



南京大學

NANJING UNIVERSITY

## 课程作业报告

大数据分析 02

---

学生姓名：左皓升

学生学号：211250074

2023 年 9 月

# 大数据分析云平台环境配置与测试

使用阿里云 EMR on ECS 搭建云平台环境

## 1. 编写 hello world 程序

首先在本地搭建 scala 环境 (本地有 java 环境, 过程较为简单), 使用 IDEA 创建 scala 项目并编写 hello world 程序。

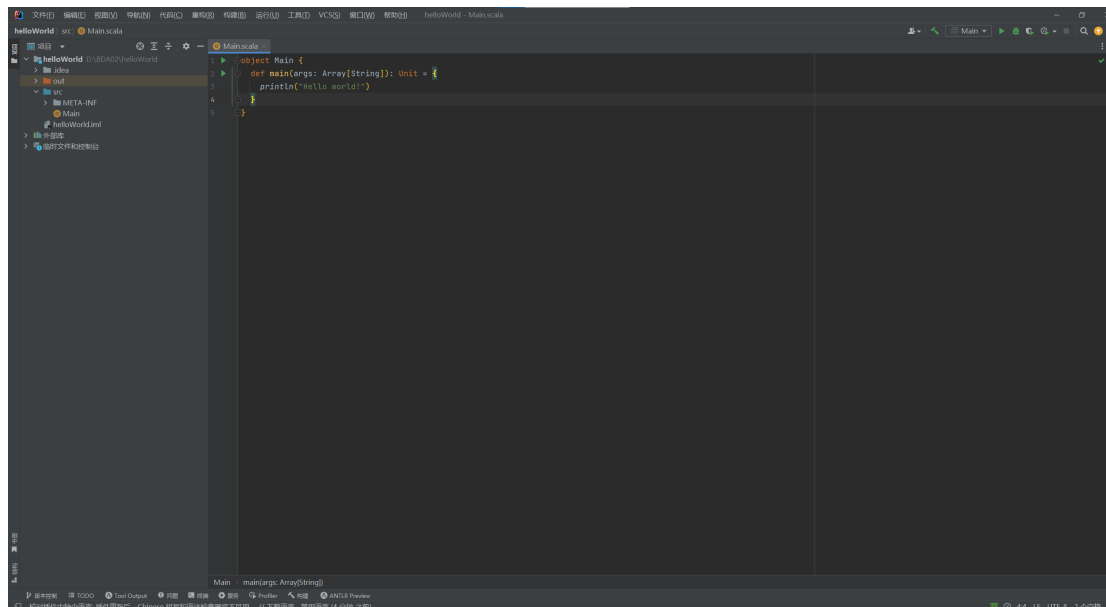


图 1: hello world 程序

## 2. 打 jar 包

### 2.1 使用 IDEA 将项目打成 jar 包

