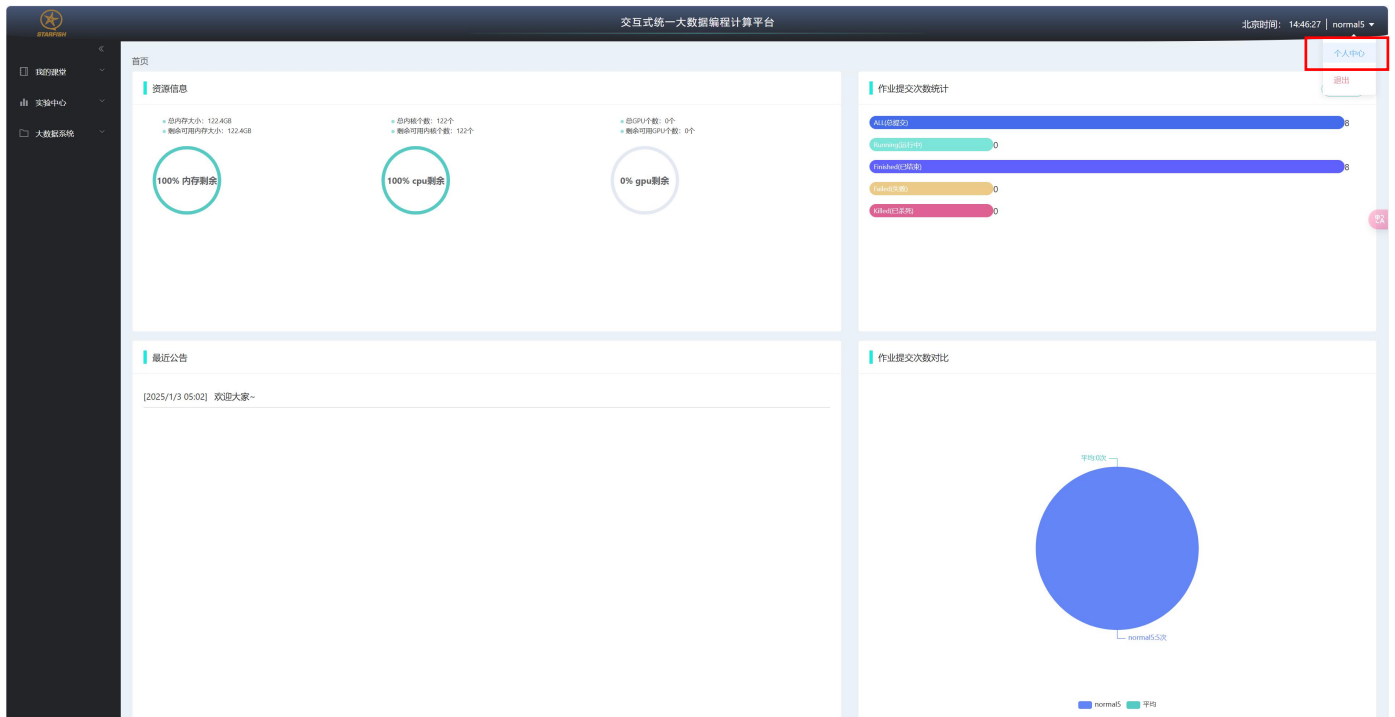


大数据平台使用手册

一. 登录

平台地址为 <http://114.212.130.202:8082/>，账号为：学生学号+stu，初始密码为：123456

登录后在个人中心修改密码



