访客visit分析

# 归总

1.案例代码总结

distinct的去重是在reduce端通过整成一个reducet分区再去重的,应该用group by在map端去重,对大量的数据来说代码更好

-- 代码优化后

代码感悟

1.想要两个或是多个字段做比较(加减乘除,大小等运算),那就需要关联其他的表(包括自己),

也就是说要先把需要做比较的两列用临时表的方式给弄到一起,

然后再select 表1.字段和表2.字段做比较就可以看

2.业务分析

1.人均访问频次,

使用所有的独立访问的人，即独立的session个数除以所有的去重IP即可

2.回头访客

查询同一个ip出现的次数大于1就是回头访客

# -- 回头/单次访客统计

-- 在访客表中按照ip分组,如果组成员数据大于1的说明就是来过来回, 也就是回头访问

select remote\_addr ,count(remote\_addr) ipcount

from ods\_click\_stream\_visit

group by remote\_addr

having ipcount > 1

-- 查询今日所有回头访客及其访问次数。

drop table dw\_user\_returning;

create table dw\_user\_returning(

day string,

remote\_addr string,

acc\_cnt string)

partitioned by (datestr string);

insert overwrite table dw\_user\_returning partition(datestr='20130918')

select '20130918',remote\_addr ,count(remote\_addr) ipcount

from ods\_click\_stream\_visit

group by remote\_addr

having ipcount > 1

-- 太复杂了

select tmp.day,tmp.remote\_addr,tmp.acc\_cnt

from

(select '20130918' as day,remote\_addr,count(session) as acc\_cnt from ods\_click\_stream\_visit

group by remote\_addr) tmp

where tmp.acc\_cnt>1;

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

-- 人均访问频次,使用所有的独立访问的人，即独立的session个数除以所有的去重IP即可

-- 计算出一共访问了多少次，除以一共有多少人

-- 频次表示一共访问了多少次，所有页面访问的次数累加即可

select sum(pagevisits)

from ods\_click\_stream\_visit

76

53

select count(distinct remote\_addr) from ods\_click\_stream\_visit

select count(tmp.total\_count)

from (

select remote\_addr,count(1) as total\_count

from ods\_click\_stream\_visit

group by remote\_addr

) tmp

53

-- 人均访问的频次，频次表示我们来了多少个session

-- 次数都是使用session来进行区分，一个session就是表示一次

select count(session)/count(distinct remote\_addr) from ods\_click\_stream\_visit where datestr='20130918';

-- distinct的去重是在reduce端通过整成一个reducet分区再去重的,应该应group by现在map端去重,对大量的数据来说代码更好

-- 代码优化后

-- 代码感悟

-- 1.想要两个或是多个字段做比较(加减乘除,大小等运算),那就需要关联其他的表(包括自己),

-- 也就是说要先把需要做表的两列用临时表的方式给弄到一起,

-- 然后再select 表1.字段和表2.字段做比较就可以看

SELECT t.se/te.ip

FROM

(SELECT COUNT(1) ip

FROM(

SELECT COUNT(1) ip,remote\_addr

FROM ods\_click\_stream\_visit

WHERE datestr='20130918'

GROUP BY remote\_addr

)p

) te

LEFT JOIN

(SELECT COUNT(SESSION) se

FROM ods\_click\_stream\_visit

WHERE datestr='20130918'

) t

select count(1)

from ods\_click\_stream\_visit

where datestr ='20130918'

# -- 人均页面浏览量，所有的页面点击次数累加除以所有的独立去重IP总和即可

select sum(pagevisits)/count(distinct remote\_addr) from ods\_click\_stream\_visit where datestr='20130918';