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 实验二：云服务实现大数据

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

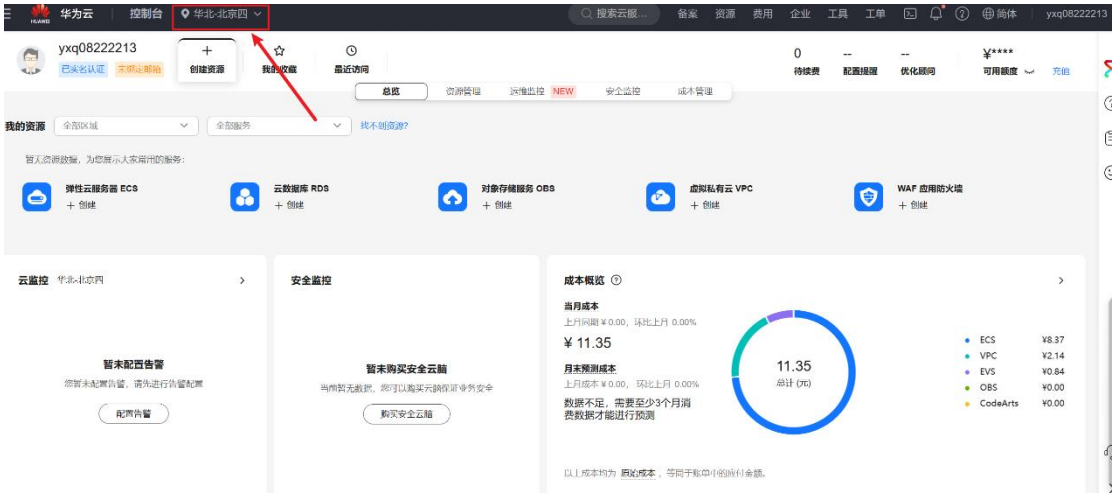
实验二：云服务实现大数据

1 实验环境准备

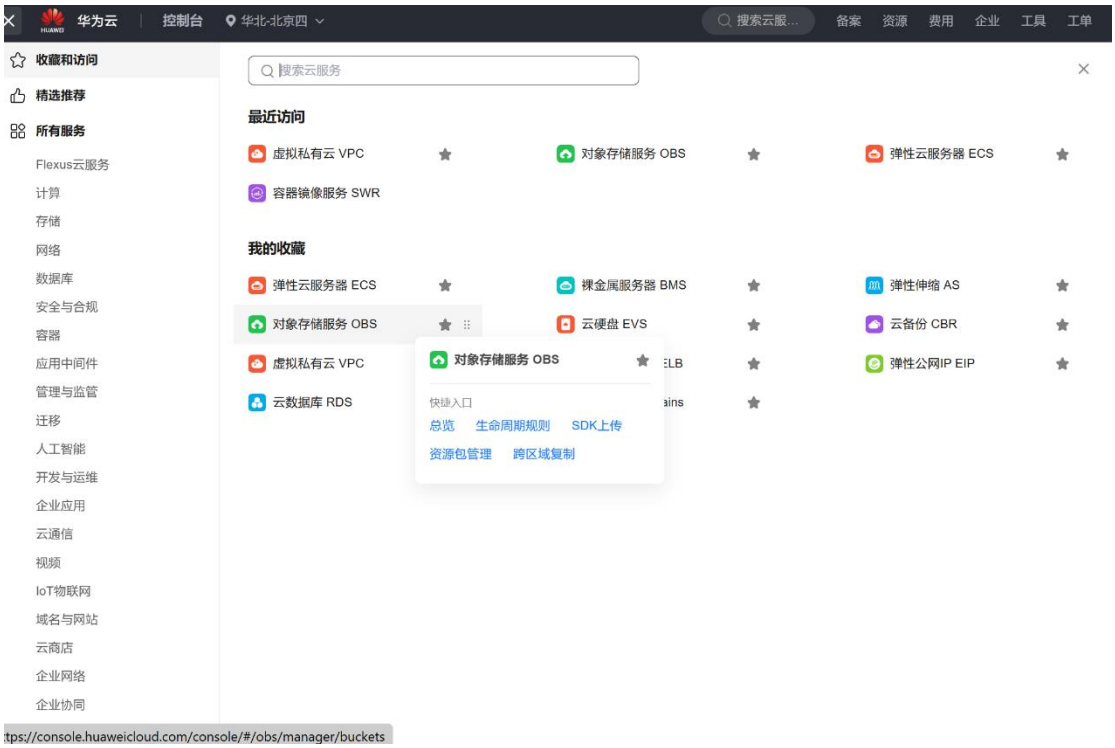
1.1 服务购买

1.1.1 登陆控制台

打开华为云官网首页（https://www.huaweicloud.com/），点击“登录”按钮后输入账号信息进行登录，选择区域为“北京四”