个人中心

基本信息

● 当前在线

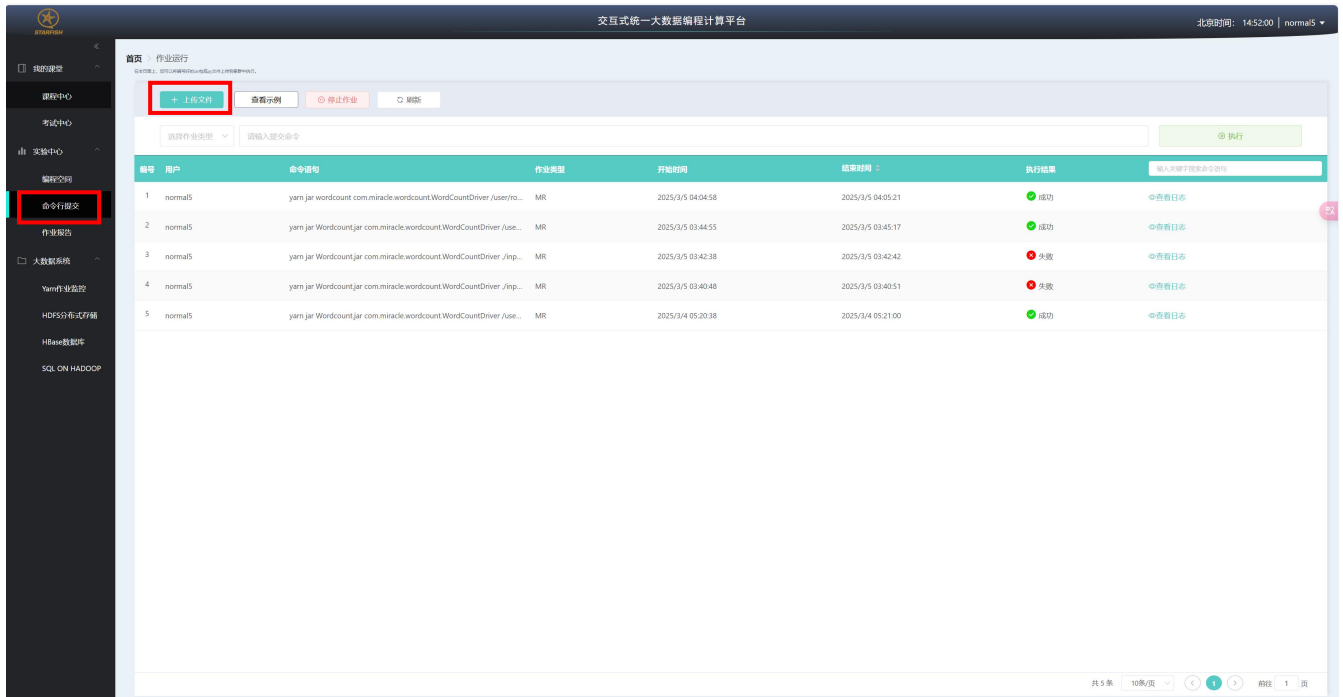
- 用户名称: normal5
- 用户角色: 学生
- 登录次数: 15
- 最后登录: 2025-03-06 14:45:31

