

《云综合实训》报告二

题目：云主机实现大数据

学 院\_计算机科学与技术

班 级 计科20-4班

姓 名 边浩宇

学 号 12206112

中国矿业大学

2023年06月

课程设计报告格式要求：

1.报告中应包含一、二级标题；

2.正文宋体小4号，1.5倍行距；

3.报告正文不少于1000字；

4.上交报告电子版，各班学委统一收齐后打包发邮件到wjyang@cumt.edu.cn。

课程设计报告评分要点：

1. 报告中云计算技术的应用情况（20%）
2. 报告中功能有效及完整性（40%）
3. 报告中的个人总结及体会情况（20%）
4. 报告格式规范程度（20%）

评语：

评分： 日期：

目录

[一 实验环境准备 3](#_Toc7825)

[1.1购买 OBS 桶 3](#_Toc202)

[1.2购买 MRS 服务 4](#_Toc21074)

[1.3 购买弹性公网 IP 6](#_Toc13007)

[2 云服务实现大数据 7](#_Toc6409)

[2.1程序及数据准备 7](#_Toc31354)

[2.2执行程序 10](#_Toc22592)

[2.3查看作业日志 12](#_Toc15696)

[2.4使用组件管理 12](#_Toc17636)

[2.5思考题 15](#_Toc13323)

[3 资源清理释放 16](#_Toc5729)

[3.1 释放 MapReduce 服务 16](#_Toc23471)

[3.2 释放对象存储服务 OBS 16](#_Toc5065)

[3.3 释放弹性公网 IP 17](#_Toc11422)

[4 实验心得 17](#_Toc21288)

# 

# 一 实验环境准备

## 1.1购买 OBS 桶

鼠标放到控制台左上角三根横线处，弹出服务列表，点击“存储”分类里的“对象存储服务 OBS。在对象存储界面点击“创建桶”。 区域为“华北-北京四”，冗余策略为“多 AZ 存储”。输入桶名称“obs-a32e”。存储类别“标准存储”，桶策略“私有”，默认加密、归档数据直读“关闭”。



弹出对话框点击“确认”。



OBS 桶创建成功。



## 1.2购买 MRS 服务

在服务列表中点击“EI 企业智能”分类下的“MapReduce 服务”。 在现有集群界面点击“购买集群”。 选择“自定义购买”，区域选择“华北-北京四”。集群名称“mrs\_a32e”，版本“1.9.2”，类型为“分析集群”，组件默认勾选 Hadoop 。



计费模式“按需计费”，可用区、虚拟私有云、子网默认，安全组“自动创建”，弹性公网 IP “暂不绑定”。



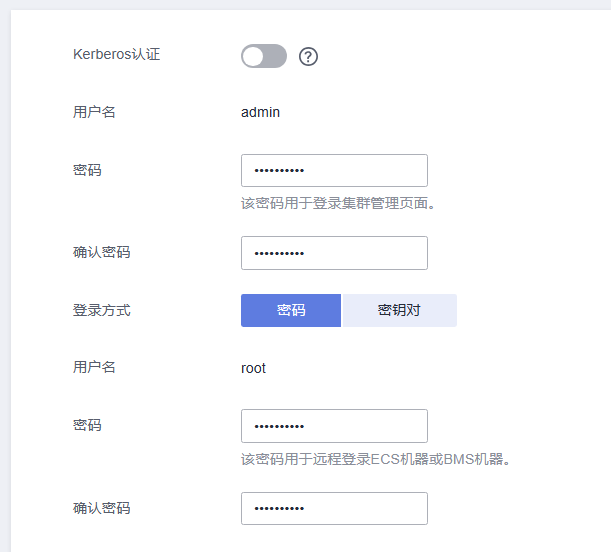
CPU 架构选择“鲲鹏计算”，集群节点默认。



标签、弹性伸缩、引导操作默认。 委托、数据盘加密默认，告警“关闭”。



关闭 Kerberos 认证（“Kerberos 认证”关闭时，普通用户可使用 MRS 集群的所有功能，建议单 用户场景下使用），输入密码（该密码用于登录集群管理页面）。登录方式“密码”，输入密码（该密码用于远程登录集群节点的 ECS 机器）



