**《云计算平台技术实践》**

**本科生课程实验报告**

|  |  |
| --- | --- |
| **题目：** | **BigDataPro集群搭建实验** |
| **专业：** | **软件工程** |
| **班级：** | **软工一班** |
| **姓名：** | **李西洋** |
| **学号：** | **2020302111004** |

**武汉大学计算机学院**

**2023年 5 月 15 日**

个人型实验报告学术声明示例：

**郑 重 声 明**

（宋体粗体2号居中）

本人呈交的实验报告，是在指导老师的指导下，独立进行实验工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本实验报告不包含他人享有著作权的内容。对本实验报告做出贡献的其他个人和集体，均已在文中以明确的方式标明。本实验报告的知识产权归属于培养单位。

（宋体4号）

本人签名： 李西洋 日期： 2023/5/15

1. **实验1-基于华为云搭建BigDataPro集群。**
   1. **实验概述**

【实验名称】：基于华为云搭建BigDataPro集群。

【实验目的】：

掌握Hadoop集群搭建方法

掌握Hadoop集群与华为云OBS互联方法

掌握Spark集群搭建

掌握Spark集群与OBS互联，实现存算分离

【实验环境】：华为云平台

【参考资料】：华为云BigDataPro集群搭建指南

* 1. **实验内容**

【实验设计】：

实验资源：

|  |  |
| --- | --- |
| **设备名称** | **设备型号** |
| 华为云平台ECS服务器  node-0001 | 华为云计算，通用计算增强型，c3ne.large.4 | 2vCPUs | 8GB |
| 华为云平台ECS服务器  node-0002 | 华为云计算，通用计算增强型，c3ne.large.4 | 2vCPUs | 8GB |
| 华为云平台ECS服务器  node-0003 | 华为云计算，通用计算增强型，c3ne.large.4 | 2vCPUs | 8GB |
| 华为云平台对象存储服务OBS  bigdatapro-obs-lxy | 多AZ存储 |
| 华为云平台并行文件系统  bigtata-lxy | 多AZ存储 |

【实验过程】：

一、资源购买：

1.1购买华为云ECS云服务器：

点击控制台：



在页面左上角，选择区域“北京四”，单击“服务列表”，选择“计算 -> 弹性云服务器ECS”，进入弹性云服务器列表页面。



单击页面右上角的“购买弹性云服务器”。



进入弹性云服务器的基础配置页面，选择基础配置（按需计费，华为云计算，X86架构，通用计算增强型，2核8G内存，CentOS7.6，超高IO 200GiB），单击“下一步：网络配置”。





进入弹性云服务器的网络配置页面，其中网络选择“vpc-default –> subnet-default -> 自动分配地址”，安全组选择“sg-FullAccess”，公网带宽选择“按流量计费”和“100”Mbit/s带宽大小，单击“下一步：高级配置”。



进入弹性云服务器的高级配置页面，高级配置如下图所示：

云服务器名称：node

输入自定义密码，密码需大于8位并带有特殊字符，再次确认密码

单击“下一步：确认配置”。

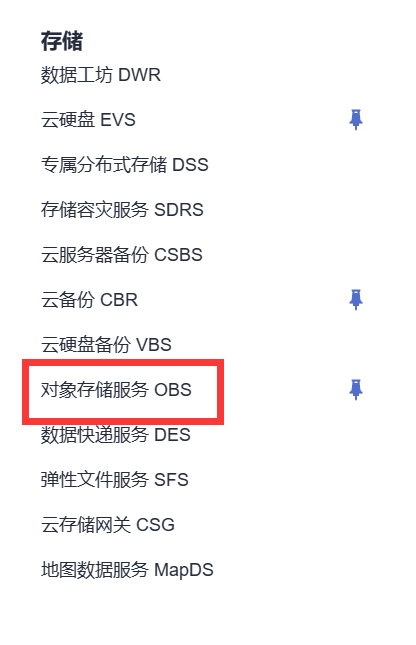


进入弹性云服务器的确认配置页面，核对信息无误后，选择购买三台，并勾选下方的“我已经阅读并同意《华为镜像免责声明》”，如果有企业项目使用下拉框选择企业项目，单击“立即购买”，完成云服务器购买：



1.2购买对象存储服务OBS

1.2.1选择对象存储服务OBS

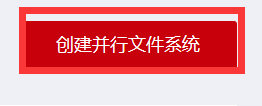


进入OBS服务后，点击右侧“创建桶”



参照下图购买OBS

1.2.2选择“并行文件系统”