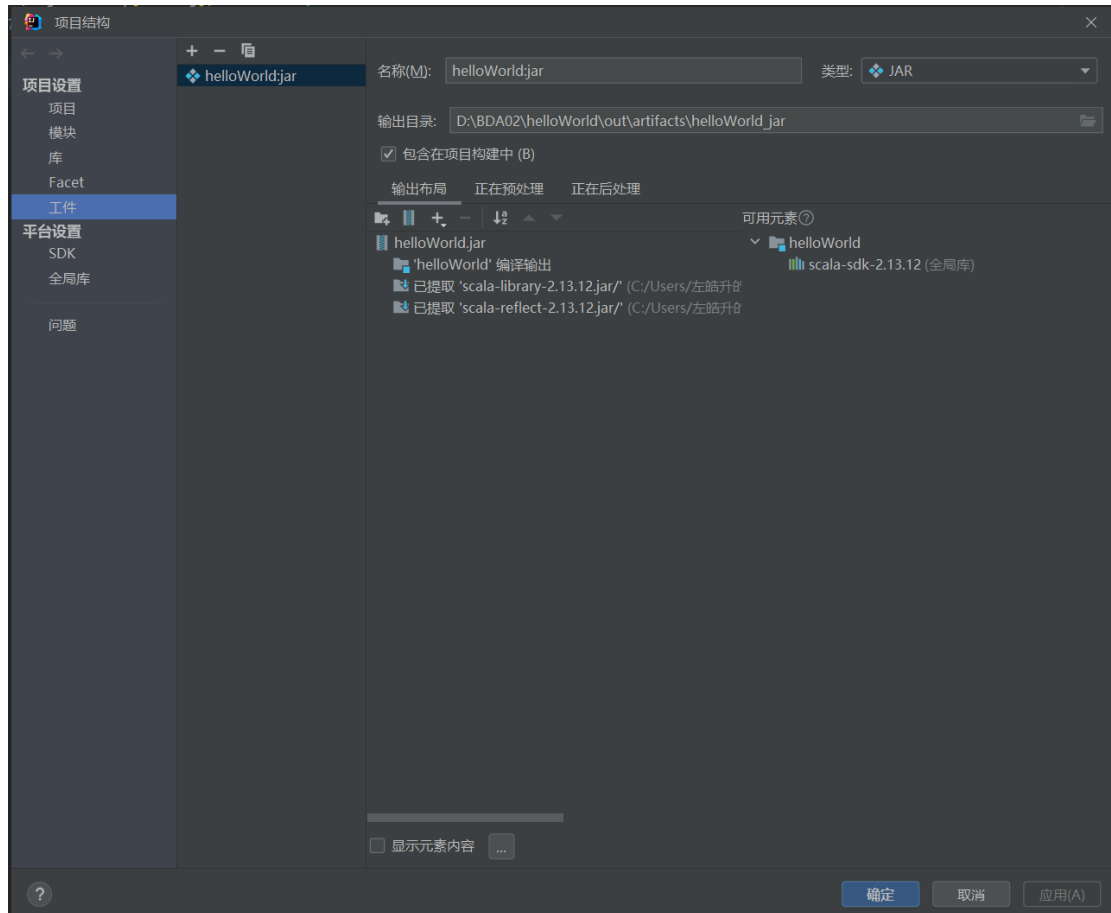


图 2: 使用 IDEA 打 jar 包

## 2.2 本地测试运行 jar 包

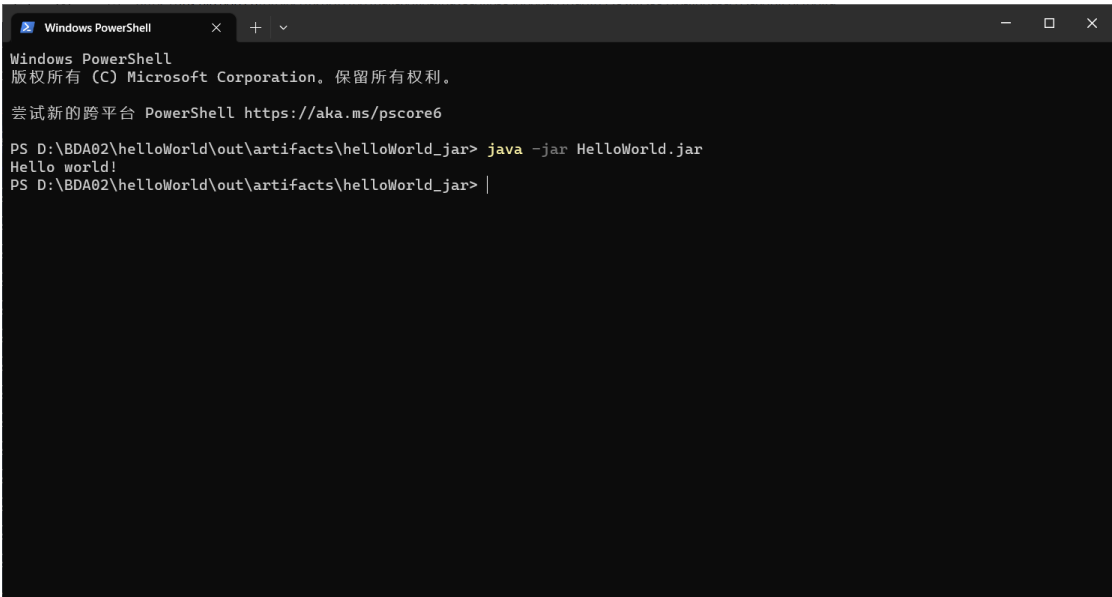


图 3: 本地运行 jar 包

## 3. 使用阿里云搭建 EMR on ECS 服务器

### 3.1 配置搭建过程