1.1.2 购买 OBS 桶

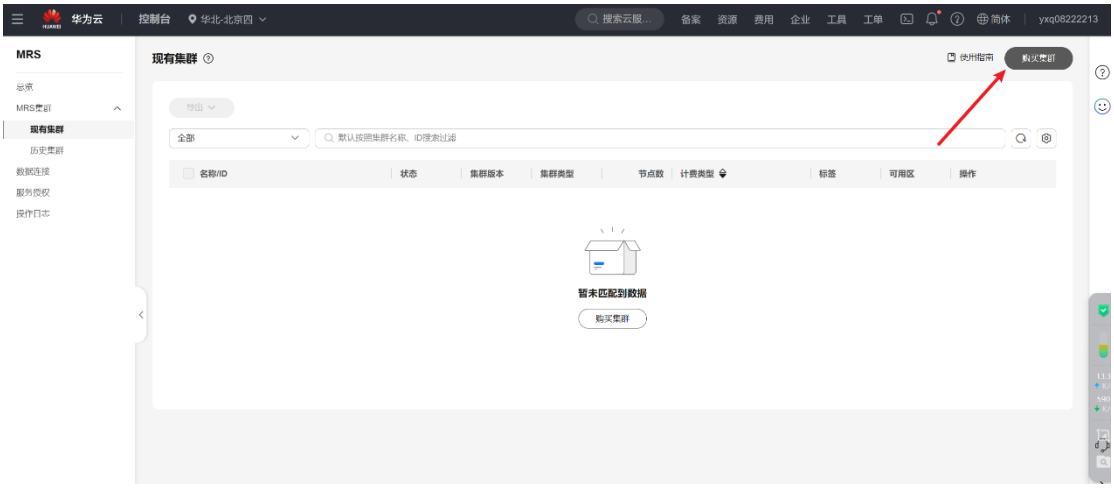


区域为“华北-北京四”，冗余策略为“多 AZ 存储”，按照规则输入桶名称，存储类别“标准存储”，桶策略“私有”，默认加密、归档数据直读“关闭”

