1024：TC代码大奖赛

任务规则说明

**截止时间**：2019年10月31日  
**参与方式**：个人或组队均可，将答案发送至TC@youxin.com  
**评选时间**：2019年11月初，由技术委员会根据答案内容评分，评分最高者获胜（相同分数以提交时间先后为准）  
**获奖金额**：RMB 3,000

如有疑问，请联系技术委员会秘书 窦琪熠 （[douqiyi@youxin.com](mailto:douqiyi@youxin.com) ）

任务内容

用户每一次借款根据期数会生成对应多期还款计划,记录在mysql中还款计划表repay\_plan中, 数据库主要字段如下所示，

1 放款后初始状态是待还款，用户到期如果未还状态会变为待还款

2 用户每次还款可以选择一期进行还款，也可以一次还多期，但只能从待还款状态最小一起开始还（待还款状态包括： 0待还款 1 部分还款 3 逾期）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | *还款计划Key* | *借款信息Key* | *借款人* | *应还期次* | 逾期日期 | *该期状态*   0待还款1 部分还款 2 已还清 3 逾期 | *剩余未还本金* |
| *id* | *r*epayPlanKey | loanKey | userKey | period | due\_date | status | *leftPrincipal* |
| 1 | 4ab6... | 73a9... | 7sd8... | 1 | 2019-09-24 | 2 | 16832.55 |
| 2 | asd6... | 73a9... | 7sd8... | 2 | 2019-10-24 | 1 | 1000 |
| 3 | 3kas... | 73a9... | 7sd8... | 3 | 2019-11-24 | 0 | 15987.30 |
| 4 | asd6... | 88s5... | 7u8d... | 1 | 2019-09-22 | 3 | 1256.2 |
| 5 | 5s7f... | 88s5... | 7u8d... | 2 | 2019-10-22 | 3 | 1262.2 |
| 6 | 9ss9... | 88s5... | 7u8d... | 3 | 2019-11-22 | 0 | 1286.2 |

假如数据量50万，回答以下2个问题：

**1 请用一条sql，找出所有*借款*信息中最小一期未还的并且最小一期*剩余未还本金*最大的Top100 还款信息（基础分60）**

   重点：sql基本语法的使用

结果：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | *还款计划Key* | *借款信息Key* | *借款人* | *应还期次* | 逾期日期 | *该期状态*   0待还款1 部分还款 2 已还清 3 逾期 | *剩余未还本金* |
| *id* | *r*epayPlanKey | loanKey | userKey | period | due\_date | status | *leftPrincipal* |
| 2 | asd6... | 73a9... | 7sd8... | 2 | 2019-10-24 | 1 | 1000 |
| 4 | asd6... | 88s5... | 7u8d... | 1 | 2019-09-22 | 3 | 1256.2 |

**2 请用一条sql，找出所有*借款*信息中最小第二期未还的并且最小第二期*剩余未还本金*最大的Top100 还款信息（高级分80）**

  重点：灵活使用sql完成复杂业务需求，有多种实现方案，要考虑效率

返回结果：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | *还款计划Key* | *借款信息Key* | *借款人* | *应还期次* | 逾期日期 | *该期状态*   0待还款1 部分还款 2 已还清 3 逾期 | *剩余未还本金* |
| *id* | *r*epayPlanKey | loanKey | userKey | period | due\_date | status | *leftPrincipal* |
| 3 | 3kas... | 73a9... | 7sd8... | 3 | 2019-11-24 | 0 | 15987.30 |
| 5 | 5s7f... | 88s5... | 7u8d... | 2 | 2019-10-22 | 3 | 1262.2 |

**3 随着业务量增加，每天晚上mysql会同步到hive的表中（**repay\_plan\_snapshot\_1024**）保留一份snapshot，假如每天数据量达到10亿条，现在要从hive数据库中快速找出q2的答案，有哪些方案？（100分）**

  重点： 每天10亿条数据都必须计算一次才能获取到答案，开放式答案，可以考虑流式计算框架，比如采用SPARK，flink，（需要描述计算中涉及到的方法）。可以使用java8中流式计算框架stream简单描述计算过程：

输入：List<RepayPlan> replayPlanList;

List<RepayPlan> = replayPlanList.stream…….;