

新用户数和日活跃数

题目链接：

<https://www.nowcoder.com/practice/dbbc9b03794a48f6b34f1131b1a903eb?tpId=268&tqId=2285346&ru=/exam/oj&qru=/ta/sql-factory-interview/question-ranking&sourceUrl=%2Fexam%2Foj%3Ftab%3DSQL%25E7%25AF%2587%26topicId%3D268>

描述

用户行为日志表tb_user_log

id	uid	artical_id	in_time	out_time	sign_cin
1	101	9001	2021-10-31 10:00:00	2021-10-31 10:00:09	0
2	102	9001	2021-10-31 10:00:00	2021-10-31 10:00:09	0
3	101	0	2021-11-01 10:00:00	2021-11-01 10:00:42	1
4	102	9001	2021-11-01 10:00:00	2021-11-01 10:00:09	0
5	108	9001	2021-11-01 10:00:01	2021-11-01 10:00:50	0
6	108	9001	2021-11-02 10:00:01	2021-11-02 10:00:50	0
7	104	9001	2021-11-02 10:00:28	2021-11-02 10:00:50	0
8	106	9001	2021-11-02 10:00:28	2021-11-02 10:00:50	0
9	108	9001	2021-11-03 10:00:01	2021-11-03 10:00:50	0
10	109	9002	2021-11-03 11:00:55	2021-11-03 11:00:59	0
11	104	9003	2021-11-03 11:00:45	2021-11-03 11:00:55	0
12	105	9003	2021-11-03 11:00:53	2021-11-03 11:00:59	0
13	106	9003	2021-11-03 11:00:45	2021-11-03 11:00:55	0

(uid-用户ID, artical_id-文章ID, in_time-进入时间, out_time-离开时间, sign_in-是否签到)

问题：统计每天的日活数及新用户占比

注：

- 新用户占比=当天的新用户数÷当天活跃用户数（日活数）。
- 如果in_time-进入时间和out_time-离开时间跨天了，在两天里都记为该用户活跃过。
- 新用户占比保留2位小数，结果按日期升序排序。

输出示例：

示例数据的输出结果如下

dt	dau	uv_new_ratio
2021-10-30	2	1.00
2021-11-01	3	0.33
2021-11-02	3	0.67
2021-11-03	5	0.40

解释：

2021年10月31日有2个用户活跃，都为新用户，新用户占比1.00；

2021年11月1日有3个用户活跃，其中1个新用户，新用户占比0.33；

示例1

输入：

```

1 DROP TABLE IF EXISTS tb_user_log;
2 CREATE TABLE tb_user_log (
3     id INT PRIMARY KEY AUTO_INCREMENT COMMENT '自增ID',
4     uid INT NOT NULL COMMENT '用户ID',
5     artical_id INT NOT NULL COMMENT '视频ID',
6     in_time datetime COMMENT '进入时间',
7     out_time datetime COMMENT '离开时间',
8     sign_in TINYINT DEFAULT 0 COMMENT '是否签到'
9 ) CHARACTER SET utf8 COLLATE utf8_bin;
10
11 INSERT INTO tb_user_log(uid, artical_id, in_time, out_time, sign_in) VALUES
12     (101, 9001, '2021-10-31 10:00:00', '2021-10-31 10:00:09', 0),
13     (102, 9001, '2021-10-31 10:00:00', '2021-10-31 10:00:09', 0),
14     (101, 0, '2021-11-01 10:00:00', '2021-11-01 10:00:42', 1),
15     (102, 9001, '2021-11-01 10:00:00', '2021-11-01 10:00:09', 0),
16     (108, 9001, '2021-11-01 10:00:01', '2021-11-01 10:01:50', 0),
17     (108, 9001, '2021-11-02 10:00:01', '2021-11-02 10:01:50', 0),
18     (104, 9001, '2021-11-02 10:00:28', '2021-11-02 10:00:50', 0),
19     (106, 9001, '2021-11-02 10:00:28', '2021-11-02 10:00:50', 0),
20     (108, 9001, '2021-11-03 10:00:01', '2021-11-03 10:01:50', 0),
21     (109, 9002, '2021-11-03 11:00:55', '2021-11-03 11:00:59', 0),
22     (104, 9003, '2021-11-03 11:00:45', '2021-11-03 11:00:55', 0),
23     (105, 9003, '2021-11-03 11:00:53', '2021-11-03 11:00:59', 0),
24     (106, 9003, '2021-11-03 11:00:45', '2021-11-03 11:00:55', 0);

```

复制

输出：

```

1 2021-10-31|2|1.00
2 2021-11-01|3|0.33
3 2021-11-02|3|0.67
4 2021-11-03|5|0.40

```

分析

```
1  SELECT * FROM tb_user_log
2  # 分析: tb_user_log 里的数据里 用户 u_id 每天(in_time)可以进入多次, 所以 u_id 可以有
   多条记录。
3
4  SELECT count(DISTINCT uid) 人数 FROM tb_user_log
5
6  # union : 合并相同的列, 每列不允许重复
7  # union all : 合并相同的列每列允许重复
8
9  # 查找每天的新用户数 : 新用户: 首次使用该软件或者首次进入
10 SELECT uid, DATE(min(in_time)) 进入时间 FROM tb_user_log GROUP BY uid # 每个用
   户首次进入的时间
11
12
13 -- 每天的新用户数:
14 DROP VIEW IF EXISTS V_NewUser;
15 CREATE VIEW V_NewUser
16 AS
17 SELECT dt, count(uid) NewUser
18       FROM (SELECT uid, DATE(min(in_time)) dt FROM tb_user_log GROUP
   BY uid) a
19            GROUP BY dt
20
21 SELECT * FROM V_NewUser
22 -- 每日活跃用户数(DAU): 日期分组    用户计数
23 DROP VIEW IF EXISTS v_dau;
24 CREATE VIEW v_dau
25 as
26 SELECT dt, COUNT(uid) dau
27       FROM
28       (select uid, date(in_time) dt from tb_user_log
29        union
30        select uid, date(out_time) dt from tb_user_log) b
31 GROUP BY b.dt
32
33 SELECT * FROM v_dau
34 -- 先统计新用户, 再统计日活数(DAU), 再计算新用户占比。
35
36 SELECT a.dt,
37        b.dau,
38        IFNULL(ROUND(a.newuser/b.dau, 2), 0) uv_new_ratio
39 FROM v_dau b
40 LEFT JOIN V_NewUser a
41 ON a.dt = b.dt
42 ORDER BY a.dt
43
44
45 SELECT * FROM V_NewUser;
46 SELECT * FROM v_dau;
```

答案

```
1  SELECT b.dt,
```

```
2         b.dau,
3         IFNULL(ROUND(a.newuser/b.dau, 2) ,0) uv_new_ratio
4 FROM
5
6 (SELECT dt, count(uid) NewUser
7  FROM (SELECT uid, DATE(min(in_time)) dt FROM tb_user_log GROUP
8  BY uid) a
9  GROUP BY dt) a      -- 新用户
10
11 Right JOIN
12 (SELECT dt, COUNT(uid) dau
13  FROM
14  (select uid, date(in_time) dt from tb_user_log
15  union
16  select uid, date(out_time) dt from tb_user_log) b
17  GROUP BY b.dt) b      -- 日活跃
18 ON a.dt = b.dt
19 ORDER BY b.dt
```