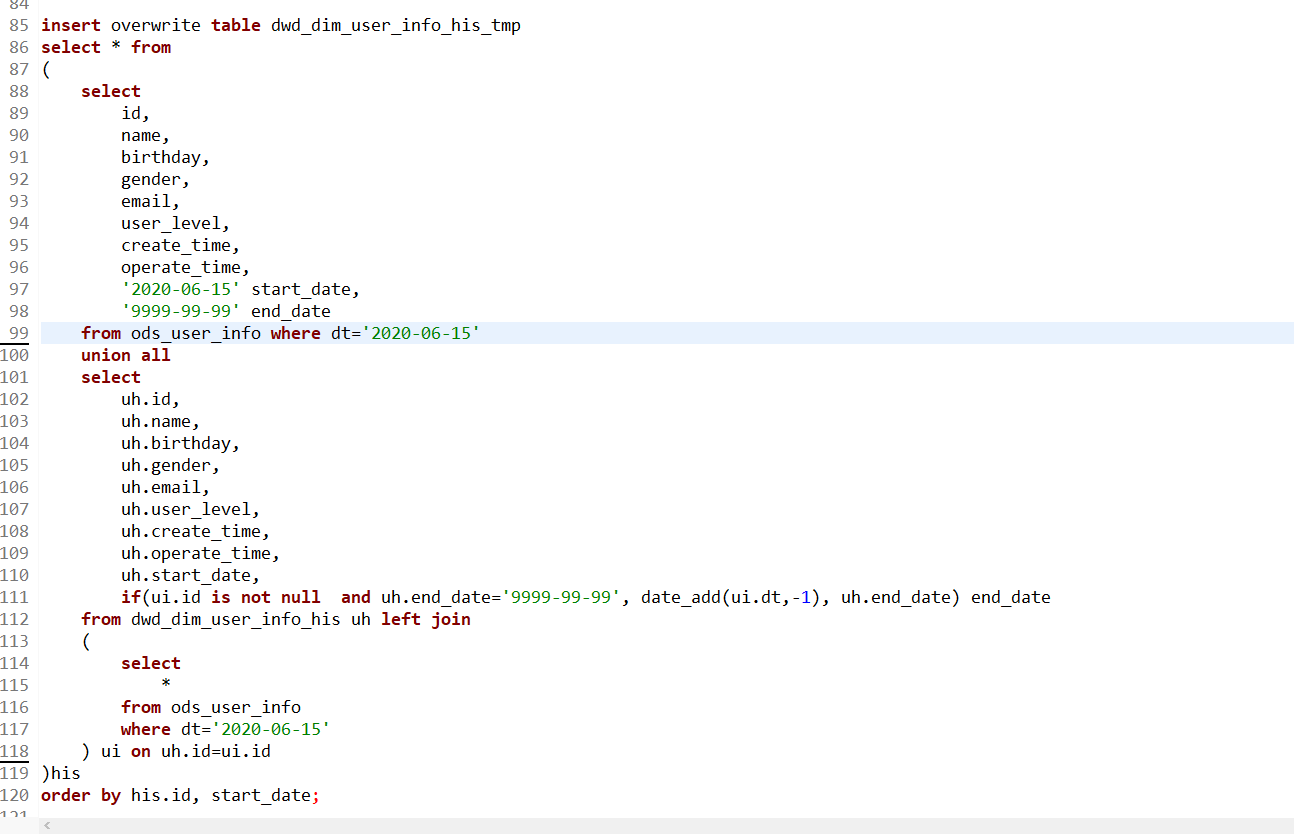
**方案一：**

问题：拉链表重复跑某一天数据错误

原始hql：

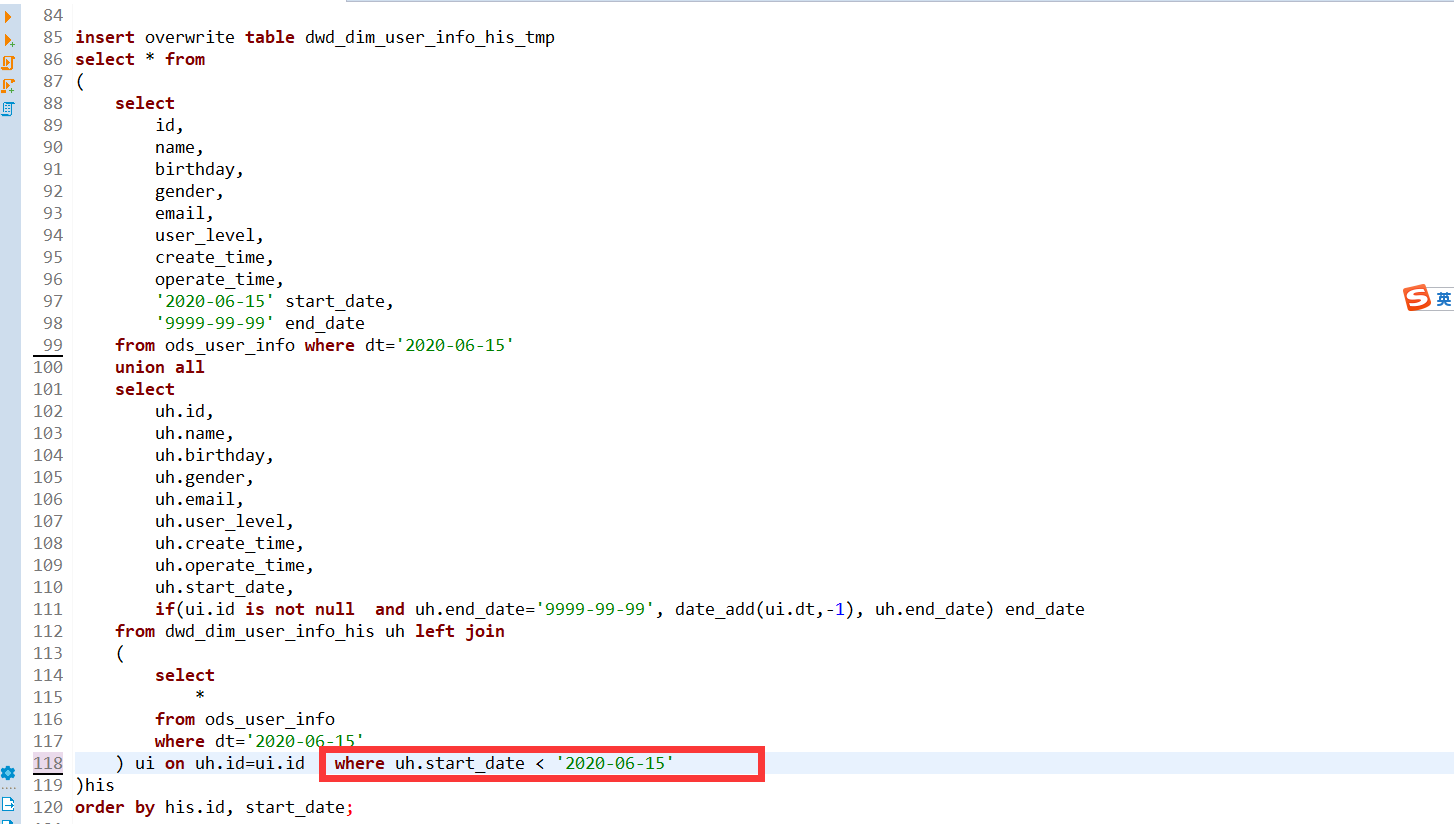


数据错误原因：

多次跑同一天数据，dwd\_dim\_user\_info\_his表中当天数据会被当做历史数据无差别更新

解决思路：

每次跑目标日数据时，dwd\_dim\_user\_info\_his表中数据只取非当前日期以前数据



目前看暂时是解决了，引发思考 如果跑的是历史中的某一天重跑 就不能用 “<”目标日期作为where条件，这样该天后的数据就会丢失而这里如果使用“!=” ，目标日数据的end\_date无法确定。 所以该思路如果用于重跑历史数据，需要从目标日一直跑到最新一天数据。

