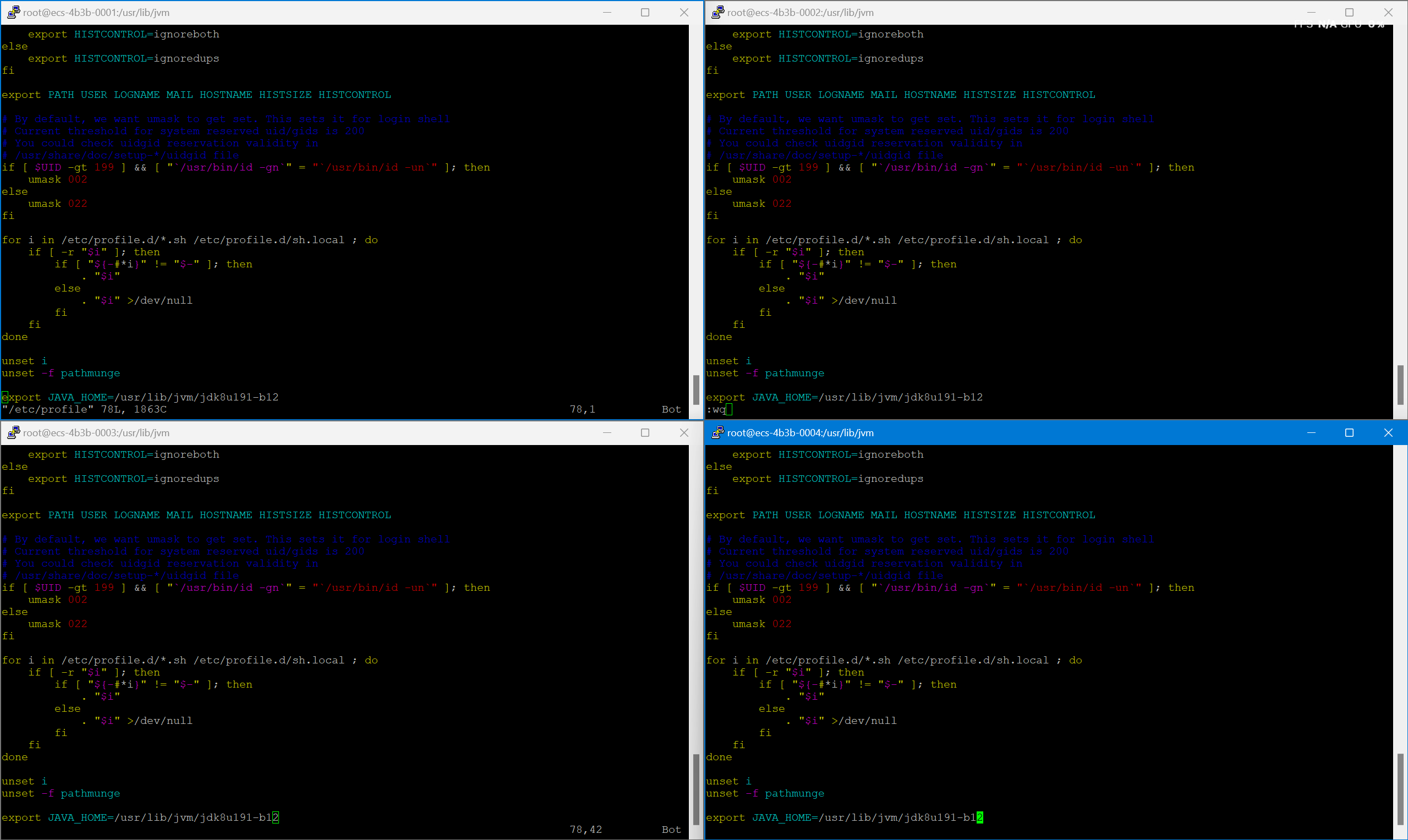
安装OpenJDK

在node1~node4四个节点分别执行命令

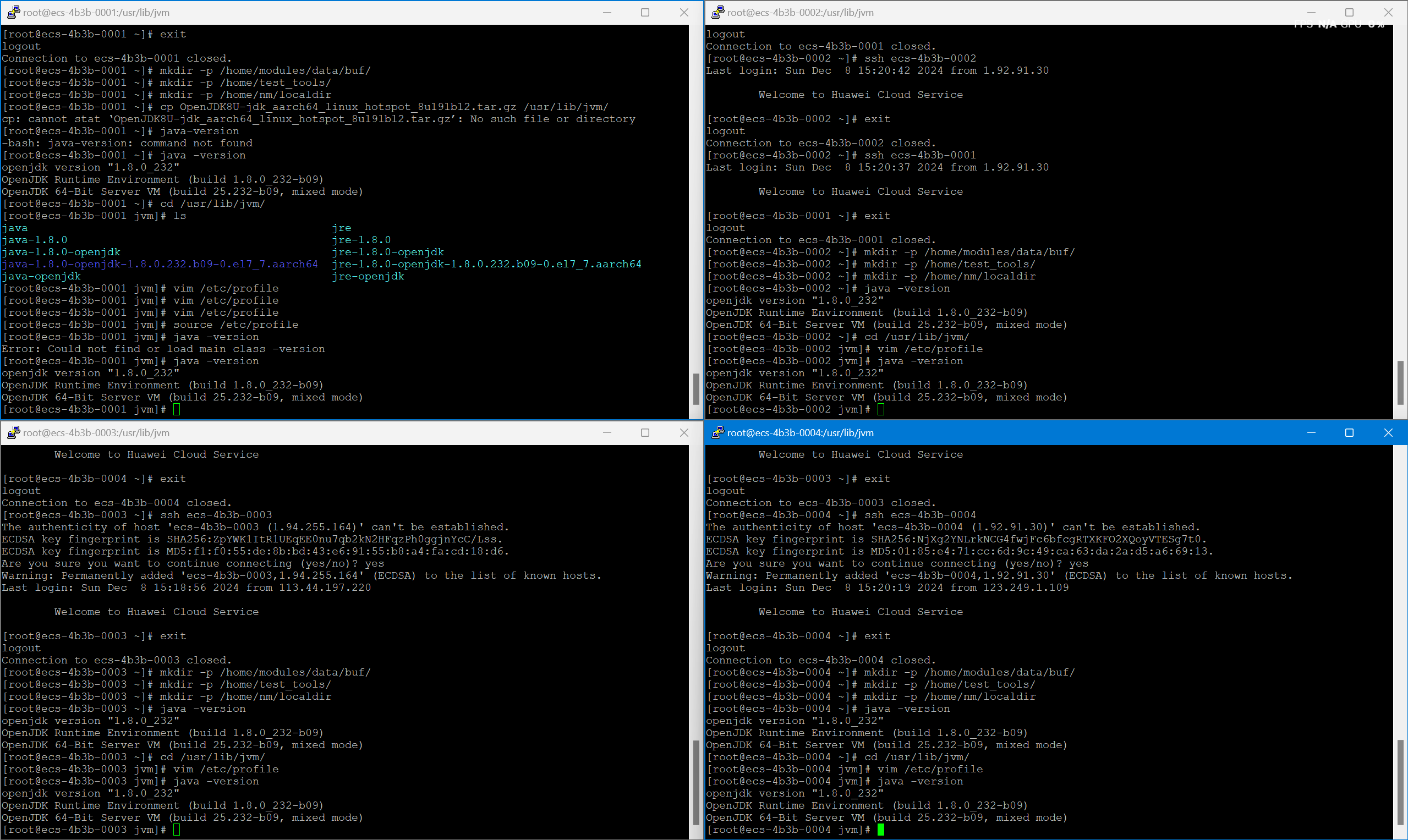


在node1~node4四个节点上编辑/etc/profile增加如下的配置





确认Java版本



### 搭建集群

### 登录node1节点，解压hadoop安装包

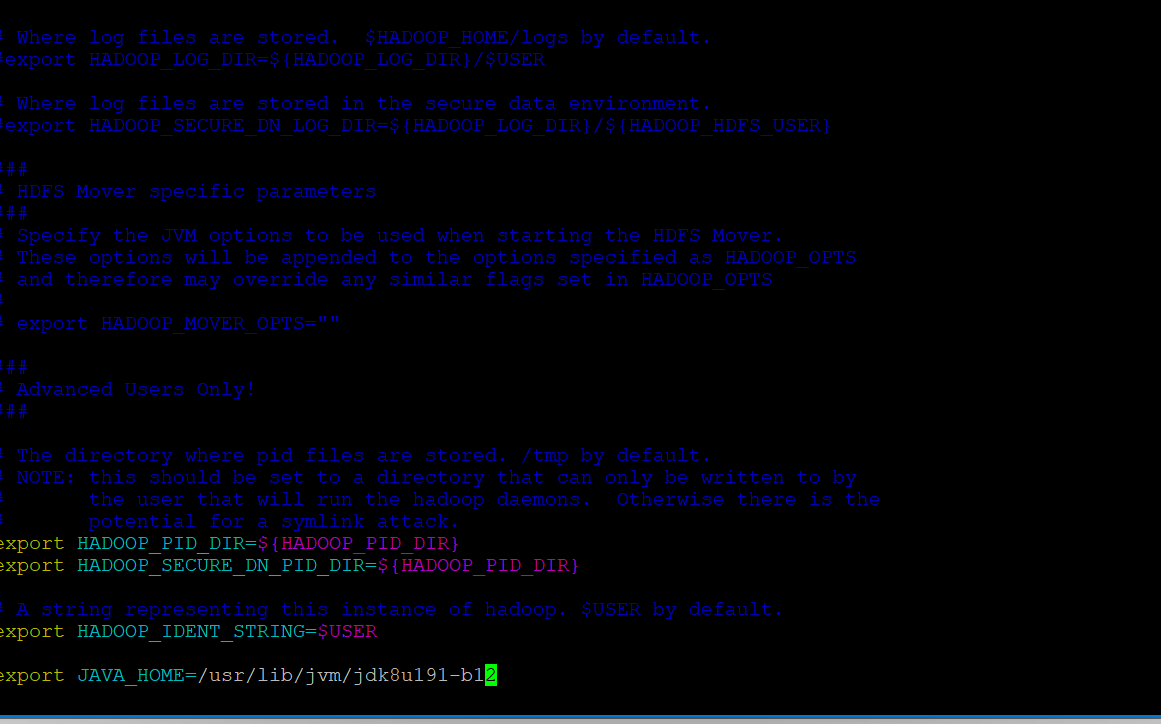
cp hadoop-2.8.3.tar.gz /home/modules/

cd /home/modules/

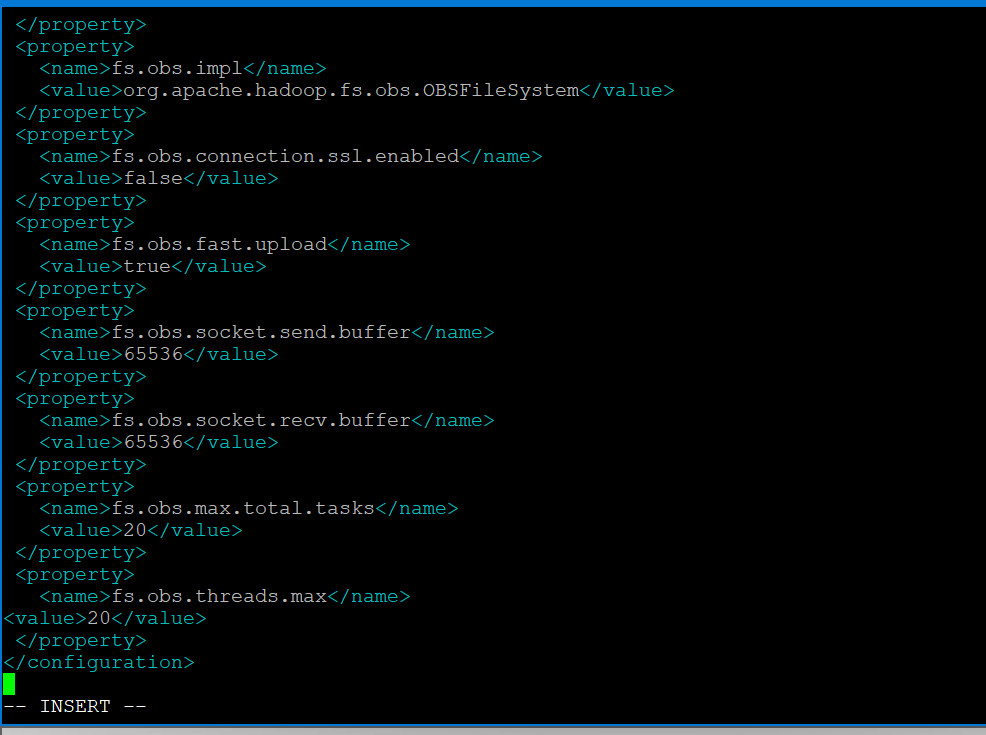
tar zxvf hadoop-2.8.3.tar.gz

### 

配置hadoop环境变量



修改xml文件:



### 配置site.sml文件

### vim /home/modules/hadoop-2.8.3/etc/hadoop/hdfs-site.xml

### 

### 配置yarn.site.xml文件

vim /home/modules/hadoop-2.8.3/etc/hadoop/yarn-site.xml