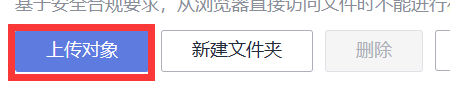


参照下图购买文件并行系统

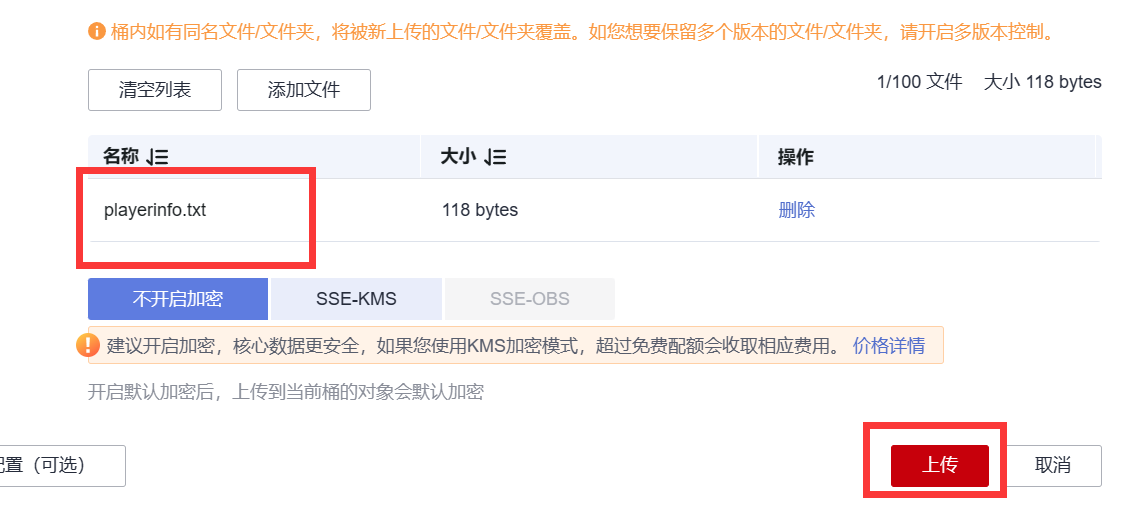


1.2.3进入创建的OBS桶，上传文件

点击上传对象按钮



选择playerinfo.txt文件上传



结果如下：



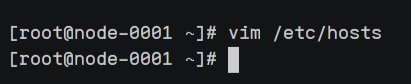
二、Hadoop集群搭建

2.1配置ECS

2.1.1将hadoop-2.8.3.tar.gz、hadoop-huaweicloud-2.8.3-hw-39.jar拖入root文件夹

2.1.2配置/etc/hosts文件

在各节点执行：



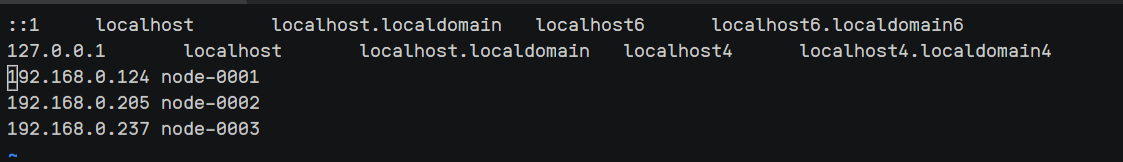
添加内容：

192.168.0.124 node-0001

192.168.0.205 node-0002

192.168.0.237 node-0003