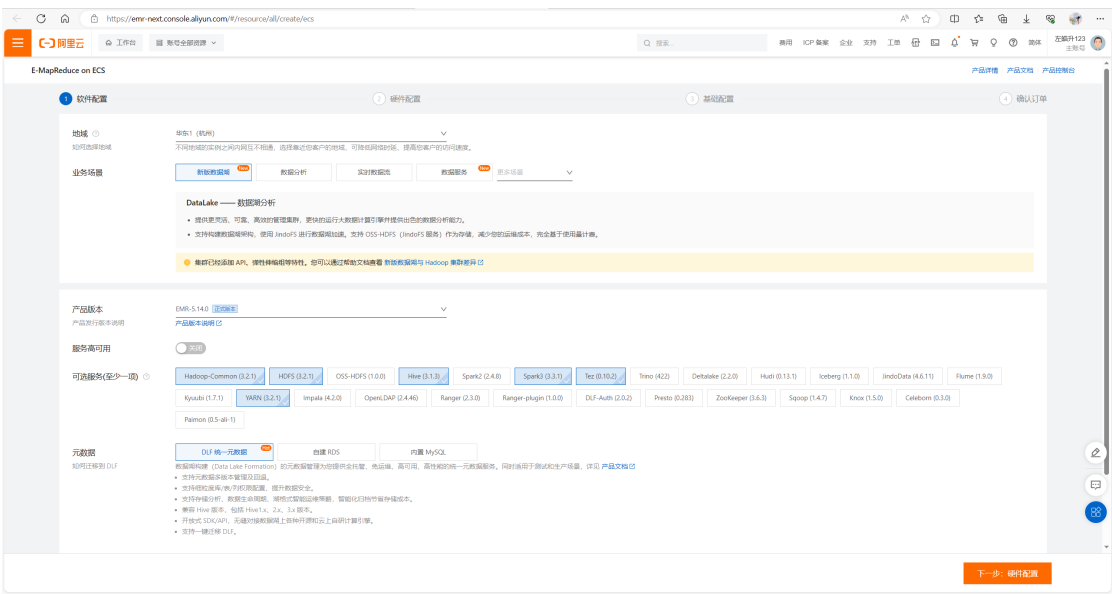


图 4: 选择需要的服务

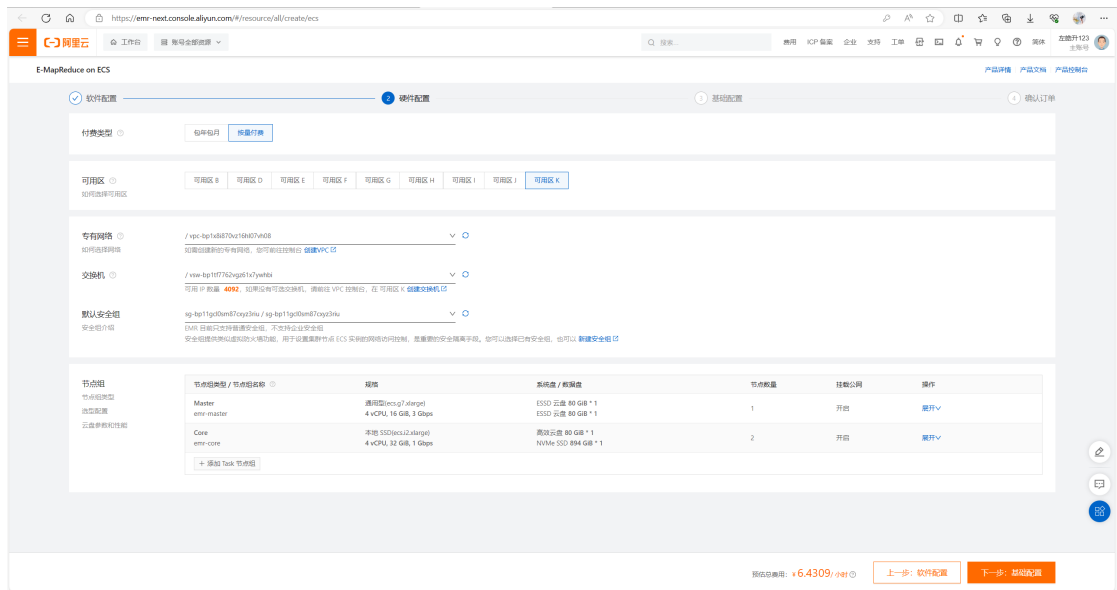


图 5: 配置交换机，设置公网 IP 便于 SSH 访问

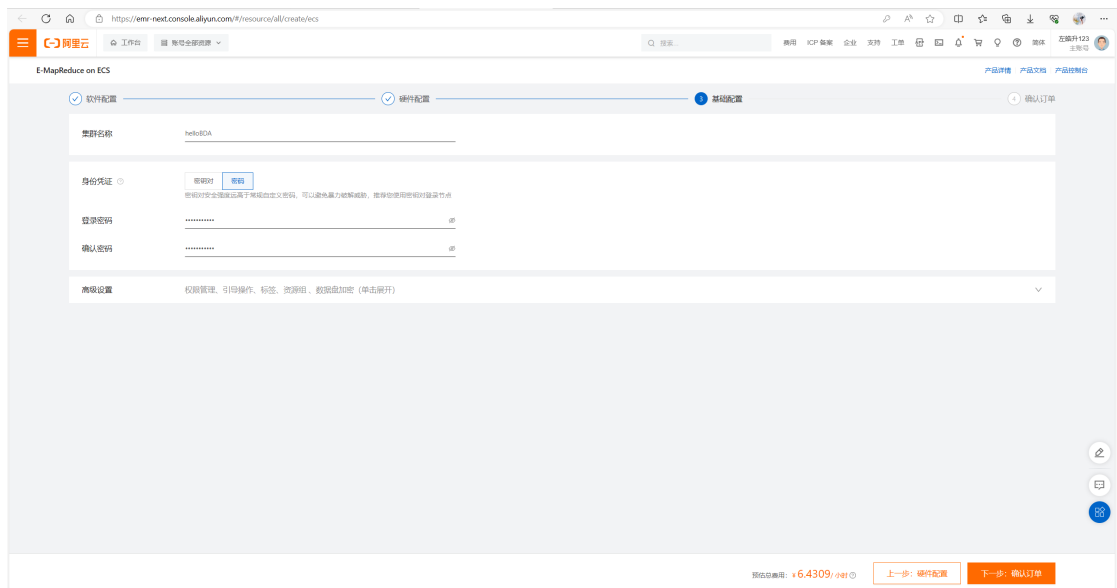


图 6: 设置密码