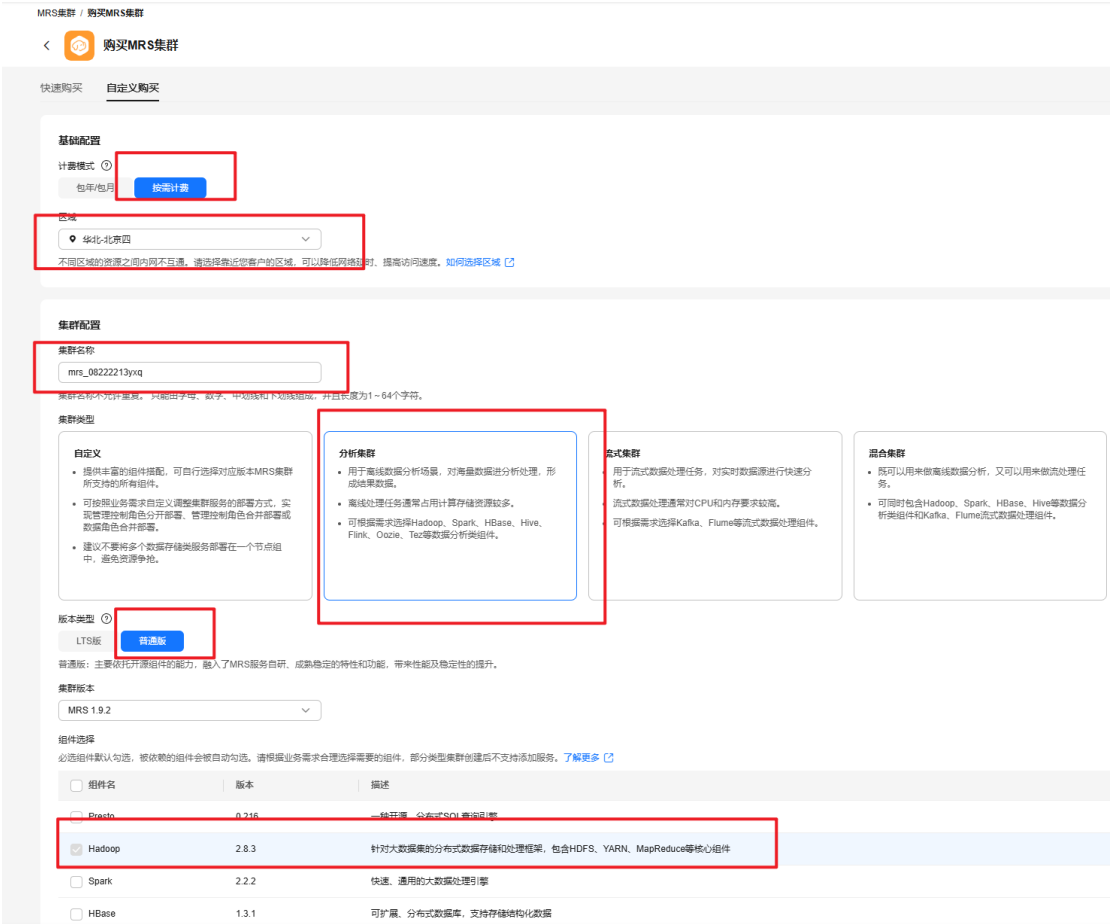
OBS 桶创建成功

1.1.3 购买 MRS 服务

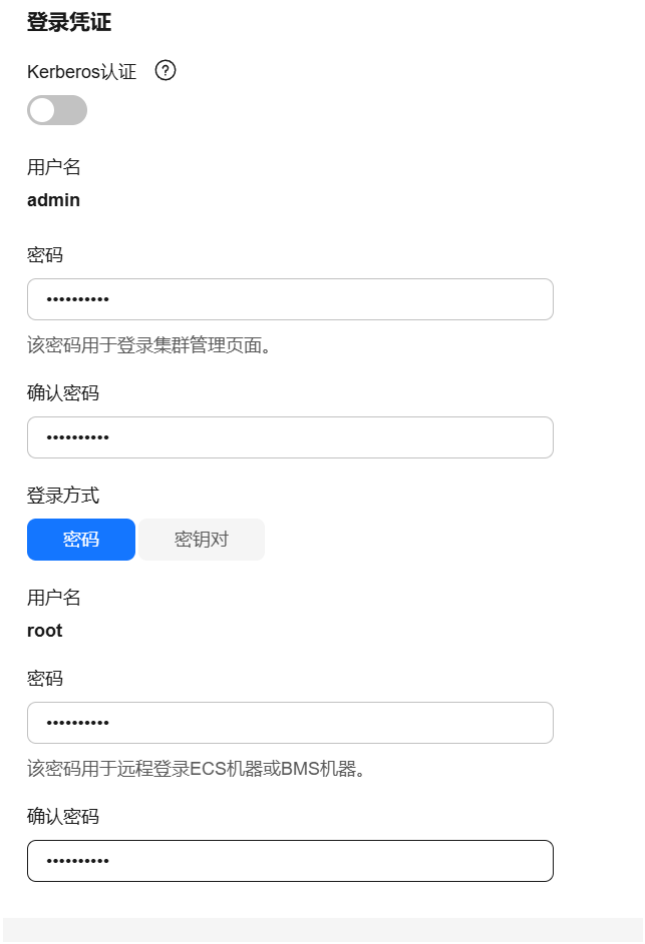
在现有集群界面点击“购买集群”