勾选“确认授权”。



集群状态变为“运行中”，创建成功。



## 1.3 购买弹性公网 IP

在服务列表中点击“网络”分类下的“弹性公网 IP EIP”。点击“购买弹性公网 IP”。选择“按需计费”，区域“华北-北京四”。 线路“全动态 BGP”，公网带宽“按流量计费”，带宽大小“50”。



点击“立即购买”，进入确认页面点击“提交”。回到弹性公网 IP 页面后点击刷新按钮，可以看到已经购买的弹性公网 IP。

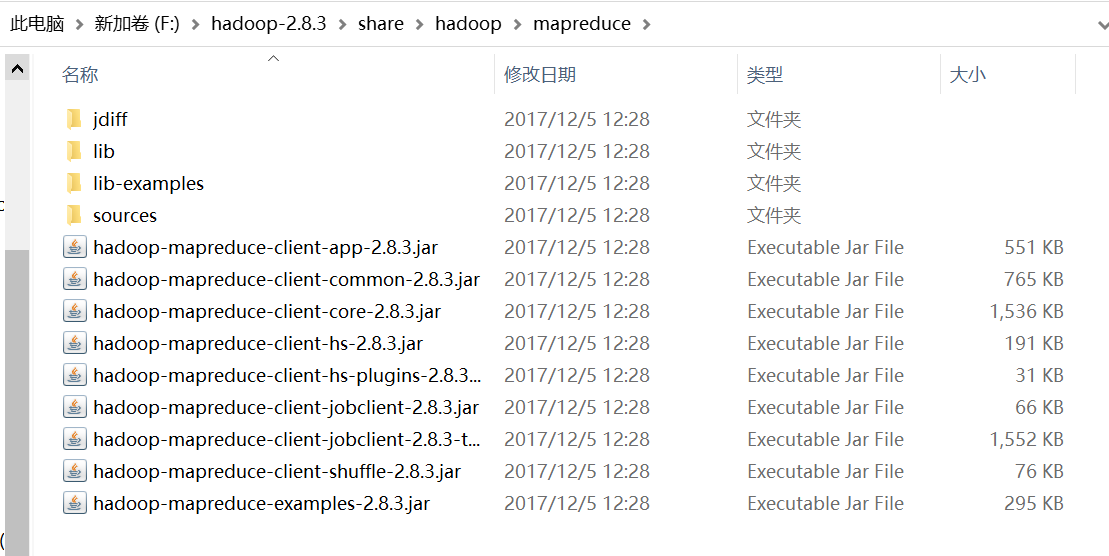


# 2 云服务实现大数据

## 2.1程序及数据准备

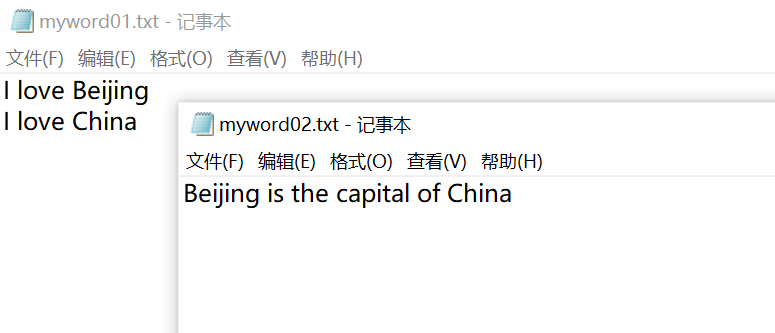
步骤 1 准备示例程序

下载 hadoop 2.8.3安装包



步骤 2 准备数据文件

准备两个文本文件



步骤 3 存储到 OBS

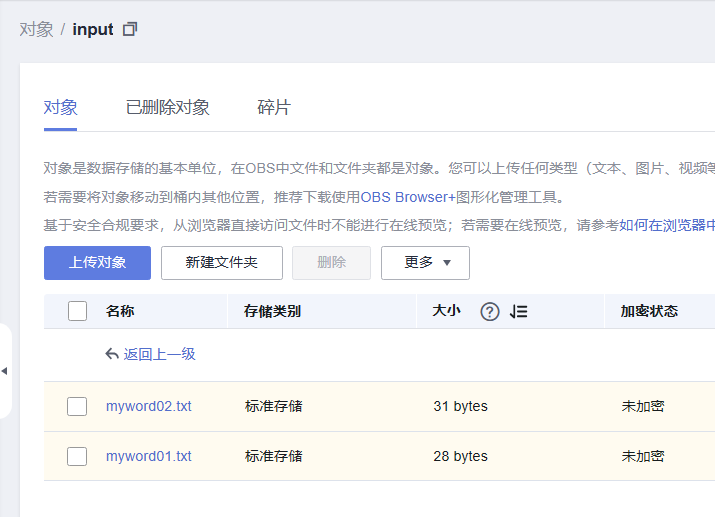
