# 第1题

我们有如下的用户访问数据

|  |  |  |
| --- | --- | --- |
| userId | visitDate | visitCount |
| u01 | 2017/1/21 | 5 |
| u02 | 2017/1/23 | 6 |
| u03 | 2017/1/22 | 8 |
| u04 | 2017/1/20 | 3 |
| u01 | 2017/1/23 | 6 |
| u01 | 2017/2/21 | 8 |
| u02 | 2017/1/23 | 6 |
| u01 | 2017/2/22 | 4 |

要求使用SQL统计出每个用户的累积访问次数，如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| 用户id | 月份 | 小计 | 累积 |
| u01 | 2017-01 | 11 | 11 |
| u01 | 2017-02 | 12 | 23 |
| u02 | 2017-01 | 12 | 12 |
| u03 | 2017-01 | 8 | 8 |
| u04 | 2017-01 | 3 | 3 |

数据：

u01 2017/1/21 5

u02 2017/1/23 6

u03 2017/1/22 8

u04 2017/1/20 3

u01 2017/1/23 6

u01 2017/2/21 8

u02 2017/1/23 6

u01 2017/2/22 4

1）创建表

create table action

(userId string,

visitDate string,

visitCount int)

row format delimited fields terminated by "\t";

load data local inpath '/opt/module/hive/datas/action.txt' into table action;

答案

select

t2.userId,

t2.visitdate,

t2.visitsum,

sum(t2.visitsum)over(partition by userId order by t2.visitdate)

from (select

t1.userId,

t1.visitdate,

sum(t1.visitCount) visitsum

from (select

userId,

date\_format(regexp\_replace(visitdate,"/","-"),'yyyy-MM') visitdate,

visitCount

from action)t1

group by t1.userId,t1.visitdate)t2;

# 第2题

有50W个京东店铺，每个顾客访问任何一个店铺的任何一个商品时都会产生一条访问日志，访问日志存储的表名为Visit，访客的用户id为user\_id，被访问的店铺名称为shop，访问时间为visit\_time。数据样例：'huawei','1001','2017-02-10'，'apple','1001','2017-02-11'……

请统计：

1. 每个店铺的UV（访客数）

2）每个店铺访问次数top3的访客信息。输出店铺名称、访客id、访问次数

建表：

drop table if exists Visit;

create table Visit(

shop string COMMENT '店铺名称',

user\_id string COMMENT '用户id',

visit\_time string COMMENT '访问时间'

)

row format delimited fields terminated by '\t';

插入数据：

huawei 1005 2017-02-10

huawei 1005 2017-02-10

huawei 1005 2017-02-10

huawei 1005 2017-02-10

huawei 1004 2017-02-10

huawei 1004 2017-02-10

huawei 1003 2017-02-10

huawei 1003 2017-02-10

huawei 1001 2017-02-10

huawei 1002 2017-02-10

huawei 1006 2017-02-10

apple 1001 2017-02-10

apple 1001 2017-02-10

apple 1001 2017-02-10

apple 1001 2017-02-10

apple 1002 2017-02-10

apple 1002 2017-02-10

apple 1005 2017-02-10

apple 1005 2017-02-10

apple 1006 2017-02-10

apple 1004 2017-02-10

meizu 1006 2017-02-10

meizu 1006 2017-02-10

meizu 1006 2017-02-10

meizu 1006 2017-02-10

meizu 1003 2017-02-10

meizu 1003 2017-02-10

meizu 1003 2017-02-10

meizu 1002 2017-02-10

meizu 1002 2017-02-10

meizu 1004 2017-02-10

第一问答案

select

shop,

count(\*) shop\_uv

from visit

group by shop

order by shop\_uv desc;

第二问答案

select

t2.shop,

t2.user\_id,

t2.shop\_user\_uv

from (select

t1.shop,

t1.user\_id,

t1.shop\_user\_uv,

rank()over(partition by shop order by shop\_user\_uv desc ) rk

from (select

shop,

user\_id,

count(\*) shop\_user\_uv

from visit

group by shop,user\_id)t1)t2

where t2.rk <=3;

# 第3题

以下表记录了用户每天的蚂蚁森林低碳生活领取的记录流水。

table\_name：user\_low\_carbon

user\_id  data\_dt   low\_carbon

用户      日期       减少碳排放（g）

蚂蚁森林植物换购表，用于记录申领环保植物所需要减少的碳排放量

table\_name:  plant\_carbon

plant\_id  plant\_name  low\_carbon

植物编号 植物名 换购植物所需要的碳

----题目

1.蚂蚁森林植物申领统计

问题：假设2017年1月1日开始记录低碳数据（user\_low\_carbon），假设2017年10月1日之前满足申领条件的用户都申领了一颗p004-胡杨，

剩余的能量全部用来领取“p002-沙柳” 。

统计在10月1日累计申领“p002-沙柳” 排名前10的用户信息；以及他比后一名多领了几颗沙柳。

得到的统计结果如下表样式：

user\_id  plant\_count less\_count(比后一名多领了几颗沙柳)

u\_101    1000         100

u\_088    900          400

u\_103    500          …

2、蚂蚁森林低碳用户排名分析

问题：查询user\_low\_carbon表中每日流水记录，条件为：

用户在2017年，连续三天（或以上）的天数里，

每天减少碳排放（low\_carbon）都超过100g的用户低碳流水。

需要查询返回满足以上条件的user\_low\_carbon表中的记录流水。

例如用户u\_002符合条件的记录如下，因为2017/1/2~2017/1/5连续四天的碳排放量之和都大于等于100g：

seq（key） user\_id data\_dt  low\_carbon

xxxxx10    u\_002  2017/1/2  150

xxxxx11    u\_002  2017/1/2  70

xxxxx12    u\_002  2017/1/3  30

xxxxx13    u\_002  2017/1/3  80

xxxxx14    u\_002  2017/1/4  150

xxxxx14    u\_002  2017/1/5  101

备注：统计方法不限于sql、procedure、python,java等

提供的数据说明：

user\_low\_carbon：

u\_001 2017/1/1 10

u\_001 2017/1/2 150

u\_001 2017/1/2 110

u\_001 2017/1/2 10

u\_001 2017/1/4 50

u\_001 2017/1/4 10

u\_001 2017/1/6 45

u\_001 2017/1/6 90

u\_002 2017/1/1 10

u\_002 2017/1/2 150

u\_002 2017/1/2 70

u\_002 2017/1/3 30

u\_002 2017/1/3 80

u\_002 2017/1/4 150

u\_002 2017/1/5 101

u\_002 2017/1/6 68

...

plant\_carbon：

p001 梭梭树 17

p002 沙柳 19

p003 樟子树 146

p004 胡杨 215

...

1.创建表

create table user\_low\_carbon(user\_id String,data\_dt String,low\_carbon int) row format delimited fields terminated by '\t';

create table plant\_carbon(plant\_id string,plant\_name String,low\_carbon int) row format delimited fields terminated by '\t';

2.加载数据

load data local inpath "/opt/module/hive/datas/user\_low\_carbon.txt" into table user\_low\_carbon;

load data local inpath "/opt/module/hive/datas/plant\_carbon.txt" into table plant\_carbon;

3.设置本地模式

set hive.exec.mode.local.auto=true;