1. **算法说明：**

推荐学员列表，预测学员报班。

1. **部署方式：**

****

**2.1定时任务：**

1. 拷贝preprocess6month.jar文件至指定文件夹。
2. 修改preprocess6month.sh 最后一行preprocess6month.jar位置，如果大数据环境变量有修改，做相应修改。
3. 拷贝preprocess6month.sh至指定文件脚本文件夹，并使用chmod 修改为可执行文件。

chmod +x preprocess6month.sh

1. 添加定时任务，时间应该定于上一月数据全部入库后的某个凌晨。

【示例中使用每月2日凌晨2点; /zhouzhou/preprocess6month.sh 部分需要根据3）进行修改。】

示例：

# crontab –e

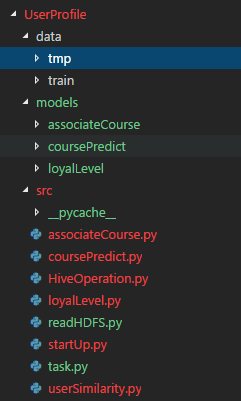
0 2 2 \* \* /zhouzhou/preprocess6month.sh

**或手动运行方式：**

./spark-submit --class "cn.xdf.www.analysis.etl.preprocess6month" preprocess6month.jar

**2.2 定时持久化至hive**

1. 获取代码（如下图所示），并将代码（按照目前结构），放置服务器相应位置（比如：/root下面）



1. 在本地文件中删除不必要的文件/文件夹（可选）：

UserProfile/models/associateCourse

UserProfile/models/ loyalLevel

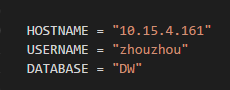
UserProfile/src/ associateCourse.py

UserProfile/src/ loyalLevel.py

UserProfile/src/ startUp.py

UserProfile/src/ userSimilarity.py

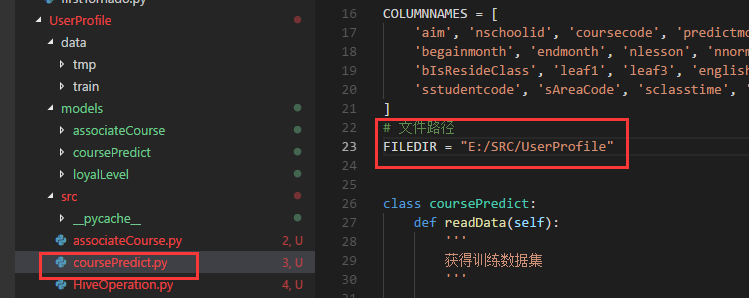
1. 编辑UserProfile/src/ HiveOperation.py，如下图所示



修改host,用户名以及数据库名

1. 修改模型路径：

在此处只需要修改课程预测的模型文件地址，如下图所示



修改FILEDIR为绝对路径，至UserProfile目录，在此处，修改为

FILEDIR=”/root/UserProfile”

1. 配置定时任务

Crontab –e

0 4 2 \* \* /Anaconda3/bin/python /root/UserProfile/src/task.py

在此处配置为每月2号凌晨4点,具体可以修改，

/Anaconda3/bin/python 为安装的python环境绝对路径，视具体而定

/root/UserProfile/src/task.py 为执行文件的绝对路径，视具体而定，但是必须指定task.py文件为执行文件

1. **查询方式：**

**表名为**dw. dw\_student\_predecit

建表脚本：

**学生课程推荐：**

* + 1. **结果**

|  |  |
| --- | --- |
| nschoolid | 学校id |
| coursecode | 课程代码 |
| predictmonth | 预测月份 |
| birthday | 出生年份 |
| gender | 性别 |
| begainmonth | 开课月份 |
| endmonth | 节课月份 |
| nlesson | 课程数 |
| nnormalcount | 班容 |
| bvip | VIP标识 |
| bIsResideClass | 住宿标识 |
| leaf1 | 课程一级分类 |
| leaf3 | 课程三级分类 |
| english | 英语课程标识 |
| steplevel | 学习阶段 |
| Enrollmonth | 报名年份 |
| sstudentcode | 学生代码 |
| sAreaCode | 校区代码 |
| sclasstime | 上课时间段 |
| bvalid | 退班标识 |
| predict | 预测结果 |