### 

配置mapred-sit.xml

cd /home/modules/hadoop-2.8.3/etc/hadoop/

mv mapred-site.xml.template mapred-site.xml

vim /home/modules/hadoop-2.8.3/etc/hadoop/mapred-site.xml

### 

### 

### 配置jar包

### 

### 

### 分发:

### 

### 配置环境变量:

### 

### 

### 启动HDFS

### 

### 

## 测试与OBS互联MapReduce程序实验

### Node1节点启动yarn执行命令

start-yarn.sh

### 测试文件:

* palyerinfo.txt内容如下：

Alex James Lax Genu

Kerry Mary Olivia William

Hale Edith Vera Robert

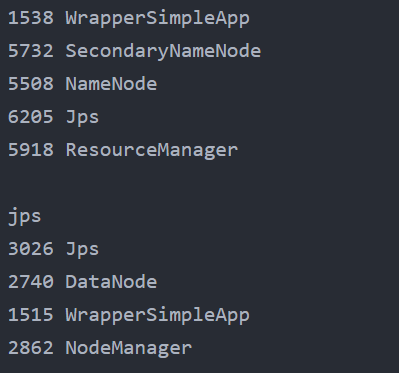
Mary Olivia James Lax

Edith Vera Robertm Genu

### 执行 hadoop wordcount

hadoop jar /home/modules/hadoop-2.8.3/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.3.jar wordcount obs://wsc/ /output





## 释放云服务资源

### 释放弹性云服务器ECS

### 

### 

### 删除OBS桶

### 

# 实验结果或结论

## 实验结果

实验成功完成,配置了四台云服务器,成功完成了hadoop实验集群搭建

* 本次实验旨在通过华为云搭建 Hadoop 集群，成功配置了四台云服务器，并完成了 Hadoop 集群的搭建。实验的过程包括以下步骤：

1. **创建云服务器**：
   * 使用华为云平台创建了四台云服务器，作为 Hadoop 集群的节点。
2. **配置环境**：
   * 在每台云服务器上安装并配置了 Hadoop 2.8.3 版本，确保各节点环境一致。
   * 配置了 Hadoop 集群的基本参数，确保集群内的节点可以互相通信。
3. **配置主机文件**：
   * 修改了 /etc/hosts 文件，以便各节点能够通过主机名互相访问。
4. **SSH 配置**：
   * 配置了 SSH 免密登录，以确保从任意节点可以无密码登录到其他节点，这对于集群间的操作至关重要。
5. **启动 Hadoop 集群**：
   * 在主节点上启动了 Hadoop 集群，包括 NameNode 和 DataNode。成功启动后，集群处于健康状态。
6. **验证集群是否正常运行**：
   * 使用 jps 命令在各节点上验证 Hadoop 服务是否成功启动，确保各节点的 Hadoop 服务正常运行。
   * 在 Hadoop Web UI 界面检查集群的状态，确保数据的存储和处理都能正常进行。

## 结论

* 通过本次实验，我成功完成了华为云上的 Hadoop 集群搭建。四台云服务器的配置和联通顺利完成，集群的安装和配置都按预期进行。整个实验过程中没有出现任何严重的错误，所有节点能够正常工作，集群功能得到了验证。
* 实验成功完成，为进一步深入学习和实践 Hadoop 分布式计算打下了坚实的基础。