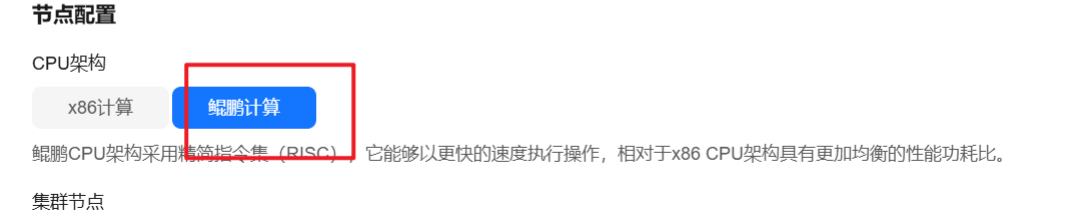
选择“自定义购买”，区域选择“华北-北京四”，版本“1.9.2”，类型为“分析集群”，组件默认勾选 Hadoop



关闭 Kerberos 认证，输入密码



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

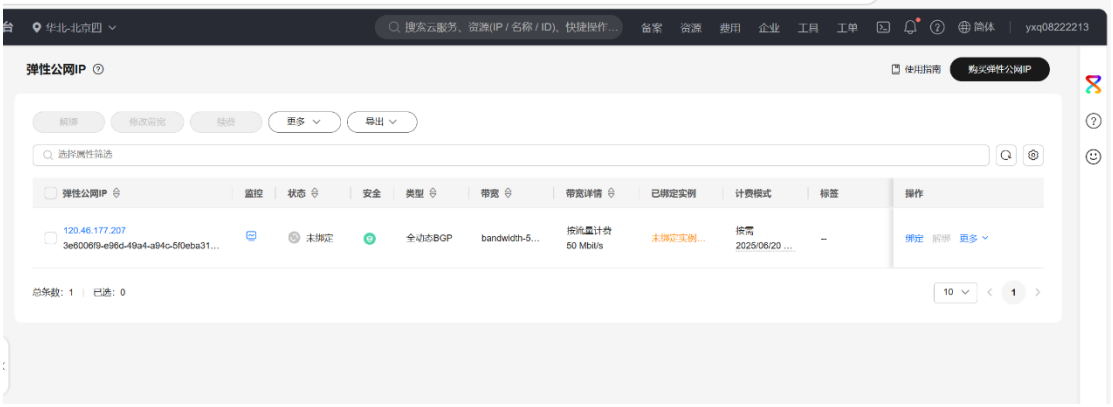


集群状态变为“运行中”，创建成功

1.1.4 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”

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

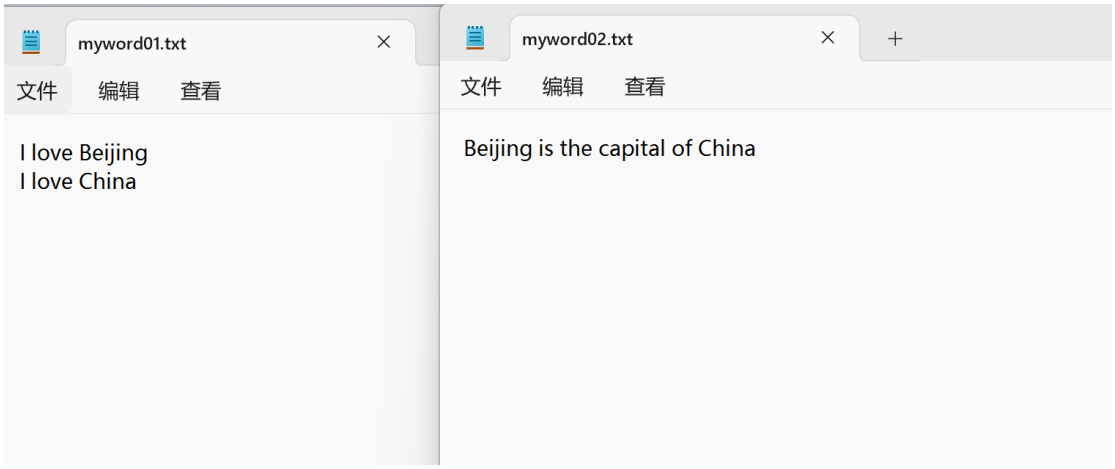


2 云服务实现大数据

2. 1 试验任务

2.1.1 程序及数据准备

步骤 1 准备示例程序 hadoop-mapreduce-examples-2.8.3.jar
步骤 2 准备数据文件
准备两个文本文件，输入如下内容



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；
点击文件夹 program 名称进入，然后点击“上传对象”，点击“添加文件”；
选择前面准备好的示例程序文件 hadoop-mapreduce-examples-3.1.3.jar，点击“打开”