参考如下：

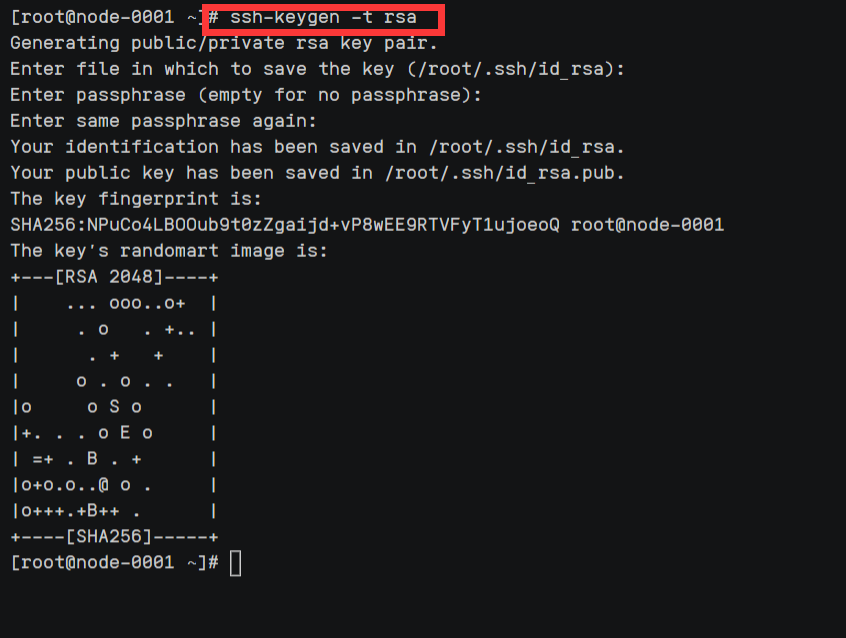


保存退出。

2.1.3配置节点互信

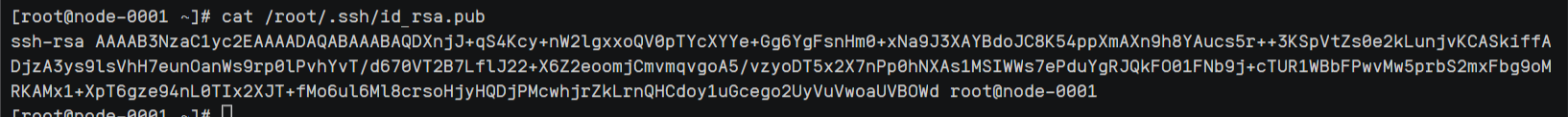
各节点执行：ssh-keygen -t rsa，

连续回车生成/root/.ssh/id\_rsa.pub文件；



各节点执行：cat /root/.ssh/id\_rsa.pub

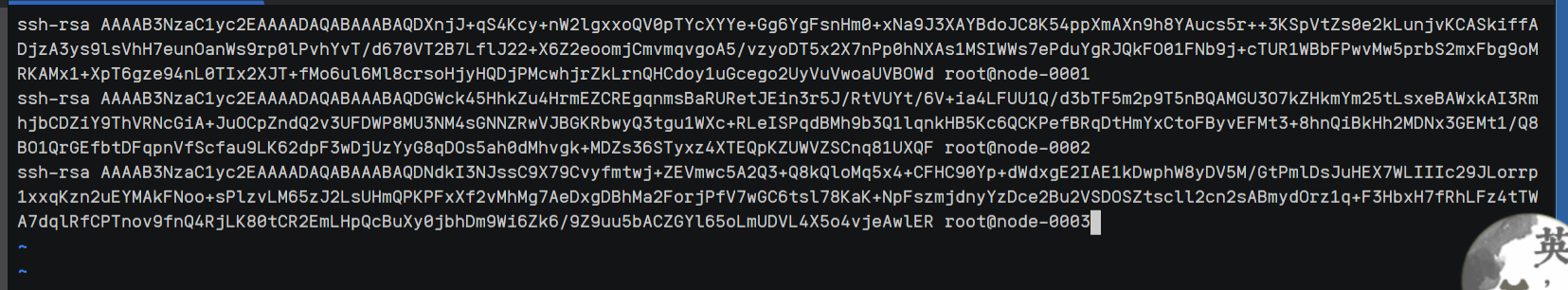
输出如下内容：



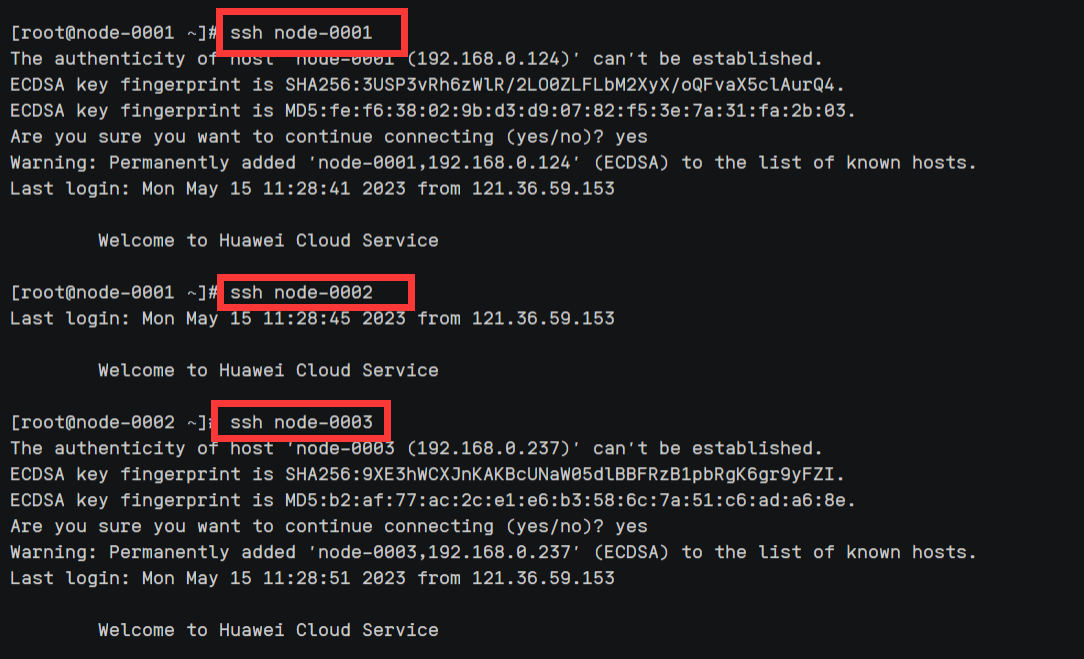
各节点执行：

vim /root/.ssh/authorized\_keys

输入各节点的复制内容，保存退出：



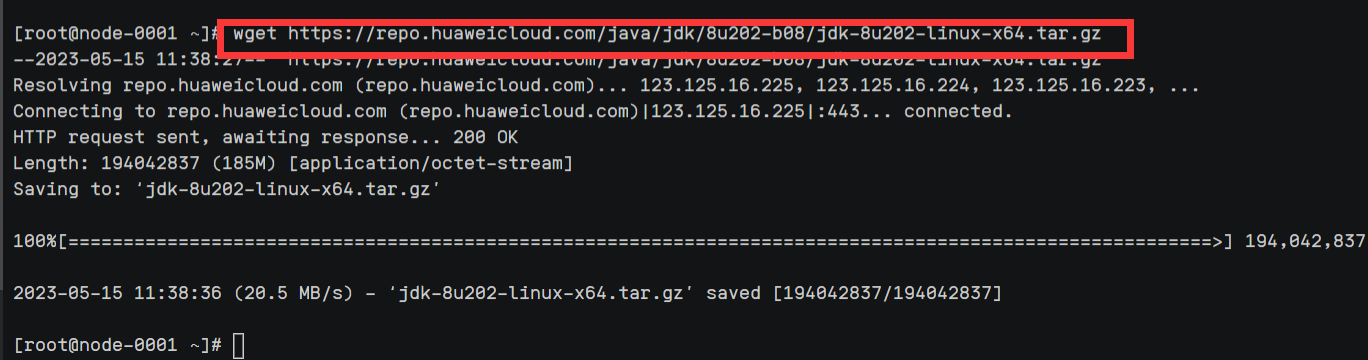
各节点执行：ssh node-0001~node-0003，选择yes后，确保能够无密码跳转到目标节点。