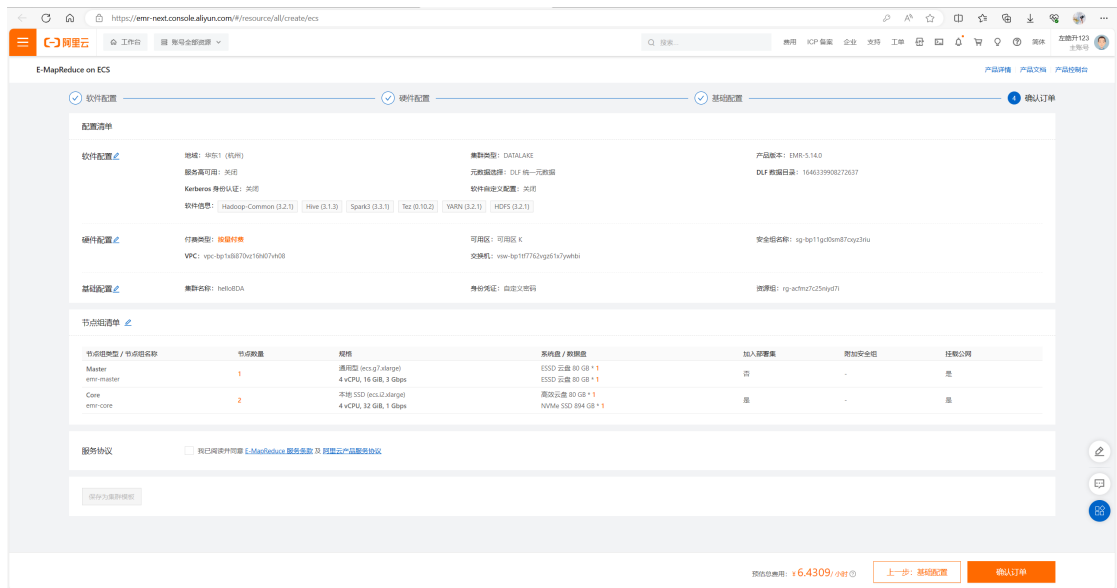


图 7: 服务器配置

## 3.2 进入工作台，可以看到服务器已经搭建成功，在实例中也可以看到

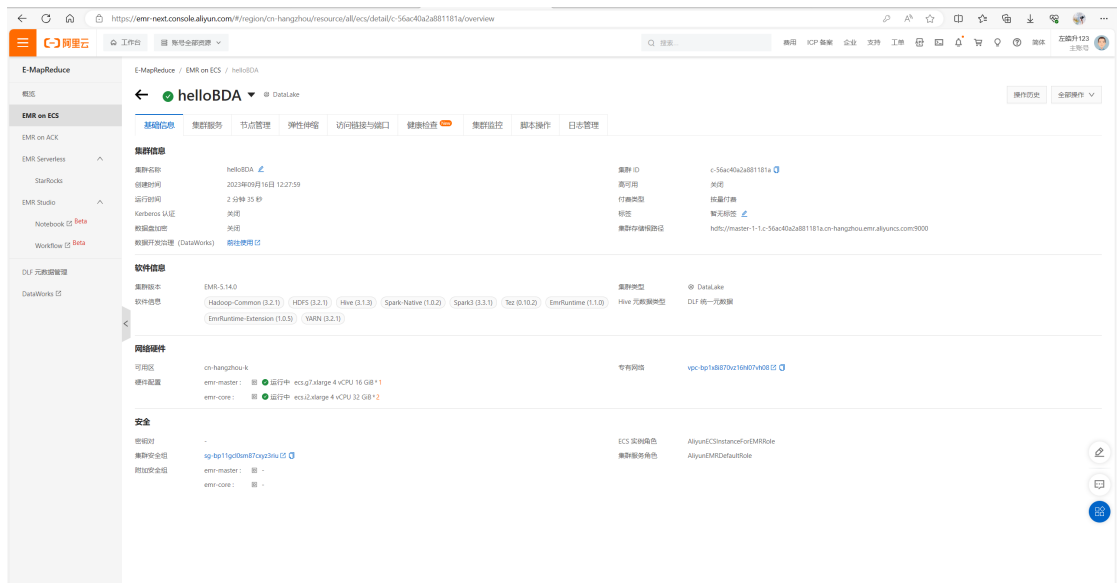
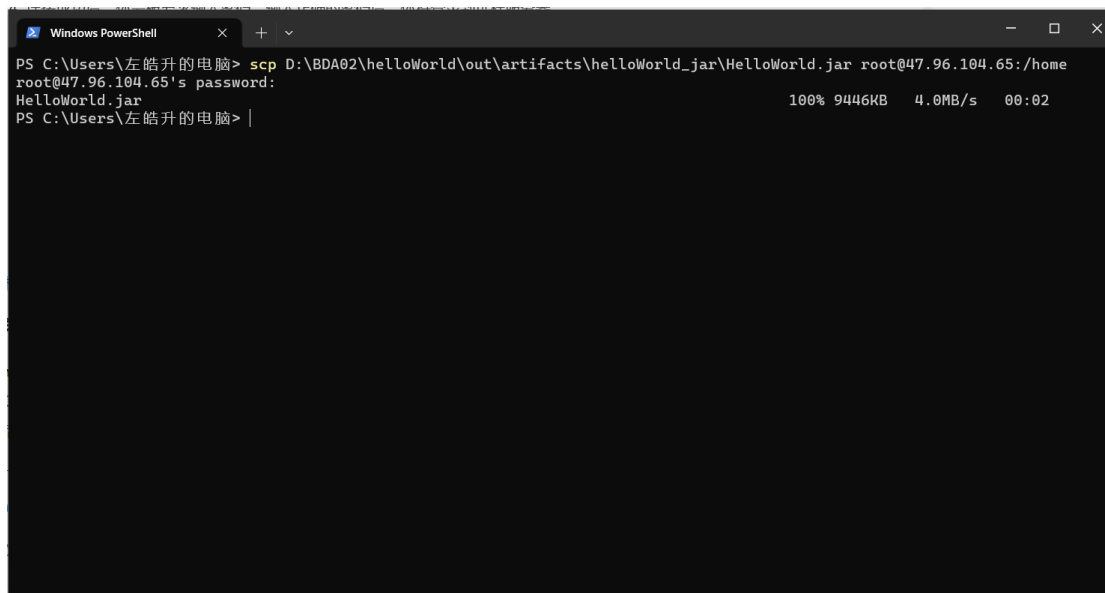