上传成功，示例程序存储到 OBS 桶的 program 目录中

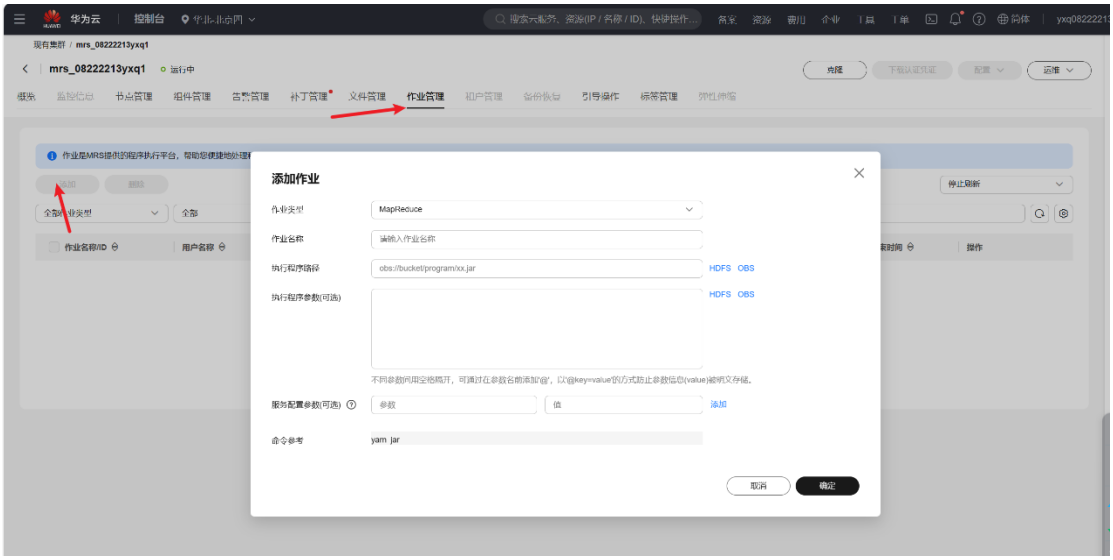
把前面准备好的数据文件上传到 input 目录中

2.1.2 执行程序

步骤 1 创建并运行作业

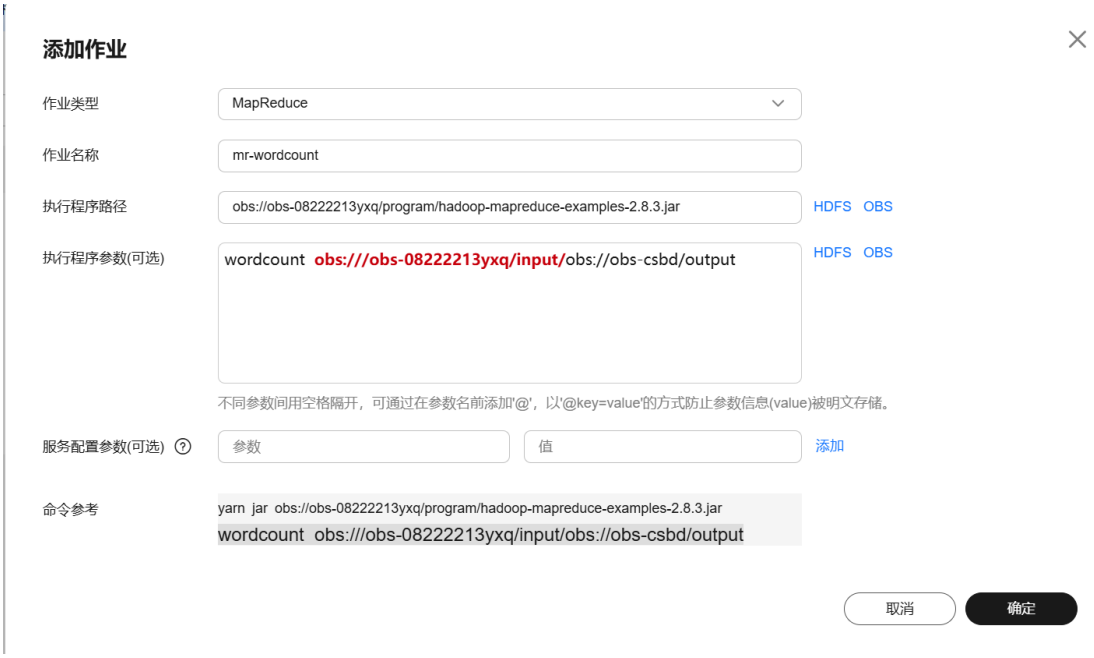
进入控制台 MRS 现有集群页面，点击前面创建的集群名称；

选择“作业管理”标签，单击“添加”按钮



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-08222213yxq/input/ obs://obs-008222213yxqoutput”



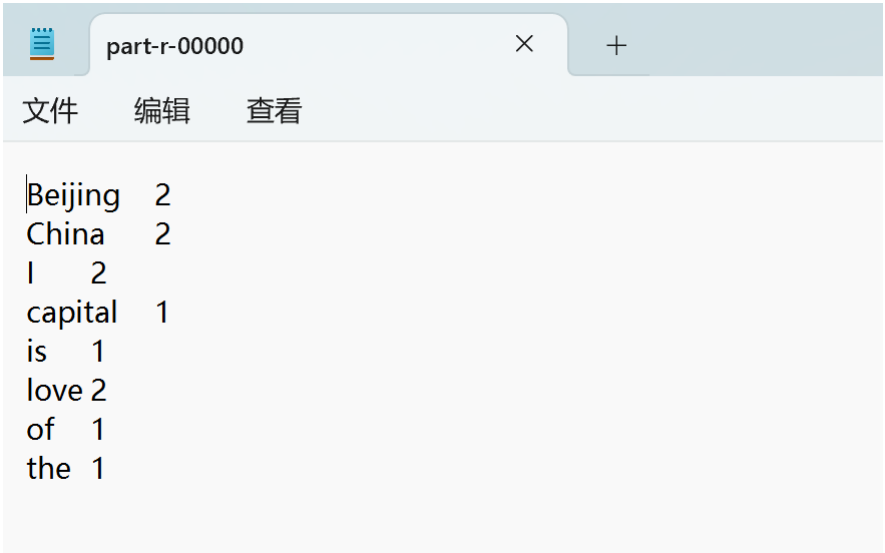
作业创建完成，状态开始为“已接受”

