****

**实验一：BigData Pro**

**实验报告**

|  |  |
| --- | --- |
| **学 期：** | 2023-2024学年 第二学期 |
| **专 业：** | 计算机科学与技术（实验班） |
| **学号姓名：** | 21071003 高立扬 |
| **完成日期：** | 2024.05 |

### 问题

1. 实验内容

本实验基于华为云OBS和华为云ECS服务构建一个存算分离的基本架构，并通过运行一个计算程序来完成存算分离架构的验证。本实验的实验数据存储在OBS中，通过在ECS上部署开源组件（Hadoop和Spark）构成计算环境，最后编写Spark程序访问存储在OBS上的数据进行计算（单词出现次数统计）并输出结果。

本实验的基本步骤包含：

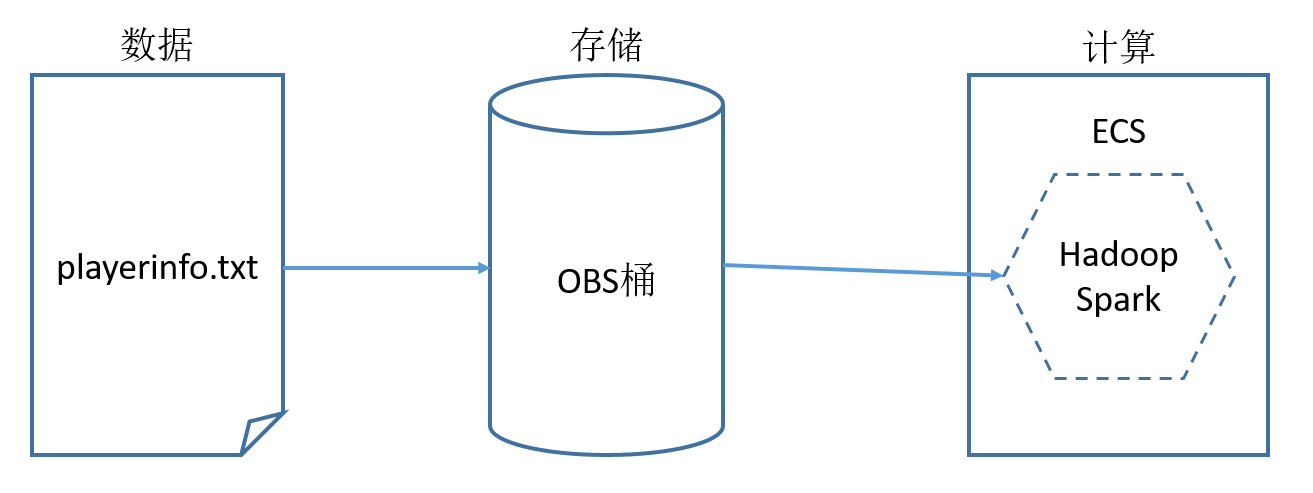
·购买并配置ECS；

·购买OBS并获取访问密钥AK/SK信息；

·搭建Hadoop集群；

·搭建Spark集群；

·编写Spark程序验证存算分离。



1. 实验目的

·掌握华为云OBS的购买和使用。

·掌握华为云ECS的购买和使用。

·掌握Hadoop\Spark环境搭建。

·掌握Spark程序读取OBS数据。

### 解决方法

1. 华为云服务器配置

1.1 弹性云服务器ECS。

选择“按需计费”，区域选择“北京四”，可用区“随机分配”。CPU架构选择“鲲鹏计算”，规格选择“鲲鹏通用计算增强型”中的“kc1.xlarge.2 4vCPUs | 8GB”。’镜像选择“openEuler”中的“20.03 64bit with ARM(40GB)）”，系统盘选择“超高IO”。

进入到网络配置页面，点击网络后面的下拉框选择VPC（后面的子网会自动选择，私有IP选择为“自动分配IP地址”），选择安全组。（如果没有VPC可以点击“前往控制台创建”）弹性公网IP（现在购买，全动态BGP，按流量计费，50）。

