查询及删除重复记录的SQL语句

1、查找表中多余的重复记录，重复记录是根据单个字段（Id）来判断

select \* from 表 where Id in (select Id from 表 group byId having count(Id) > 1)

2、删除表中多余的重复记录，重复记录是根据单个字段（Id）来判断，只留有rowid最小的记录

DELETE from 表 WHERE (id) IN ( SELECT id FROM 表 GROUP BY id HAVING COUNT(id) > 1) AND ROWID NOT IN (SELECT MIN(ROWID) FROM 表 GROUP BY id HAVING COUNT(\*) > 1);

3、查找表中多余的重复记录（多个字段）

select \* from 表 a where (a.Id,a.seq) in(select Id,seq from 表 group by Id,seq having count(\*) > 1)

4、删除表中多余的重复记录（多个字段），只留有rowid最小的记录

delete from 表 a where (a.Id,a.seq) in (select Id,seq from 表 group by Id,seq having count(\*) > 1) and rowid not in (select min(rowid) from 表 group by Id,seq having count(\*)>1)

5、查找表中多余的重复记录（多个字段），不包含rowid最小的记录

select \* from 表 a where (a.Id,a.seq) in (select Id,seq from 表 group by Id,seq having count(\*) > 1) and rowid not in (select min(rowid) from 表 group by Id,seq having count(\*)>1)

**一：重复数据根据单个字段进行判断**

1、首先，查询表中多余的数据，由关键字段（name）来查询。

select \* from OA\_ADDRESS\_BOOK where name in (select name from OA\_ADDRESS\_BOOK group by name having count(name)>1)

2、删除表中重复数据，重复数据是根据单个字段（Name）来判断，只留有rowid最小的记录

delete from OA\_ADDRESS\_BOOK where (Name) in

(select Name from OA\_ADDRESS\_BOOK group by Name having count(Name) >1)

and rowid not in (select min(rowid) from OA\_ADDRESS\_BOOK group by Name having count(Name)>1)

**二：重复数据根据多个字段进行判断**

1、首先，查询表中重复数据，由关键字段（Name,UNIT\_ID）来查询。

select \* from OA\_ADDRESS\_BOOK book1 where (book1.name,book1.unit\_id) in   
(select book2.name,book2.unit\_id from OA\_ADDRESS\_BOOK book2 group by  book2.name,book2.unit\_id  having count(\*)>1)

2、删除表中重复数据，重复数据是根据多个字段（Name,UNIT\_ID）来判断，只留有rowid最小的记录

delete from OA\_ADDRESS\_BOOK a where (a.Name,a.UNIT\_ID) in   
(select Name,UNIT\_ID from OA\_ADDRESS\_BOOK group by Name,UNIT\_ID having count(\*) > 1)   
and rowid not in (select min(rowid) from OA\_ADDRESS\_BOOK group by Name,UNIT\_ID having count(\*)>1)

3、查询表中重复数据，重复数据是根据多个字段（Name,UNIT\_ID）来判断，不包含rowid最小的记录

select name,unit\_id from OA\_ADDRESS\_BOOK a where (a.Name,a.UNIT\_ID) in   
(select Name,UNIT\_ID from OA\_ADDRESS\_BOOK group by Name,UNIT\_ID having count(\*) > 1)   
and rowid not in (select min(rowid) from OA\_ADDRESS\_BOOK group by Name,UNIT\_ID having count(\*)>1)

**1. 问题描述**

BBSCOMMENT表为BBSDETAIL的从表，记录商户评价信息。因为数据倒腾来倒腾去的，有很多重复数据。表结构如下：

COMMENT\_ID NOT NULL NUMBER　　--主键  
DETAIL\_ID NOT NULL NUMBER　　--外键，引用BBSDETAIL表  
COMMENT\_BODY NOT NULL VARCHAR2(500)　　--评价内容

--其它字段忽略

其中主键是没有重复的，重复的是DETAIL\_ID+COMMENT\_BODY+......等信息，就是某些商家的评价信息有重复。

**2. 解决步骤**

**2.1 查找表中多余的重复记录**

[复制代码](javascript:void(0);)