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

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

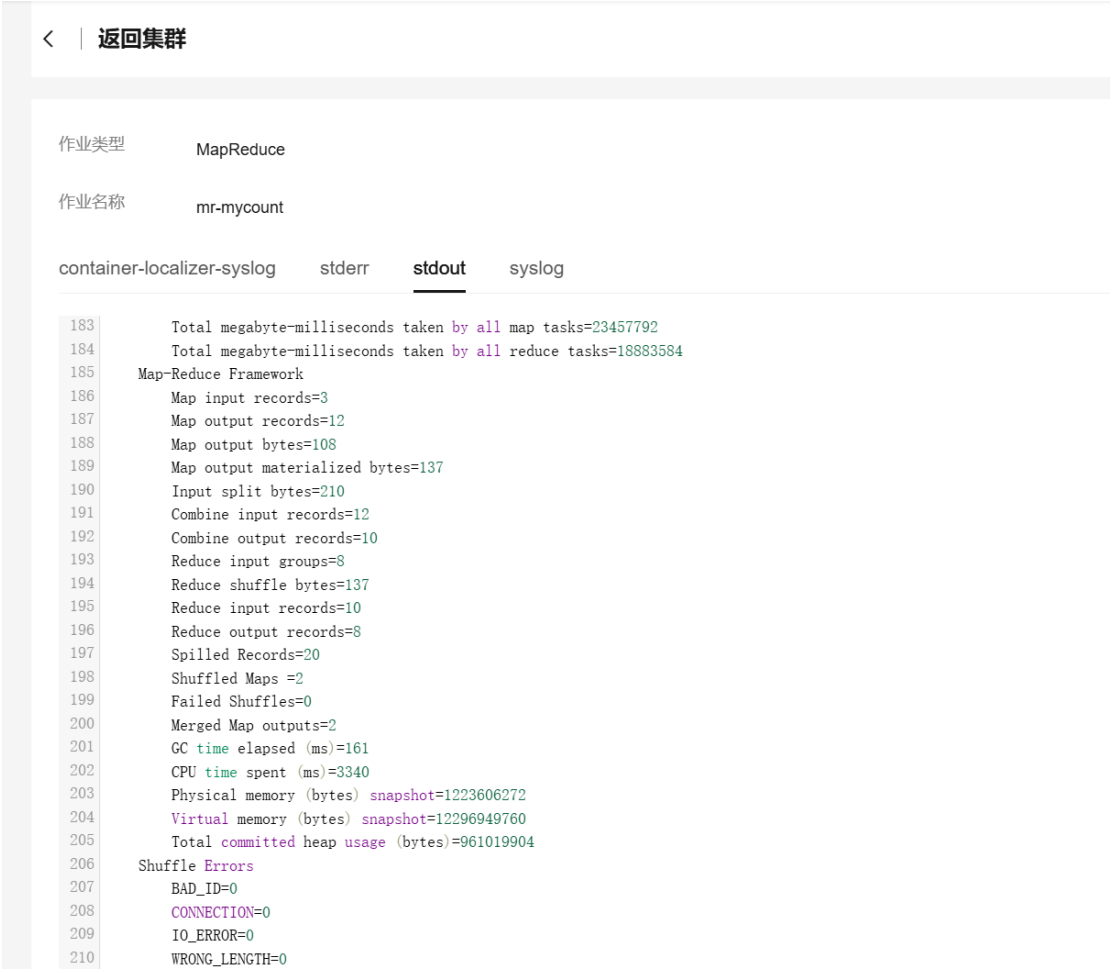
查看结果内容



2. 2 结果验证

2.2.1 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”