进入到高级配置页面，服务器名称“ecs-bigdatapro”（可自定义），登录凭证选择“密码”，输入密码和确认密码（点击密码框会有密码规则提示），其他默认。

进入确认配置界面，确认无误后购买数量选择“1”并勾选复选框。

返回弹性云服务器列表页面，可以看到刚刚购买的弹性云服务器ecs-bigdatapro，等待其状态变为“运行中”则购买成功。

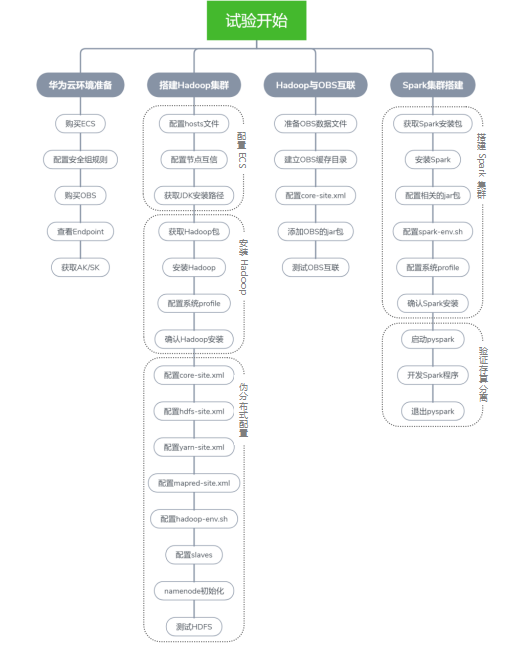
在云服务器控制台界面点击前面申请的ECS名称。进入ECS后选择“安全组”。展开Sys-default安全组，点击“更改安全组规则”。选择“入方向规则”，点击“添加规则”。点击下拉框，选择“基本协议”中的“全部放通”，然后点击“确定”。配置好安全组规则后，后面就可以通过远程工具进行连接了。如果提示规则已经存在，则无需处理。（实际上可以根据需要针对性的放开端口，这里为了试验方便选择了全部放通）

1.2 OBS服务

创建桶，区域选择“华北-北京四”（与ECS选择同一区域），根据命名规则输入桶的名称（桶的名称需要全局唯一请自行确定，本例中为obs-bigdatapro），更多配置（多AZ存储，标准存储，私有），其他默认。点击刚创建的桶名称，进入概览后下拉滚动条可以看到桶的基本信息，记录其中的Endpoint内容（IPv4地址），后面配置连接OBS时会使用。

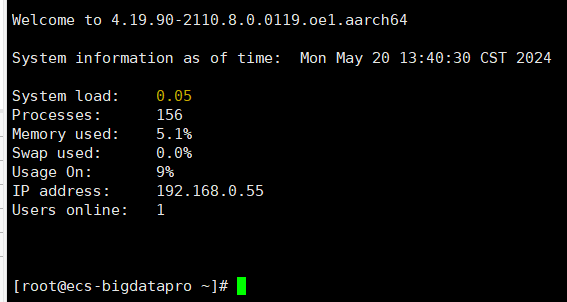
在控制台鼠标放到右上角的账号上，在弹出的菜单中点击“我的凭证”。在我的凭证界面点击“访问秘钥”，然后点击“新增访问秘钥”。新增秘钥需要身份验证，输入收到的短息验证码后点击“确定”（如手机无法使用也可以通过邮件验证）。创建成功点击“立即下载”下载秘钥文件。

