1.计算留存率

表 a 中存放着每日新用户的 id，表 b 中存放每日活跃用户的 id。

请用 HiveSQL 或 SparkSQL 实现得到如图 c 的数据结果，即得到每月的新用户数，并计算出其接下来 12 个月的每月留存。

如：2021-11 的新增用户数为 320302，该月新用户在 12 月留存率为 67%（即 11 月新增用户在 12 月继续活跃的用户数/11 月新增用户数），在 1 月的留存率为 56%，在 2 月的留存率为 42%。

表 a 和表 b 的字段示例：

p_date	p_id
2021-01-01	adb
2021-01-01	ghj
2021-01-01	ujj
…	…

图 c:

数据日期(年-月)	新用户	1月后	2月后	3月后	4月后	5月后	6月后	7月后	8月后	9月后	10月后	11月后	12月后
2022-02	176636												
2022-01	287045	67%											
2021-12	219224	67%	33%										
2021-11	320302	67%	56%	42%									
2021-10	247865	78%	33%	42%	42%								
2021-09	231188	78%	33%	42%	32%	32%							
2021-08	360942	78%	56%	42%	32%	42%	32%						
2021-07	362361	67%	44%	42%	42%	32%	32%	32%					
2021-06	336531	56%	44%	42%	42%	42%	32%	32%	42%				
2021-05	362928	56%	56%	32%	32%	42%	32%	32%	42%	32%			
2021-04	230930	44%	56%	42%	32%	42%	42%	42%	42%	42%	42%		
2021-03	172916	78%	56%	42%	42%	42%	42%	32%	42%	42%	32%	42%	
2021-02	278928	44%	33%	42%	32%	32%	42%	32%	32%	32%	32%	42%	32%
2021-01	352363	67%	44%	32%	32%	32%	42%	42%	32%	42%	32%	42%	42%

2.大表计算的优化方案

表 ods_a 中 2021-01-01 这天有 100 亿条日志如图 a，请用 HiveSQL 或 SparkSQL 得每个 id 在当天的首末条日志并标记，如图 b。由于日志量较大，请提供性能较好的实现方案。

图 a:

p_date	p_time	p_id	p_remark
2021-01-01	2021-01-01 10:33:53	adb	34
2021-01-01	2021-01-01 11:33:54	adb	35
2021-01-01	2021-01-01 12:33:54	adb	65
2021-01-01	2021-01-01 10:33:55	ujj	5454
…	…	…	…
2021-01-01	2021-01-01 10:00:00	ujj	55

图 b:

p_date	p_time	p_id	p_remark	p_label
2021-01-01	2021-01-01 10:00:00	ujj	55	first
2021-01-01	2021-01-01 10:33:55	ujj	5454	last
…	…	…	…	
2021-01-01	2021-01-01 10:33:53	adb	34	first
2021-01-01	2021-01-01 12:33:54	adb	65	last

3.实现外部数据获取（可选）

外部提供了一个 API URL（里面带日期参数如 https://***&date=YYY），通过该接口获取以天为粒度的数据，格式为压缩、加密的 sql 数据，请写一段代码实现：数据的获取并存入 mysql 库中。可以写一段 shell 脚本或其他解决方案。

4.数据处理（可选）

请写一段代码实现该需求：URL 用传参的方式通过 get 请求获取接口数据，然后判断 URL 返回值状态，status=0 为请求正常，如请求错误将完整 URL 储存到 log 文件中，如请求成功则将 json 值转成 list 并按照 level 等级进行倒序。

```
url = 'http://*****?cp=%s&game=%s&serverId=%s&lenovoId=%s'
```

请求参数为 game,cp;

请求正常结果为:

{"status":0,"message":"\u7\u0\u5channel\u5c","data":"[{“lenovoid”:110000,”gameid”:2222,”level”:20},{“lenovoid”:110000,”gameid”:2222,”level”:30},{“lenovoid”:110000,”gameid”:2222,”level”:10}]"}

;

json 解析成 list 的格式

[[('lenovoid',11000),('level',20)],[('lenovoid',11000),('level',30)],[('lenovoid',11000),('level',10)]]