登录华为云控制台，进入到对象存储服务 OBS，在对象存储界面点击前面创建的桶名称。点击“对象”，然后点击“新建文件夹”。输入文件夹名“program”并点击确定，同理再创建 input 文件夹



存入程序和数据文件，，示例程序存储到 OBS 桶的 program 目录中



数据文件上传到 input 目录中

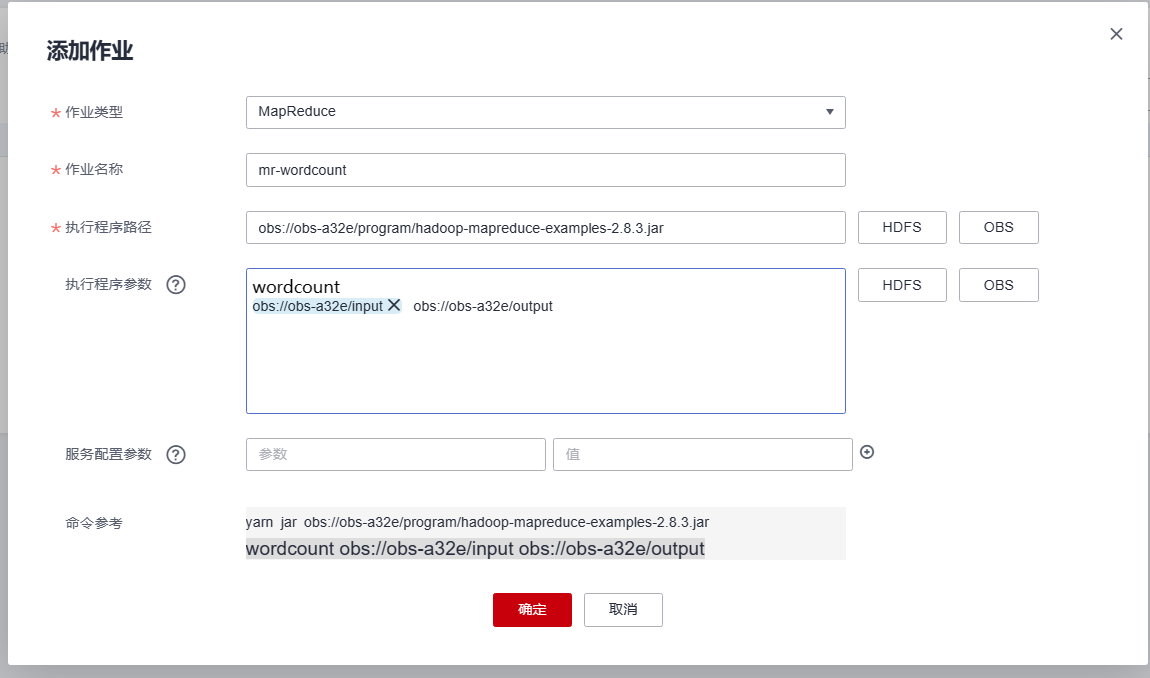


## 2.2执行程序

步骤 1 创建并运行作业

进入控制台 MRS 现有集群页面，点击前面创建的集群名称，选择“作业管理”标签，单击“添加”按钮，作业类型选择“MapReduce”，输入作业名称“mr-wordcount” ，点击程序执行路径后面的“OBS”按钮，在弹出对话框中点击桶名称、文件夹名称、文件 名，直到选择到 hadoop 样例程序，点击“是” 执行程序参数配置为“wordcount obs://obs-a32e/input/ obs://obs-a32e/output”（其