修改密码

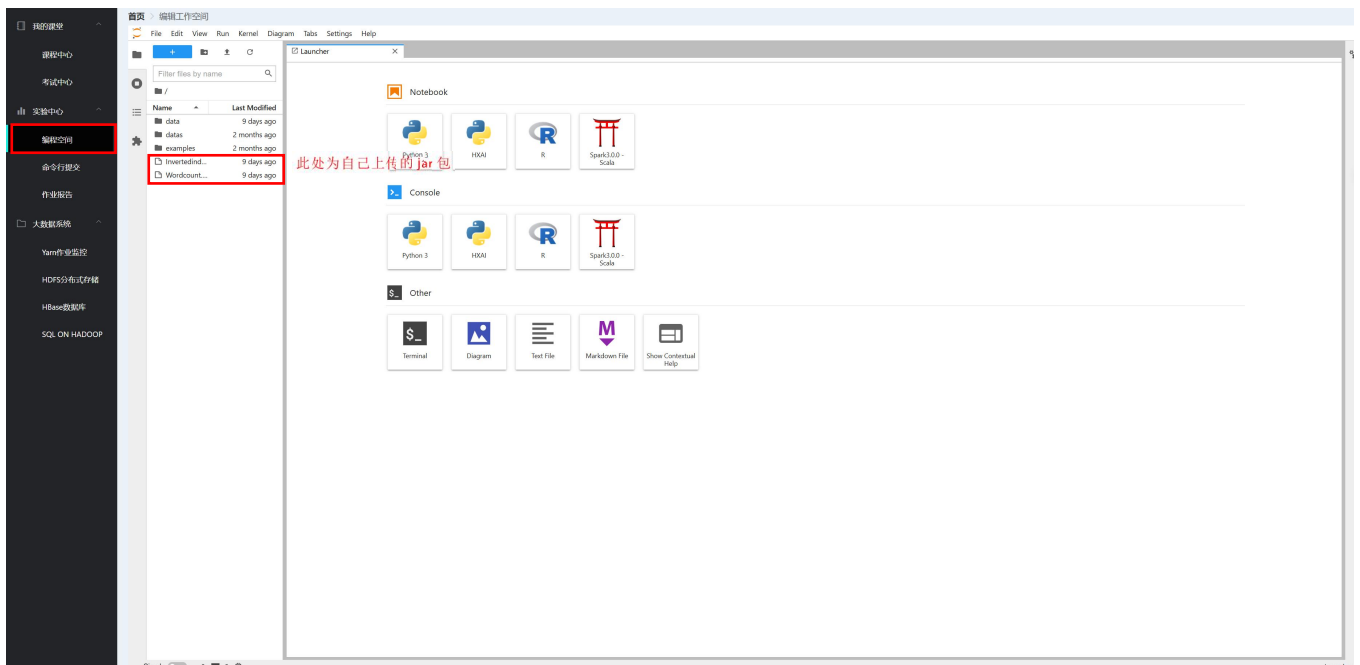
二. 功能

1. 上传 jar 包

实验中心的命令行提交中点击上传文件，上传自己的 jar 包

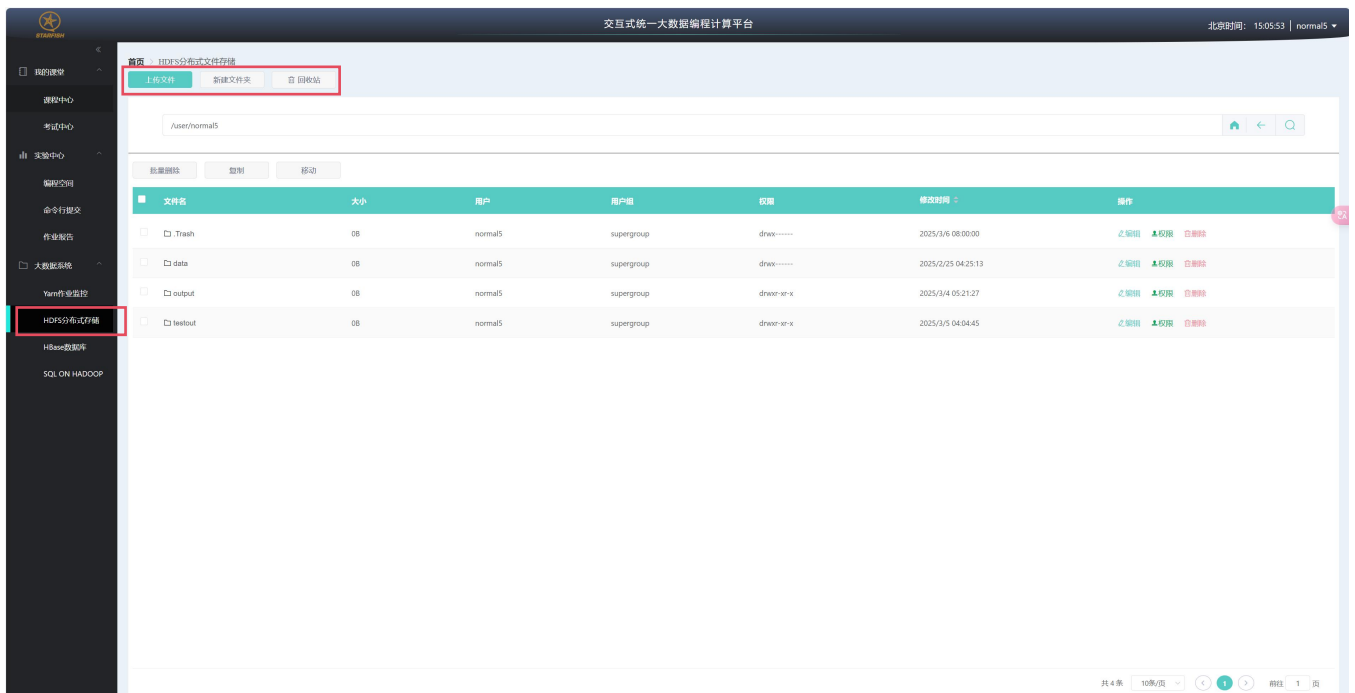


上传后的 jar 包可以在编程空间中看到



2. 上传数据文件

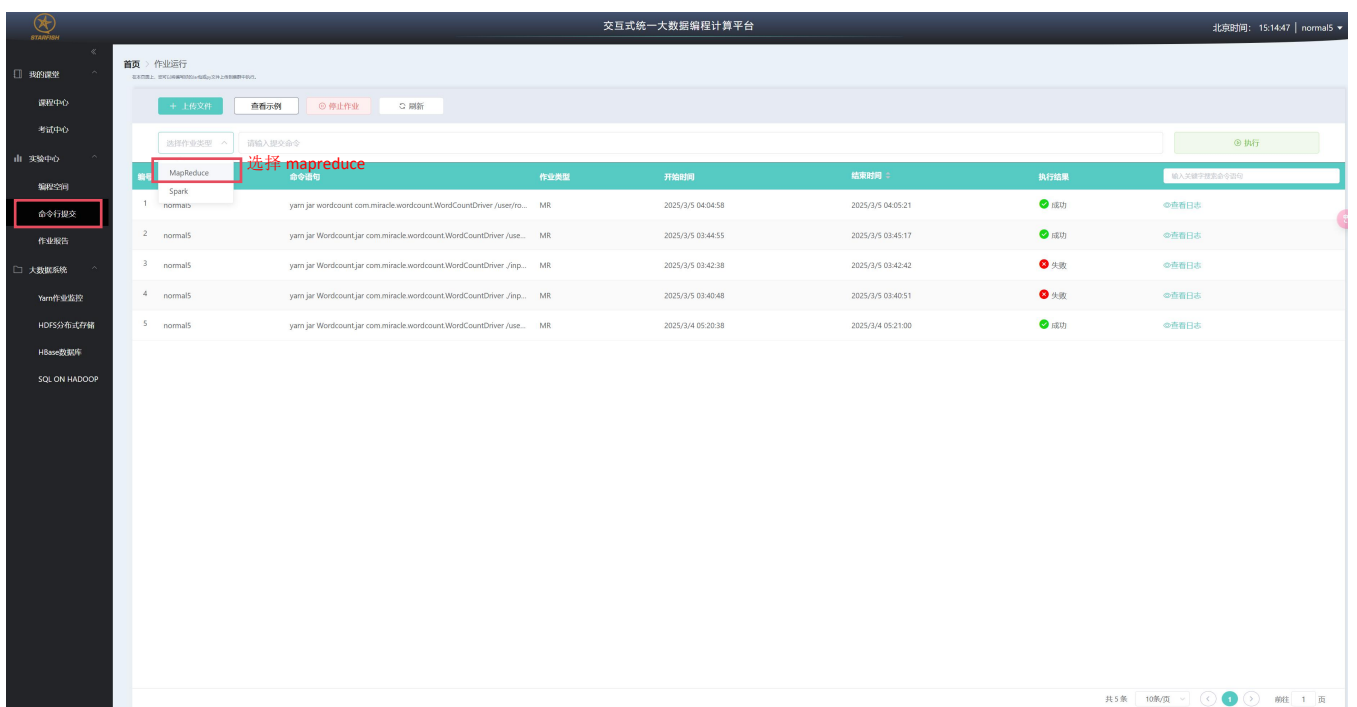
用户文件空间可从 **hdfs** 分布式存储查看，可将自己的数据文件上传到此处，以及存储执行后的输出文档。可以自己创建文件夹并将数据上传到指定文件夹。



注意此时的文件存储路径在 **hdfs://user/username** 路径下，此路径为用户自己的文件空间。

3. 提交可执行文件

上传 **jar** 包后可以从命令行提交处，使用服务器资源执行自己的文件。并可以根据自己的文件类型选择不同的提交方式（**mapreduce or spark**），在此只演示 **mapreduce** 文件的提交，关于如何写提交文件的指令可以从查看实例中看。



指令格式为 yarn jar jar 包名 主类的完整名 input_path output_path
其中 yarn jar 已经默认给出，直接按照后面的格式书写指令。
如：上传 jar 包名为 Wordcount.jar，其中要执行的主类在 com.miracle.wordcount
下，名为 WordCountDriver。在此选择最复杂的情况，输入数据存储在 root 用户
的文件夹中（完整数据集），故输入地址为/user/root/wordcount.txt
输出地址为自己用户空间的 testout 文件夹，故输出地址为./testout
完整指令： yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver
/user/root/wordcount.txt ./testout
执行后可以看到执行结果，如果失败可以查看日志找到错误原因

交互式统一大数据编程计算平台

北京时间: 15:32:30 | normal5

首页 作业运行

MapReduce yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver /user/root/wordcount.txt ./testout