2.2.2 使用组件管理

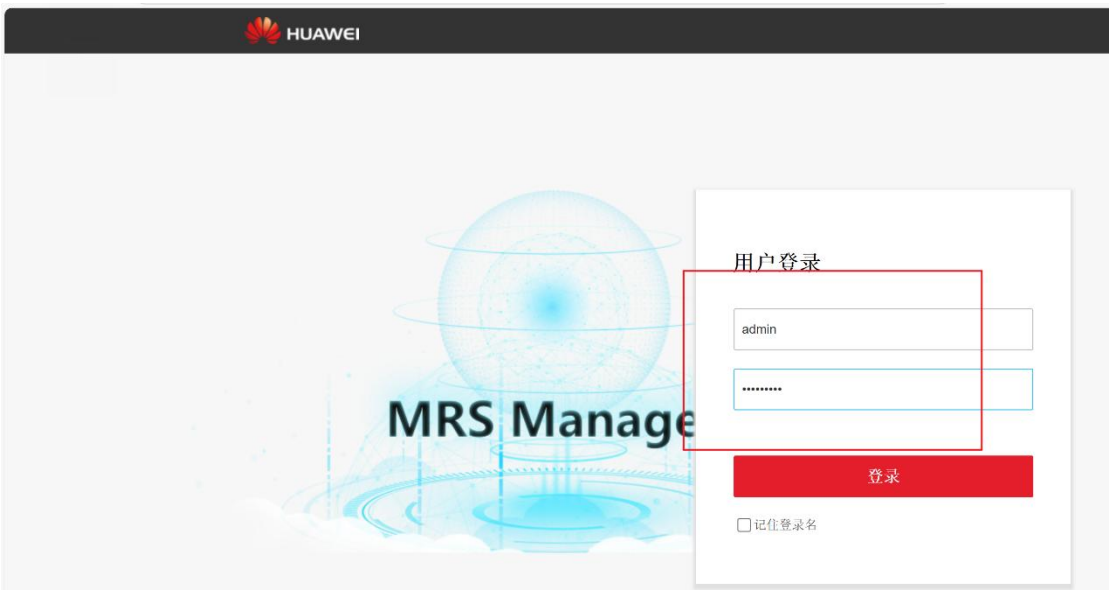
步骤 1 同步 IAM 账户

步骤 2 组件管理

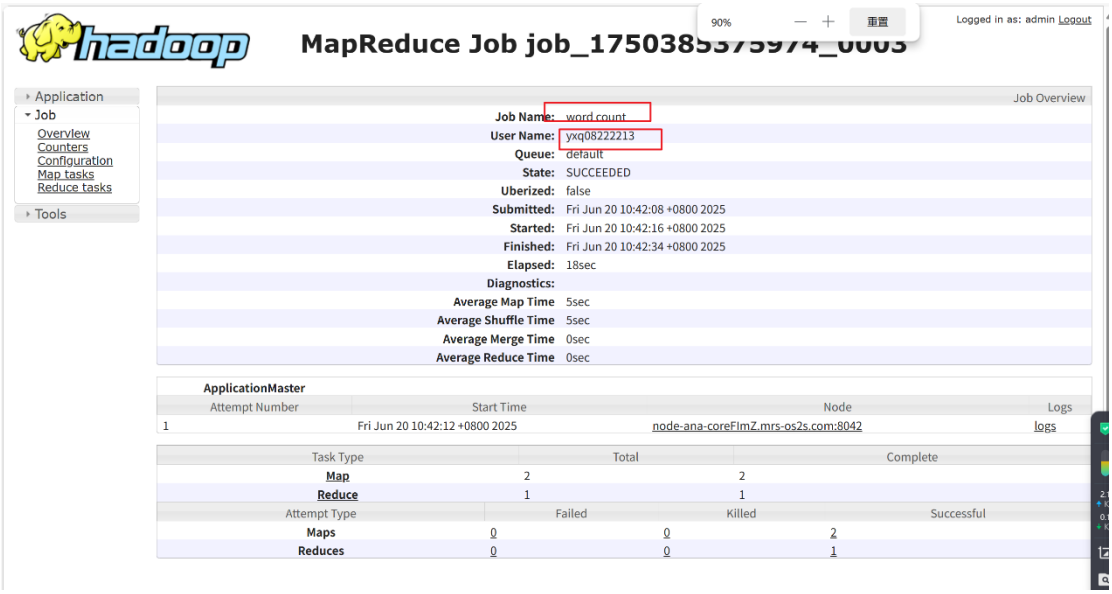
在 MapReduce 页面可以对组件进行相关管理，点击“JobHistoryServer”

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