2.2安装jdk：:

2.2.1下载jdk1.8：

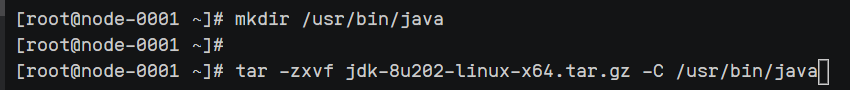
wget <https://repo.huaweicloud.com/java/jdk/8u202-b08/jdk-8u202-linux-x64.tar.gz>



2.2.2创建安装目录mkdir /usr/bin/java

2.2.3解压至安装目录

tar -zxvf jdk-8u202-linux-x64.tar.gz -C /usr/bin/java



2.2.4配置环境变量

打开文件

vim /etc/profile

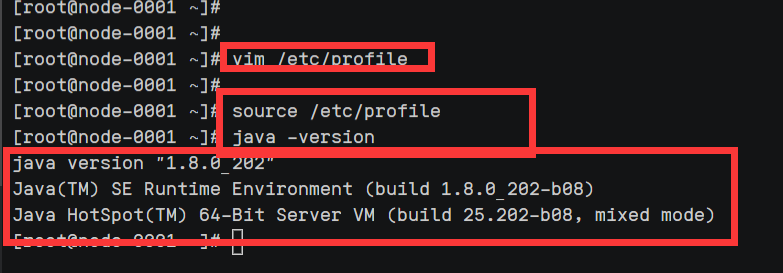
在末尾添加



2.2.5使环境变量生效

source /etc/profile

2.2.6检查 java -version



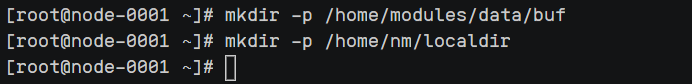
2.3搭建Hadoop集群

2.3.1创建目录

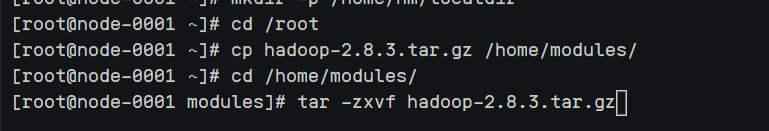
各节点执行：

mkdir -p /home/modules/data/buf

mkdir -p /home/nm/localdir



2.3.2登录node-0001节点，解压hadoop安装包



2.3.3配置hadoop core-site.xml配置文件

node-0001节点执行下列命令：

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



2.3.4配置hdfs-site.xml

node-0001节点执行下列命令：

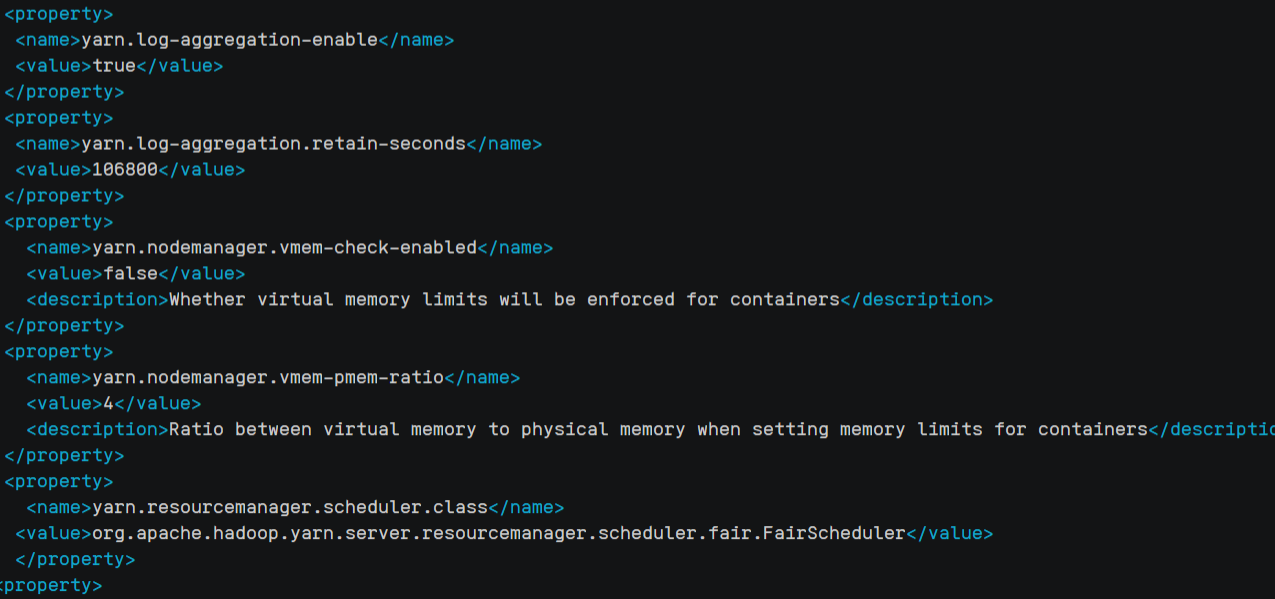
vim /home/modules/hadoop-2.8.3/etc/hadoop/hdfs-site.xml