图 8: 工作台



## 5. 运行程序

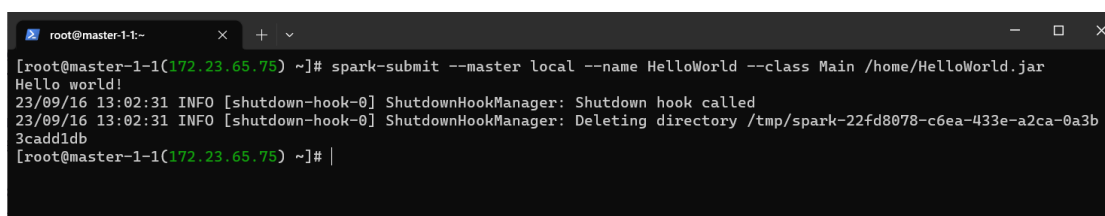
### 5.1 使用 scp 将本地的 jar 包上传



```
Windows PowerShell
PS C:\Users\左皓升的电脑> scp D:\BDA02\helloWorld\out\artifacts\helloWorld_jar\HelloWorld.jar root@47.96.104.65:/home
root@47.96.104.65's password:
HelloWorld.jar                                100% 9446KB   4.0MB/s   00:02
PS C:\Users\左皓升的电脑> |
```

图 11: 实例

### 5.2 运行 jar 包



```
root@master-1-1:~
[root@master-1-1(172.23.65.75) ~]# spark-submit --master local --name HelloWorld --class Main /home/HelloWorld.jar
Hello world!
23/09/16 13:02:31 INFO [shutdown-hook-0] ShutdownHookManager: Shutdown hook called
23/09/16 13:02:31 INFO [shutdown-hook-0] ShutdownHookManager: Deleting directory /tmp/spark-22fd8078-c6ea-433e-a2ca-0a3b3cadd1db
[root@master-1-1(172.23.65.75) ~]# |
```

图 12: 实例

成功输出了 Hello, World!

## 6. 挑战以及克服

### 6.1 选用什么样的云服务

一开始不知道应该使用什么样的云服务，尝试了微软的 Azure，阿里云 ECS，最终结合自己对作业要求的理解以及作业中对 amazon EMR 的推荐，选择了阿里云的 EMR。



## 6.2 如何配置 EMR

以前没有接触过大数据和云服务器，在配置过程中不知道自己需要什么样的服务，对于配置中给出的各种名词也不了解，不知道应该做出什么样的选择，交换机该如何配置。在配置过程中，遇到不认识的就去 bing 上搜索，去了解，大概知道这是什么东西，起着什么样的作用。

## 6.3 如何在 EMR 上运行代码

一开始不知道如何在服务器上运行代码，直接把代码传上去却发现没法运行。后来去搜索才知道，运行 jar 包是一个更好的选择。于是使用 IDEA 先打好 jar 包再传上去运行。