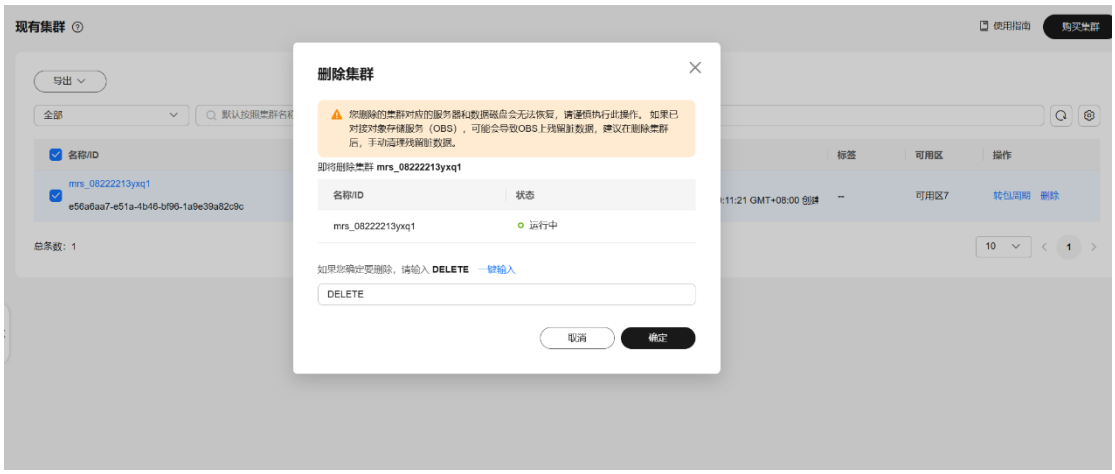
登录成功后进到 MapReduce 的 Web 管理界面，可以看到刚执行的 MapReduce 任务，点击前面执行的 Job ID，可以看到 Job 的具体执行信息



3 资源清理释放

3.1 释放 MapReduce 服务

在“现有集群”中可以看到购买的集群，点击后面的“删除”链接进行删除



3.2 释放对象存储服务 OBS

3.3 释放弹性公网 IP