2.3.5配置yarn-site.xml

node-0001节点执行下列命令：

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



2.3.6配置mapred-site.xml

node-0001节点执行下列命令：

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

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

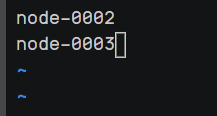
vim /home/modules/hadoop-2.8.3/etc/hadoop/mapred-site.xml

2.3.7配置slaves

node-0001节点执行下列命令：

vim /home/modules/hadoop-2.8.3/etc/hadoop/slaves

删除原有内容，添加：



2.3.8配置hadoop环境变量

node-0001节点执行下列命令：：

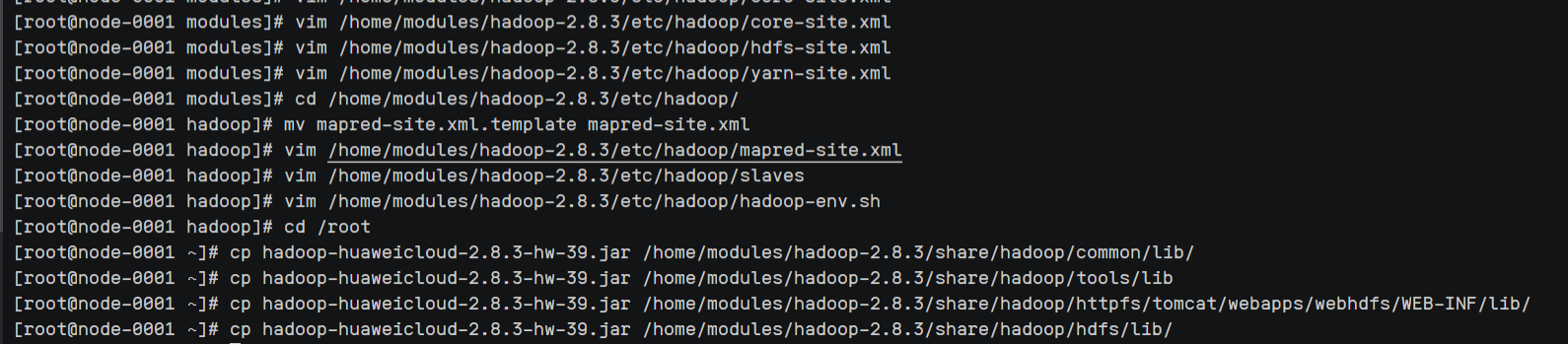
vim /home/modules/hadoop-2.8.3/etc/hadoop/hadoop-env.sh

添加如下内容：

IMG_256

2.3.9配置jar包

node-0001节点执行下列命令：



2.3.10分发hadoop包到各节点

node-0001下执行下列命令：

分发hadoop包到node2

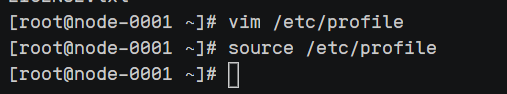
scp -r /home/modules/hadoop-2.8.3/ root@node-0002:/home/modules/

分发hadoop包到node3

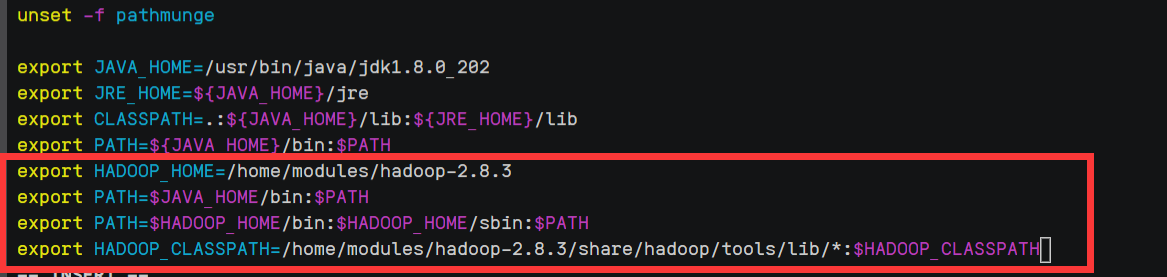
scp -r /home/modules/hadoop-2.8.3/ root@node-0003:/home/modules/