**方案二：**

张走走、:

insert overwrite table dwd\_dim\_user\_info\_his\_tmp

select \* from

(

select

id,

name,

birthday,

gender,

email,

user\_level,

create\_time,

operate\_time,

'2020-06-15' start\_date,

'9999-99-99' end\_date

from ods\_user\_info where dt='2020-06-15'

union

select

uh.id,

uh.name,

uh.birthday,

uh.gender,

uh.email,

uh.user\_level,

uh.create\_time,

uh.operate\_time,

uh.start\_date,

if(ui.id is not null and uh.end\_date='9999-99-99' and uh.start\_date !='2020-06-15', date\_add(ui.dt,-1), uh.end\_date) end\_date

from dwd\_dim\_user\_info\_his uh left join

(

select

\*

from ods\_user\_info

where dt='2020-06-15'

) ui on uh.id=ui.id

)his

order by his.id, start\_date;

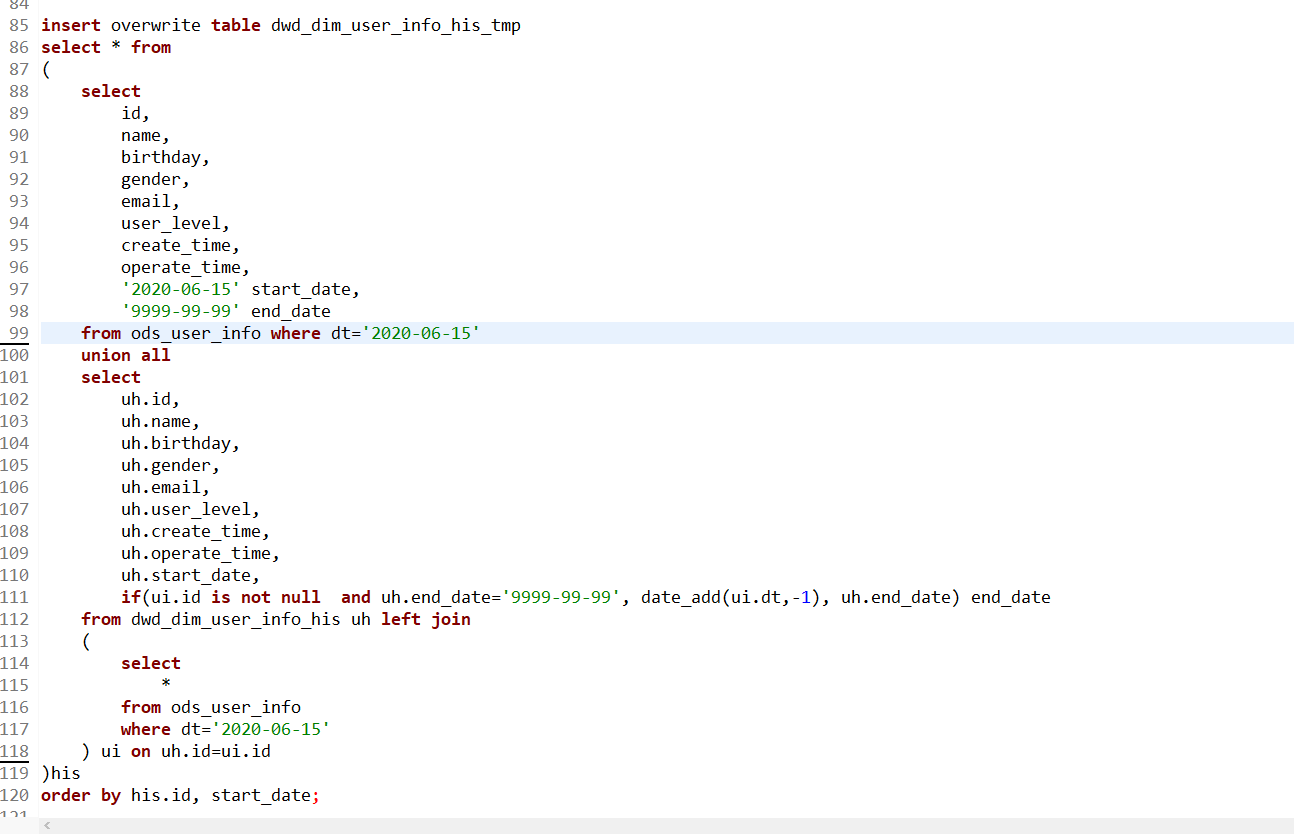
多加了个判断

然后union 去重 不用union all

**方案三**

问题：拉链表重复跑某一天数据错误

原始hql：

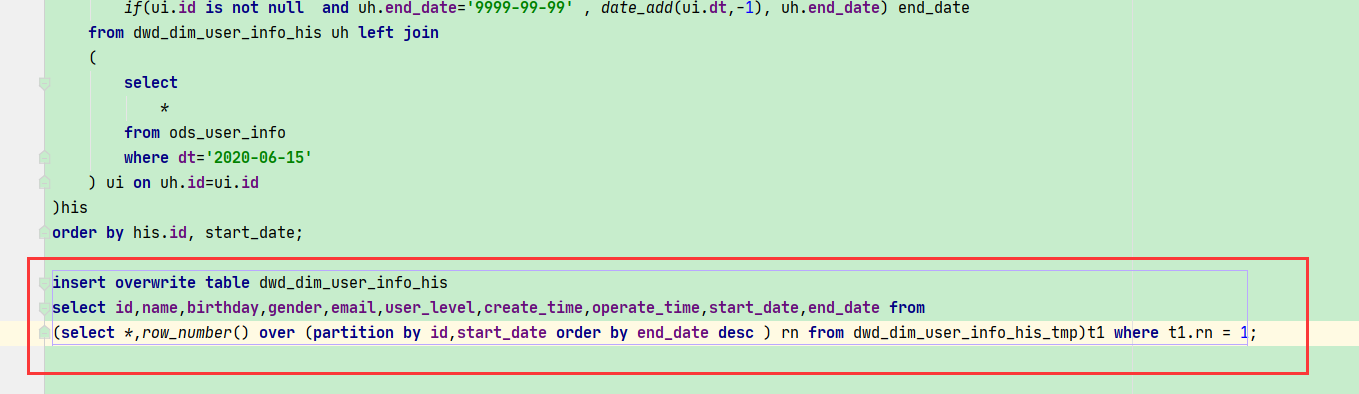


数据错误原因：

多次跑同一天数据，dwd\_dim\_user\_info\_his表中当天数据会被当做历史数据无差别更新

解决思路：

因为mysql中,同时每个user只会记录一条记录,所以同步到hive时, 相同user的start\_time是唯一的。所以同步到临时表时，有错误的数据无所谓。只需要利用开窗，只取一条的方法，去重处理即可。



目前看暂时是解决了，引发思考 如果跑的是历史中的某一天重跑 就不能用 “<”目标日期作为where条件，这样该天后的数据就会丢失而这里如果使用“!=” ，目标日数据的end\_date无法确定。 所以该思路如果用于重跑历史数据，需要从目标日一直跑到最新一天数据。

方案四：

达达:

海哥，这算一个吗，先取出每一个start\_date，再和要传入的时间进行判断，如果有，则退出程序，没有，程序继续运行