1. 工作流程

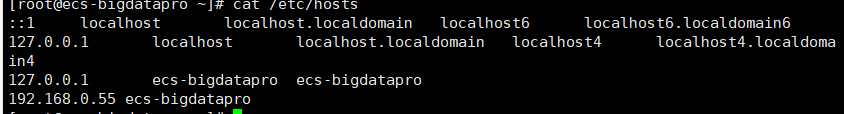


### 实验结果

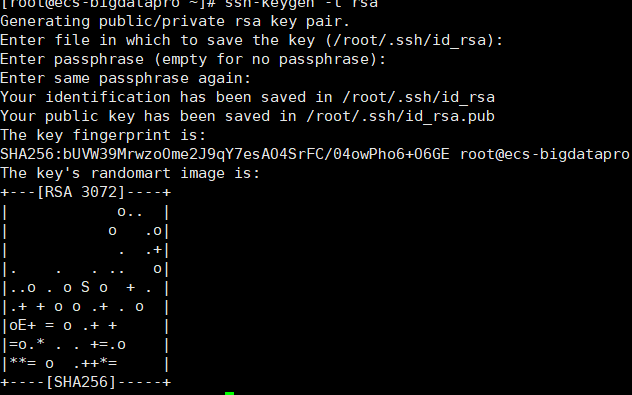
1. 登录



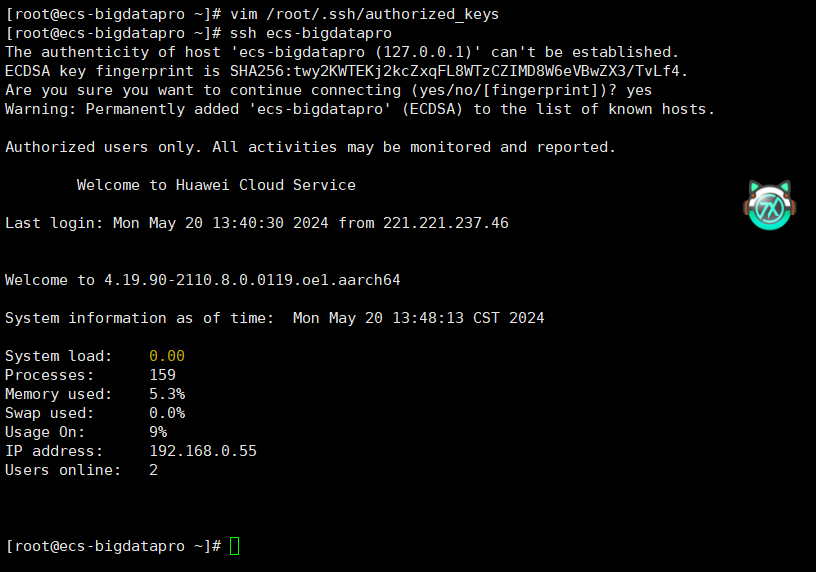
1. 步骤 2配置/etc/hosts文件



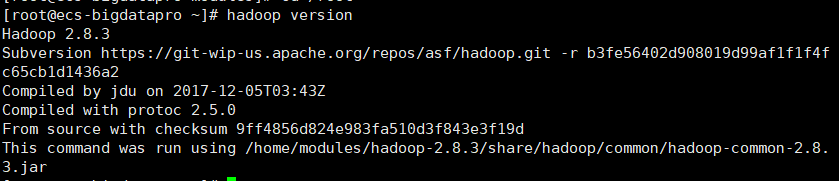
1. 配置节点互信



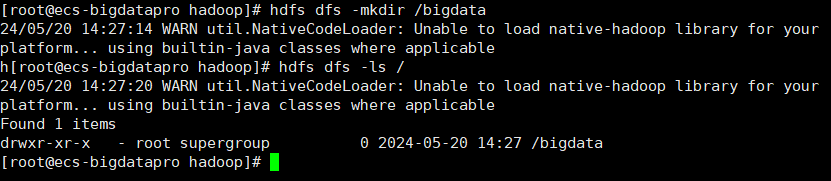
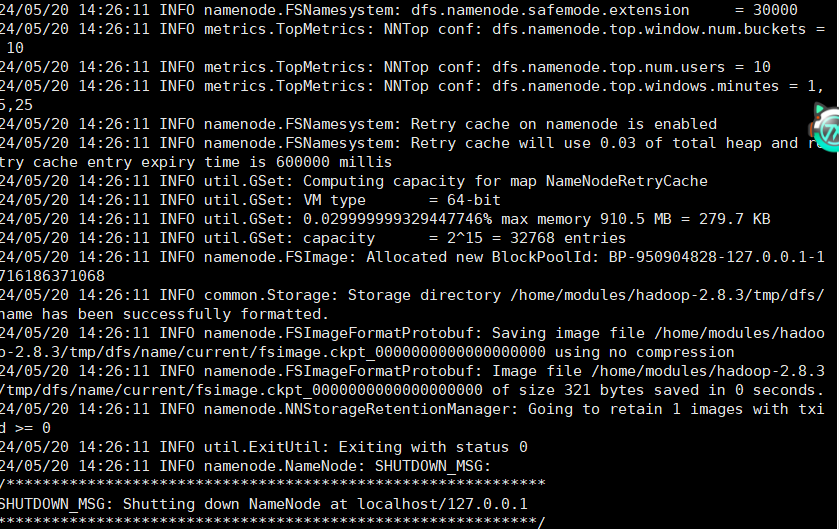
1. 验证互信



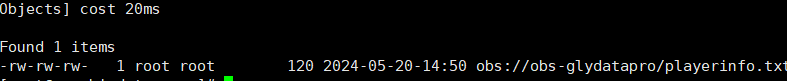
1. Hadoop安装



1. 伪分布式配置



1. Hadoop与OBS互联



1. spark