2.3.11配置环境变量

各节点执行：



在文件尾部添加如下内容：

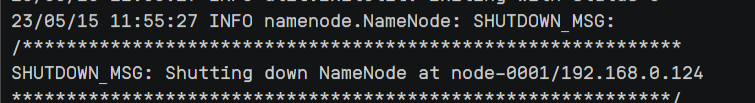


2.3.12namenode初始化

node-0001节点执行namenode初始化

执行下列命令：

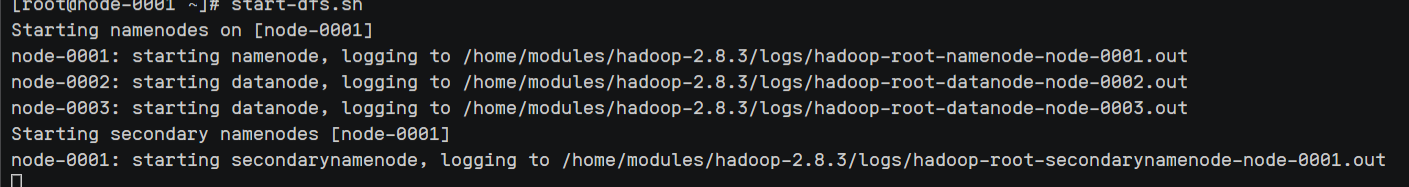
hdfs namenode -format



2.3.13启动HDFS

node-0001节点执行

IMG_256

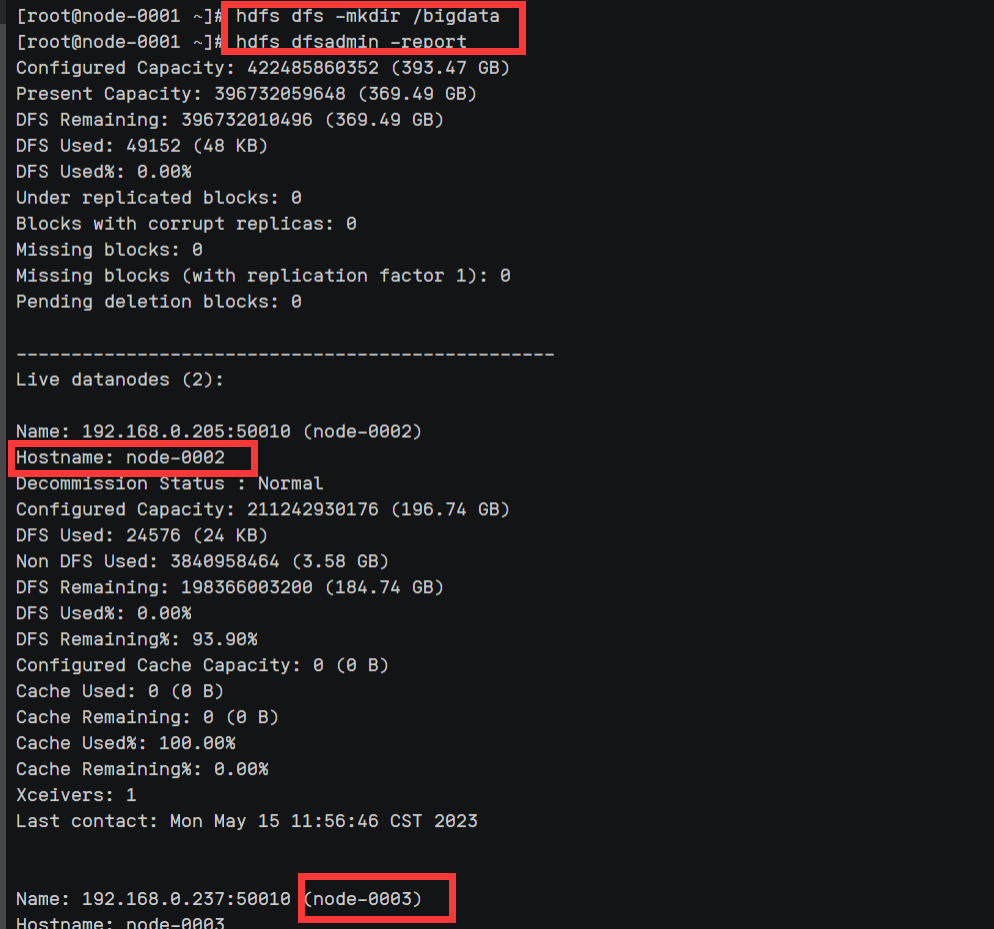


2.3.14执行hdfs命令

hdfs dfs -mkdir /bigdata

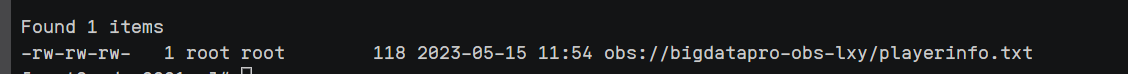
2.3.15查看集群部署报告

hdfs dfsadmin -report



2.4测试与OBS互联

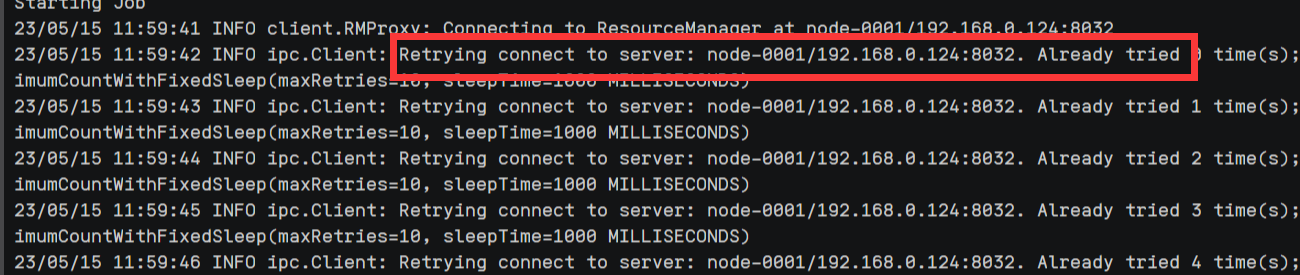
2.4.1执行hdfs命令查看OBS文件IMG_256



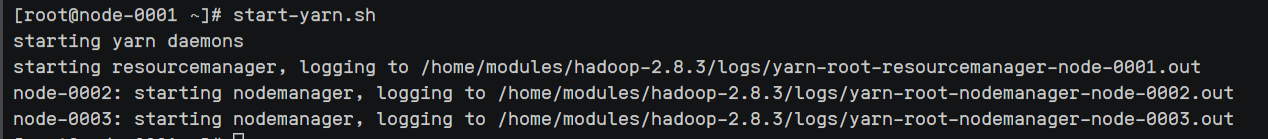
2.4.2测试能否跑任务

hadoop jar /home/modules/hadoop-2.8.3/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.3.jar pi 10 10