序号	用户	命令语句	作业类型	开始时间	结束时间	执行结果	操作
1	normal5	yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver /use...	MR	2025/3/7 04:31:20	2025/3/7 04:51:42	成功	查看日志
2	normal5	yarn jar wordcount com.miracle.wordcount.WordCountDriver /usercho...	MR	2025/3/5 04:04:58	2025/3/5 04:05:21	成功	查看日志
3	normal5	yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver /use...	MR	2025/3/5 03:44:55	2025/3/5 03:45:17	成功	查看日志
4	normal5	yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver ./inp...	MR	2025/3/5 03:42:38	2025/3/5 03:42:42	失败	查看日志
5	normal5	yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver ./inp...	MR	2025/3/5 03:40:48	2025/3/5 03:40:51	失败	查看日志
6	normal5	yarn jar Wordcount.jar com.miracle.wordcount.WordCountDriver /use...	MR	2025/3/4 05:20:38	2025/3/4 05:21:00	成功	查看日志

共 6 条 10 条/页 1 页

交互式统一大数据编程计算平台

北京时间: 15:40:30 | normal5

首页 HDFS分布式文件存储

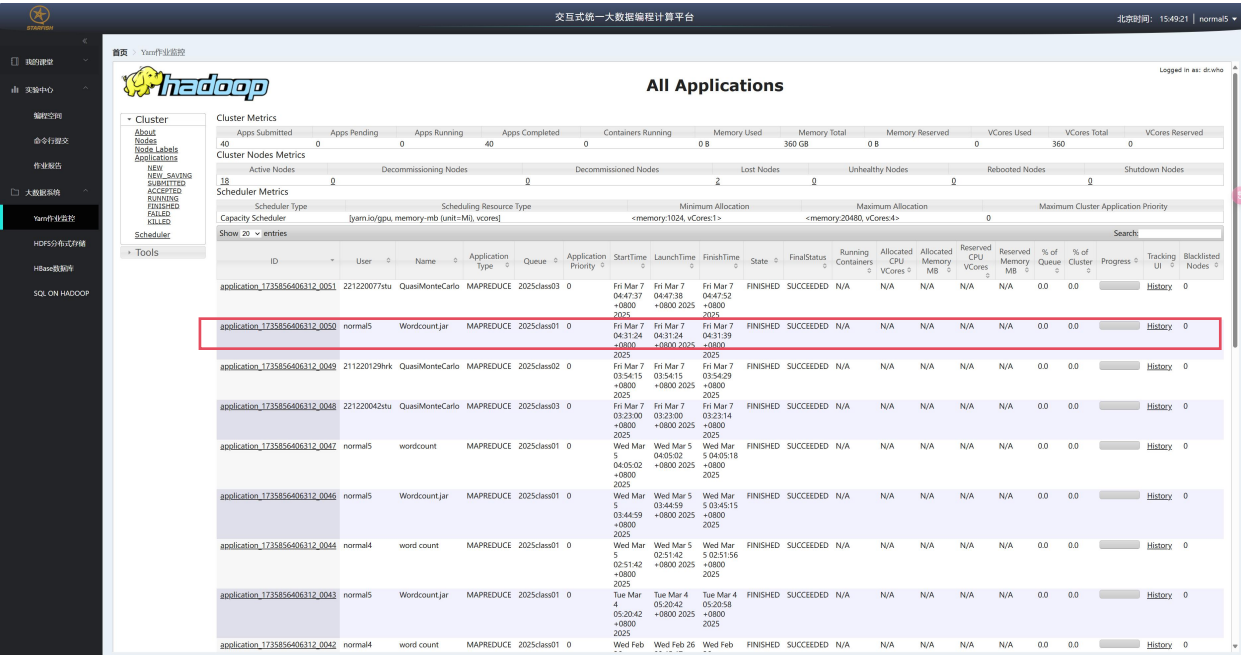
上传文件 新建文件夹 删除

/user/normal5

文件名称	大小	用户	用户组	权限	修改时间	操作
Trash	0B	normal5	supergroup	drwx-----	2025/3/7 04:25:02	查看 权限 删除
data	0B	normal5	supergroup	drwx-----	2025/2/25 04:25:13	查看 权限 删除
output	0B	normal5	supergroup	drwxr-xr-x	2025/3/4 05:21:27	查看 权限 删除
testout	0B	normal5	supergroup	drwxr-xr-x	2025/3/7 04:31:06	查看 权限 删除

共 4 条 10 条/页 1 页

可以通过 yarn 作业监控查看提交记录



注：为保证实验平台的稳定运行，请在本地进行测试，通过样例后，再将代码上传至实验平台运行完整数据集，请勿在实验平台进行测试，导致资源的浪费。