复制代码

--查询出所有有重复的数据

select DETAIL\_ID,COMMENT\_BODY,count(\*)

from BBSCOMMENT

group by DETAIL\_ID,COMMENT\_BODY

having count(\*)>1

order by DETAIL\_ID, COMMENT\_BODY; --1955条

复制代码

[复制代码](javascript:void(0);)

**2.2 显示了所有的非冗余的数据**

--这一条命令显示了所有的非冗余的数据

select min(COMMENT\_ID) as COMMENT\_ID,DETAIL\_ID,COMMENT\_BODY

from BBSCOMMENT

group by DETAIL\_ID,COMMENT\_BODY; --21453条，之所以此值不等于表总记录数-1955，是因为1955条记录中，有的重复了不止一次。

**2.3 如果记录数量少（千级别），可以把上面的语句做成子查询然后直接删除**

[复制代码](javascript:void(0);)

复制代码

--如果表数据量不是很大（1千条以内），可以把上面的语句做成子查询然后直接删除

delete from BBSCOMMENT where COMMENT\_ID not in(

select min(COMMENT\_ID)

from BBSCOMMENT

group by DETAIL\_ID,COMMENT\_BODY

); --782秒，在我这里，2万条记录，重复记录2千多（太慢了！！）

复制代码

[复制代码](javascript:void(0);)

**2.4 另一种删除方法**

[复制代码](javascript:void(0);)

复制代码

--这条语句也能够实现上述功能，但不好测试了，数据已经被我删除了

--删除条件一：有重复数据的记录；条件二：保留最小rowid的记录。

delete from BBSCOMMENT a

where

(a.DETAIL\_ID,a.COMMENT\_BODY) in(select DETAIL\_ID,COMMENT\_BODY from BBSCOMMENT group by DETAIL\_ID,COMMENT\_BODY having count(\*) > 1)

and rowid not in (select min(rowid) from BBSCOMMENT group by DETAIL\_ID,COMMENT\_BODY having count(\*)>1);

复制代码

[复制代码](javascript:void(0);)

**2.5 大数据量还是用PL/SQL方便快捷**

[复制代码](javascript:void(0);)

复制代码

declare

--定义存储结构

type bbscomment\_type is record

(

comment\_id BBSCOMMENT.COMMENT\_ID%type,

detail\_id BBSCOMMENT.DETAIL\_ID%type,

comment\_body BBSCOMMENT.COMMENT\_BODY%type

);

bbscomment\_record bbscomment\_type;

--可供比较的变量

v\_comment\_id BBSCOMMENT.COMMENT\_ID%type;

v\_detail\_id BBSCOMMENT.DETAIL\_ID%type;

v\_comment\_body BBSCOMMENT.COMMENT\_BODY%type;

--其它变量

v\_batch\_size integer := 5000;

v\_counter integer := 0;

cursor cur\_dupl is

--取出所有有重复的记录

select COMMENT\_ID, DETAIL\_ID, COMMENT\_BODY

from BBSCOMMENT

where(DETAIL\_ID, COMMENT\_BODY) in (

--这些记录有重复

select DETAIL\_ID, COMMENT\_BODY

from BBSCOMMENT

group by DETAIL\_ID, COMMENT\_BODY

having count(\*) > 1)

order by DETAIL\_ID, COMMENT\_BODY;

begin

for bbscomment\_record in cur\_dupl loop

if v\_detail\_id is null or (bbscomment\_record.detail\_id != v\_detail\_id or nvl(bbscomment\_record.comment\_body, ' ') != nvl(v\_comment\_body, ' ')) then

--首次进入、换记录了，都重新赋值

v\_detail\_id := bbscomment\_record.detail\_id;

v\_comment\_body := bbscomment\_record.comment\_body;

else

--其它记录删除

delete from BBSCOMMENT where COMMENT\_ID = bbscomment\_record.comment\_id;

v\_counter := v\_counter + 1;

if mod(v\_counter, v\_batch\_size) = 0 then

--每多少条提交一次

commit;

end if;

end if;

end loop;

if v\_counter > 0 then

--最后一次提交

commit;

end if;

dbms\_output.put\_line(to\_char(v\_counter)||'条记录被删除！');

exception

when others then

dbms\_output.put\_line('sqlerrm-->' ||sqlerrm);

rollback;

end;