如果出现以下问题：

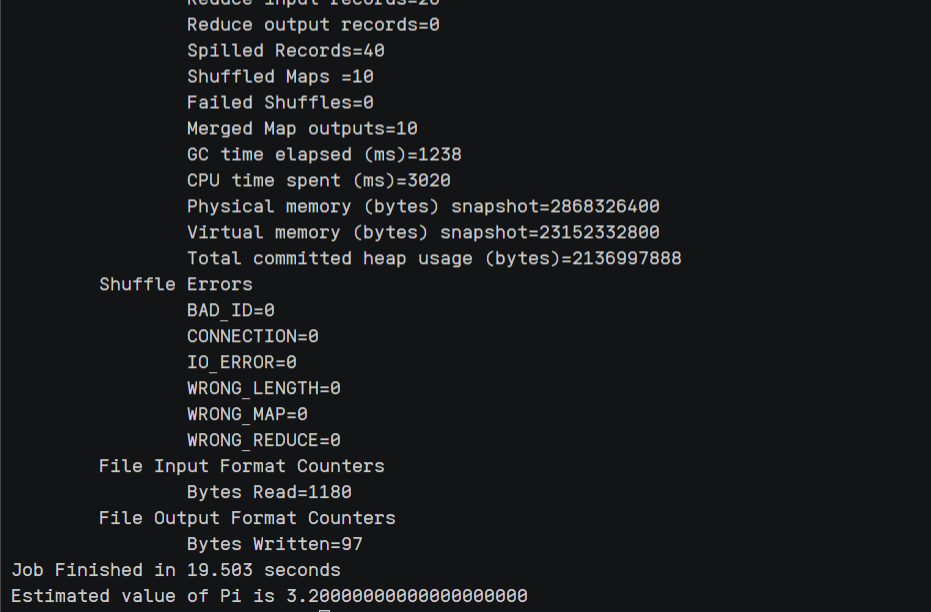


执行以下命令：



重新执行：

hadoop jar /home/modules/hadoop-2.8.3/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.3.jar pi 10 10



出现3.20成功

三、搭建Spark集群

3.1搭建Spark集群

3.1.1将安装包拖入root文件夹

3.1.2解压spark安装包

node-0001节点执行下列命令：

复制安装包到/home/modules目录下

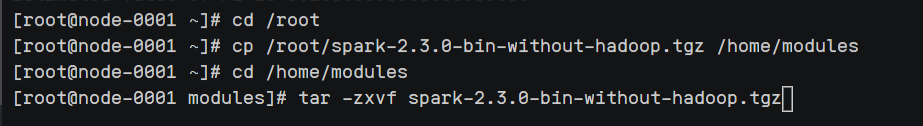
cp /root/spark-2.3.0-bin-without-hadoop.tgz /home/modules

cd /home/modules

解压安装包

tar -zxvf spark-2.3.0-bin-without-hadoop.tgz

mv spark-2.3.0-bin-without-hadoop spark-2.3.0



3.1.3配置spark jar包

在node-0001节点，复制jar包到spark/jar下

cp /root/hadoop-huaweicloud-2.8.3-hw-39.jar /home/modules/spark-2.3.0/jars/

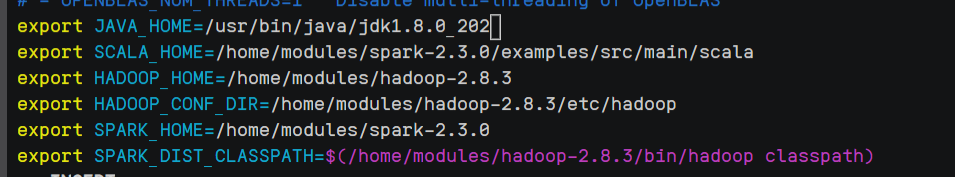
cp /home/modules/hadoop-2.8.3/share/hadoop/common/lib/snappy-java-1.0.4.1.jar /home/modules/spark-2.3.0/jars/