中 input 为输入目录，由后面的 OBS 按钮选择，output 为输出路径，手动输入一个不存在的 目录即可），服务配置参数不配置，点击“确定”



等待作业状态变为“已完成”则作业执行完毕



步骤 2 查看程序输出结果

下载结果文件，进入 OBS 控制台，点击桶名称 obs-a32e 进入桶中可以看到已经生成的 output 输出目录

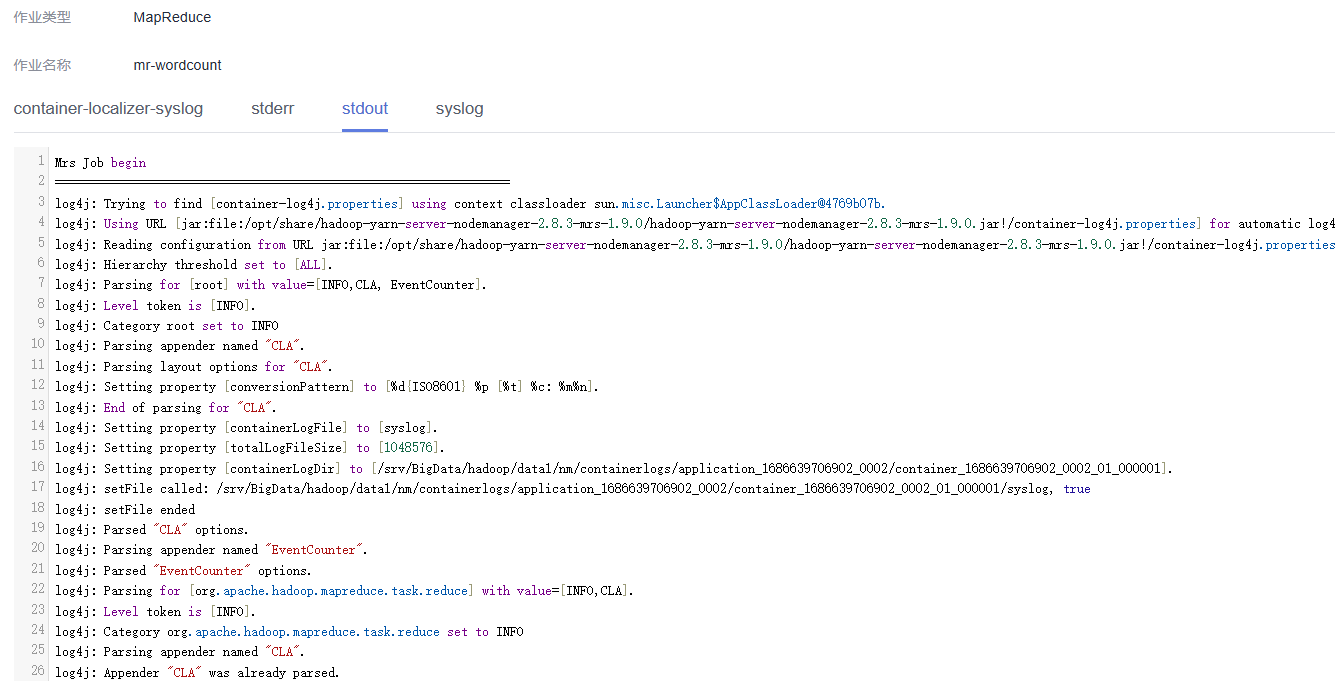


点击进入 output 目录中可以看到输出文件，点击结果文件“part-r-00000”后面的“下载”，把文件下载到本地,查看结果内容



## 2.3查看作业日志

在集群的“作业管理”界面，点击对应作业后面的“查看日志”



