

Oracle数据库删除重复记录的方法 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/269/2021_2022_Oracle_E6_95_B0_E6_c67_269820.htm

做项目的时候，一位同事导出数据的时候，不小心把一个表中的数据全都搞重了，也就是说，这个表里所有的记录都有一条重复的。这个表的数据是千万级的，而且是生产系统。也就是说，不能把所有的记录都删除，而且必须快速的把重复记录删掉。对此，总结了一下删除重复记录的方法，以及每种方法的优缺点。为了陈诉方便，假设表名为Tbl，表中有三列col1，col2，col3，其中col1，col2是主键，并且，col1，col2上加了索引。

- 1、通过创建临时表可以把数据先导入到一个临时表中，然后删除原表的数据，再把数据导回原表，SQL语句如下：

```
creat table tbl_tmp (0select distinct* from tbl).truncate table tbl.//清空表记录insert into tbl 0select * from tbl_tmp.//将临时表中的数据插回来。
```

这种方法可以实现需求，但是很明显，对于一个千万级记录的表，这种方法很慢，在生产系统中，这会给系统带来很大的开销，不可行。
- 2、利用rowid在oracle中，每一条记录都有一个rowid，rowid在整个数据库中是唯一的，rowid确定了每条记录是oracle中的哪一个数据文件、块、行上。在重复的记录中，可能所有列的内容都相同，但rowid不会相同。SQL语句如下：

```
0delete from tbl where rowid in (0select a.rowid from tbl a, tbl b where a.rowid>b.rowid and a.col1=b.col1 and a.col2 = b.col2)
```

如果已经知道每条记录只有一条重复的，这个sql语句适用。但是如果每条记录的重复记录有N条，这个N是未知的，就要考虑适用下面这种方法了。
- 3、利用max或min函数这里也要使

用rowid，与上面不同的是结合max或min函数来实现。SQL语句如下

```
delete from tbl a where rowid not in (select max(b.rowid)
from tbl b where a.col1=b.col1 and a.col2 = b.col2).//这里max使用min也可以 或者用下面的语句
delete from tbl a where
rowid from tbl b where a.col1=b.col1 and a.col2 = b.col2).//这里如果把max换成min的话，前面的where子句中需要把"" 跟上面的方法思路基本是一样的，不过使用了group by，减少了显性的比较条件，提高效率。SQL语句如下：
```

```
delete from tbl where
rowid not in (select max(rowid) from tbl t group by t.col1,
t.col2).
delete from tbl where (col1, col2) in (select col1, col2 from
tbl group by col1, col2 having count(*) > 1) and
rowid not in (select n in (rowid) from tbl group
by col1, col2 having count(*) > 1)
```

还有一种方法，对于表中有重复记录的记录比较少的，并且有索引的情况，比较适用。假定col1，col2上有索引，并且tbl表中有重复记录的记录比较少，SQL语句如下

4、利用group by，提高效率

4、Group By方法（我经常所用的一种方法）

一、查数据

```
Select count(Num), max(Name) from student --列出重复记录，并列出属性
Group By Name Having count(Num) > 1 --按Name分组后找出表中Num列重复，即出现次数大于1
```

二、删数据

```
Delete from student
Group By Name Having count(Num) > 1
```

100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com