3.1.4配置spark配置文件

node-0001节点执行下列命令：



文件末尾添加如下内容：

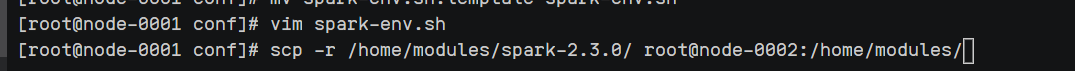


3.1.5分发Spark

node-0001节点执行下列命令：

scp -r /home/modules/spark-2.3.0/ root@node-0002:/home/modules/

scp -r /home/modules/spark-2.3.0/ root@node-0003:/home/modules/

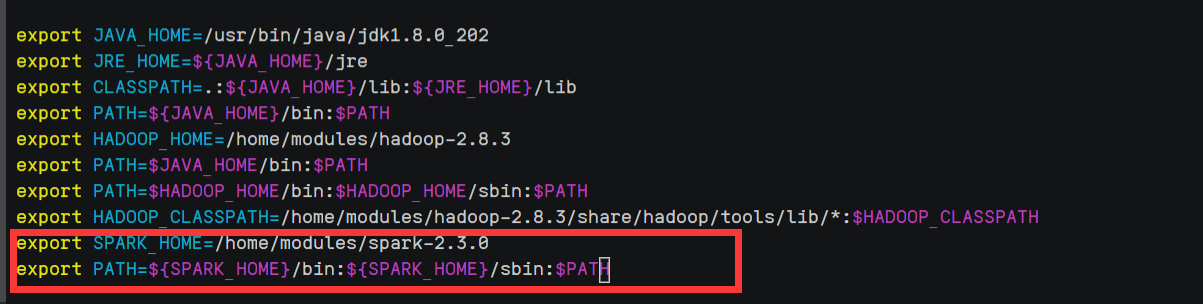


3.1.6配置环境变量

各节点执行：

vim /etc/profile

添加如下内容：

保存退出

各节点执行如下命令，使环境变量生效：

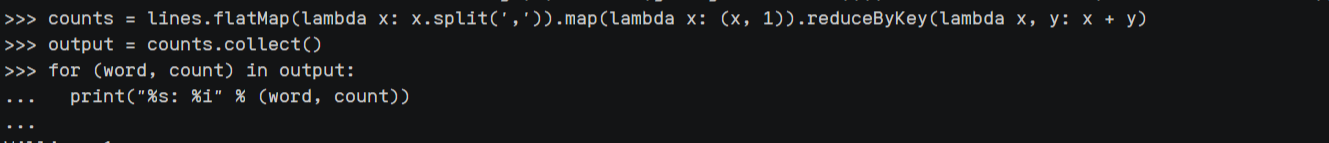
source /etc/profile

3.2验证存算分离

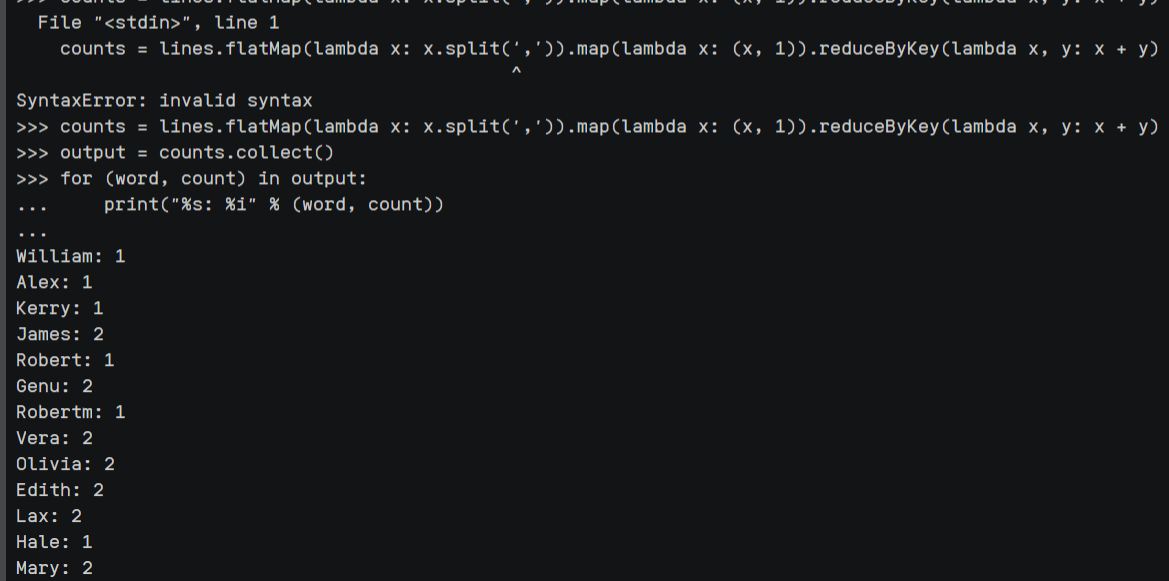
3.2.1启动pyspark



3.2.2输入下列代码：



出现以下结果则成功：



**实验总结**

【主要工作与收获】

通过本次实验，我学会了如何在华为云平台上搭建BigDataPro集群，掌握了Hadoop集群搭建方法、Hadoop集群与华为云OBS互联方法、Spark集群搭建、Spark集群与OBS互联以及存算分离的实现。

我对华为云的使用进一步了解，同时自己手动搭建集群的经历让我对云计算的概念得到进一步理解，对一组独立的计算机系统之间通过网络实现进程间的通信的过程有了初次的实践接触，培养了我对软件工程实验的兴趣。

【存在问题与展望】

在实验过程中，节点有时会突然连接不上，后来发现是format过多的原因，实际上像这种高封装的服务部署，每一个步骤都有其内在的含义，不能随随便便、粗心大意，任何遗漏都可能导致链条断裂，服务崩溃。计算机实验是非常精细的，在任何一个地方都不能含糊，一旦含糊就有可能付出极大的代价来修正错误。所以在未来的实验中，一定要小心谨慎，宁可把时间投入在过程中，也决不能浪费在纠错中。当然，错误本身也会加深我对实验的理解。

教师评语评分

评语：

评分：

评阅人：

年 月 日