## 2.4使用组件管理

步骤 1 同步 IAM 账户

进入 mrs\_test 集群页面，选择“概览”标签，点击“IAM 用户同步”后面的“点击同步”,稍后刷新页面，“组件管理”等标签变的可用



步骤 2 组件管理

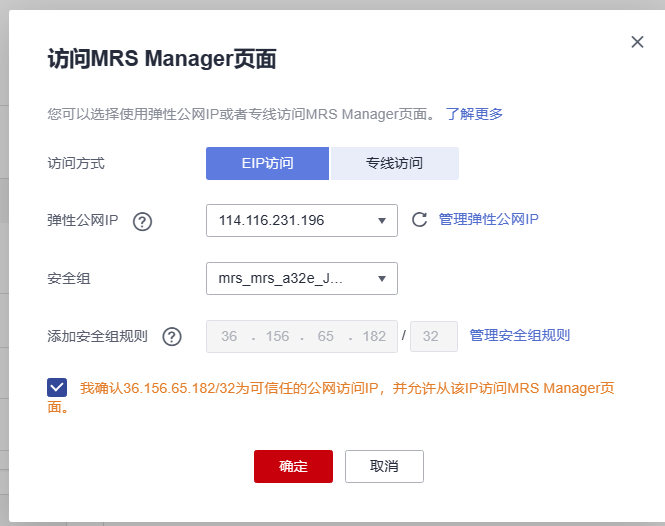
在集群页面点击“组件管理”，可以看到正在运行的组件服务，点击“MapReduce”



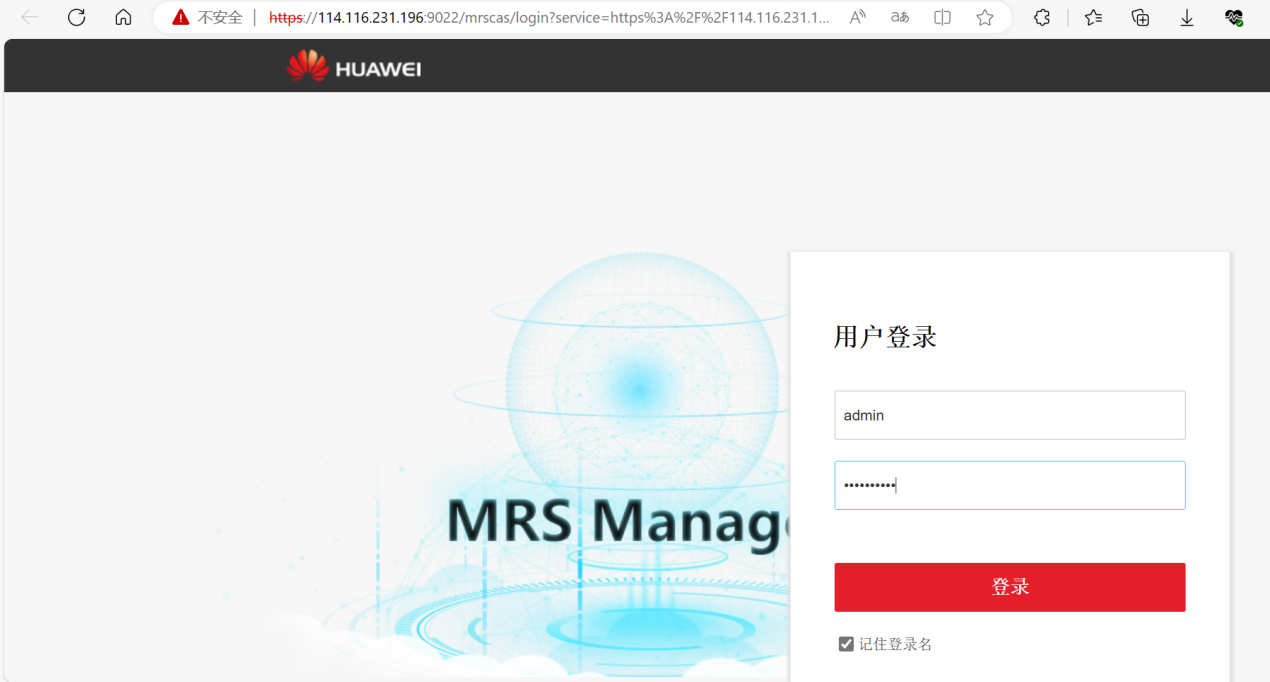
在 MapReduce 页面可以对组件进行相关管理，点击“JobHistoryServer”



