访客分析

归总

1. 就是业务名称需要解释成sql伪代码比较难

1.独立访客:其实就是访客, 也就是对应的不重复的ip的个数

2.新访客: 就是老的访客表里没有的,要新建新老访客表,然后把今日的访客表和老的访客表对比,只有老的访客表里没有的就是新的访客,然后把他加入新的访客表,最后还有加入到老的访客表中,表示他现在也算是老的访客了

# -- 独立访客

--需求：按照时间维度来统计独立访客及其产生的pv量

按照时间维度比如小时来统计独立访客及其产生的 pv 。

时间维度：时

drop table dw\_user\_dstc\_ip\_h;

create table dw\_user\_dstc\_ip\_h(

remote\_addr string,

pvs bigint,

hour string);

insert into table dw\_user\_dstc\_ip\_h

select remote\_addr,count(1) as pvs,concat(month,day,hour) as hour

from ods\_weblog\_detail

Where datestr='20130918'

group by concat(month,day,hour),remote\_addr;

--在上述基础之上，可以继续分析，比如每小时独立访客总数

select count(1) as dstc\_ip\_cnts,hour from dw\_user\_dstc\_ip\_h group by hour;

时间维度：日

select remote\_addr,count(1) as counts,concat(month,day) as day

from ods\_weblog\_detail

Where datestr='20130918'

group by concat(month,day),remote\_addr;

时间维度： 月

select remote\_addr,count(1) as counts,month

from ods\_weblog\_detail

group by month,remote\_addr;

----------------------------------------------------------------------------------------

-- 每日新访客

-- 需求：将每天的新访客统计出来。

--历日去重访客累积表

drop table dw\_user\_dsct\_history;

create table dw\_user\_dsct\_history(

day string,

ip string)

partitioned by(datestr string);

--每日新访客表

drop table dw\_user\_new\_d;

create table dw\_user\_new\_d (

day string,

ip string)

partitioned by(datestr string);

# 新老访客建立

-- 历史访客表里没有的就是新访客

-- 这里可以关联访客表和历史访客表, 左关联不上历史访问表的就是新房客

-- 就说明第一次来访,

-- 细节: 同样的ip需要把它过滤掉, 还有每日, 也需要对日期去重,精准到天

select a.remote\_addr ,a.day

from (

select remote\_addr,'20130918' as day

from ods\_weblog\_detail newIp

where datestr ='20130918'

group by remote\_addr

) a

left join dw\_user\_dsct\_history hist

on a.remote\_addr = hist.ip

where hist.ip is null;

--每日新用户插入新访客表

insert into table dw\_user\_new\_d partition(datestr='20130918')

select tmp.day as day,tmp.today\_addr as new\_ip

from

(

select today.day as day,today.remote\_addr as today\_addr,old.ip as old\_addr

from

(

select distinct remote\_addr as remote\_addr,"20130918" as day

from ods\_weblog\_detail where datestr="20130918"

) today

left outer join

dw\_user\_dsct\_history old

on today.remote\_addr=old.ip

) tmp

where tmp.old\_addr is null;

--每日新用户追加到历史累计表

insert into table dw\_user\_dsct\_history partition(datestr='20130918')

select day,ip from dw\_user\_new\_d where datestr='20130918';

验证：

select count(distinct remote\_addr) from ods\_weblog\_detail;

select count(1) from dw\_user\_dsct\_history where datestr='20130918';

select count(1) from dw\_user\_new\_d where datestr='20130918';