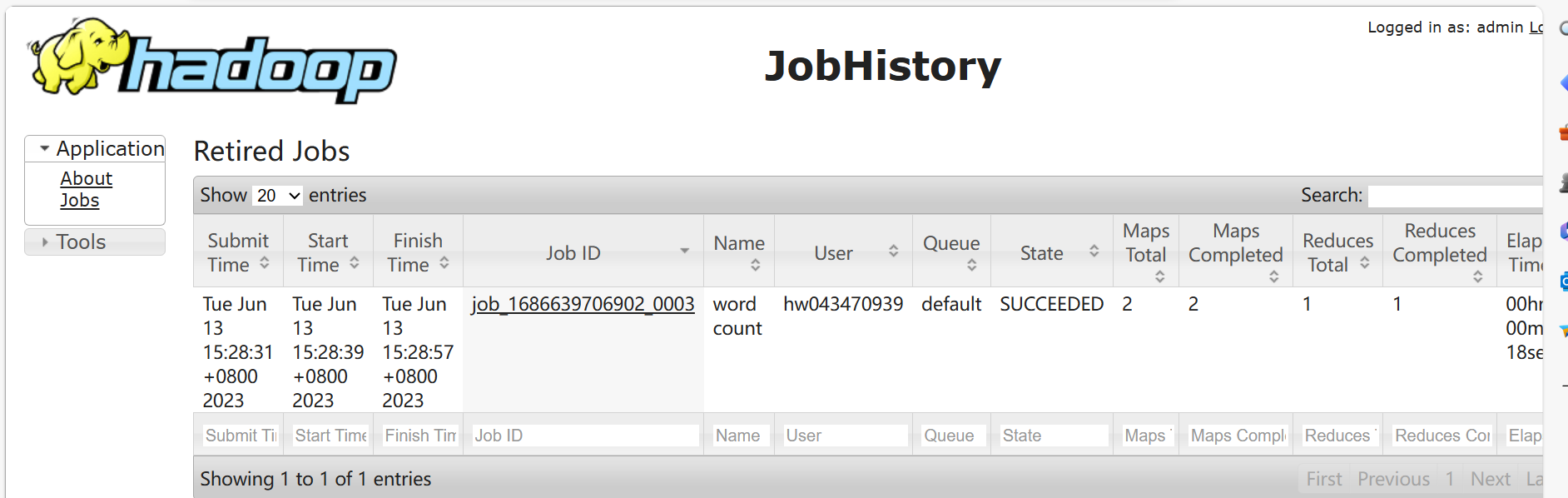
弹出访问 MRS 管理页面对话框，选择前面购买的弹性公网 IP，其他默认，点击“确定”



进入登录页面，输入 MRS 集群管理的用户名和密码



登录成功后进到 MapReduce 的 Web 管理界面，可以看到刚执行的 MapReduce 任务，点 击前面执行的 Job ID



可以看到 Job 的具体执行信息



## 2.5思考题

我们的现在使用的是 MapReduce 作业，如何使用 Spark 作业呢？

Spark使用不同的编程模型和API，因此需要重新设计和实现作业逻辑；在使用Spark之前，还需要配置和启动Spark集群；Spark支持多种数据源，需要修改作业的输入输出部分，以适应Spark的数据源接口；Spark支持多种编程语言，需要选择适合的编程语言，并进行相应的编译和打包操作；最后使用Spark提供的命令行工具或API，提交和运行Spark作业。

# 3 资源清理释放

## 3.1 释放 MapReduce 服务





## 3.2 释放对象存储服务 OBS





## 3.3 释放弹性公网 IP



# 4 实验心得

通过此次实验，深入学习了云服务及其实践，并通过实际操作来了解了云服务的一般使用流程。通过搭建基于云服务的大数据环境，我掌握了MRS的作业管理和组件管理的方法。云服务提供了灵活、可扩展且高性能的计算和存储资源，为大数据处理提供了强大的支持。在搭建大数据环境的过程中，我对云服务的配置和部署有了更深入的理解，也掌握了一些基本的管理技巧。