

如何提取出并删除数据库中重复的记录？

查询出数据表中重复的记录，应该来说用过分组查寻的，这道题并不在话下，我们先来看看这张表

```
select * from repeat|
```

	userName	userPassword	loginTimes
1	fengyan	fengyan	5
2	fengyan	fengyan	8
3	fengyan	fengyan	7
4	a	a	5
5	a	a	5
6	b	b	5
7	a	b	5
8	fengyan	fengyan	7

上图高亮部分的是重复的数据行，那么如何取出其中高亮的部分，

聚合函数

在解决这个问题前，我们必需先讲

讲sql语言中一种特殊的函数:聚合函数，例如SUM, COUNT, MAX, AVG等。

这些函数和其它函数的根本区别就是它们一般作用在多条记录上。例如上函数从左至右依次为 求总和，记录数，最大

值，

平均值！

关于它们的作用相

信很多朋友也知道！但也有可能在自

己知道的情况下，并不了解它们的学名，聚合函数。这也是我开篇写这个的目的！

另外就是group by，分组查寻。看下面结果

```
select userName,loginTimes from repeat  
select userName ,sum(loginTimes) from repeat group by userName|
```

	userName	loginTimes
1	fengyan	5
2	fengyan	8
3	fengyan	7
4	a	5
5	a	5
6	b	5
7	a	5
8	fengyan	7

	userName	(无列名)
1	a	15
2	b	5
3	fengyan	27

www.fwxt.cn

通过使用GROUP BY子句，可以让SUM 这样聚合函数对属于一组的数据起作用。当我们上面指定 GROUP BY userName时，属于同一个u

serName(用户)的一组数据将只能返回一行值，也就是说，表中所有除userName(用户)外的字段，只能通过 SUM, COUNT等聚合函数运算后返回一个值。假使我们在第二句SQL语句中添加一列,userPassword时，就会报如下错误

```
select userName ,userPassword,sum(loginTimes) from repeat group by userName
```

服务器：消息 8120，级别 16，状态 1，行 1
列 'repeat.userPassword' 在选择列表中无效，因为该列既不包含在聚合函数中，也不包含在 GROUP BY 子句中。

这应该是初学者经常遇见的错误，据上已经分析过，出现此种错误是必然，因为，我们的结果是已经分组过的，所以结果会按姓名分组，相同的姓名只返回一行结果，但并没有指定也按密码分，所以当返回姓名重复记录的时候，并不知道返回的这一行密码取哪个？因为我们没有对密码聚合，或分组，所以报错是因为密码按此逻辑，无法取到！

那么有朋友可以会想到那么group by

后面再加上密码会怎样，可以试试。学计算机，有了想法就应该去尝试，大胆尝试。电脑也试不坏！怕什么！

```
select userName,loginTimes from repeat  
|  
select userName ,userPassword,sum(loginTimes) from repeat  
group by userName,userPassword
```

	userName	loginTimes	
1	fengyan	5	
2	fengyan	8	
3	fengyan	7	
4	a	5	
5	a	5	
6	b	5	
7	a	5	
8	fengyan	7	

	userName	userPassword	(无列名)
1	a	a	10
2	a	b	5
3	b	b	5
4	fengyan	fengyan	27

www.fwxt.cn

此时程序有了结果。对比之前的图，我们可以发现结果集不一样了，多了一行

a 15

变成了

a a 10

a b 5

想想原因，因为我们现在

分组多加了密码，所以也会按姓名和密码都相同的分！因为姓名都为a的用户有二个密码为a一个密码为b，所以在按密码分时，会继续拆分！也就出现如上结果！

Having

那么Having又是有什么作用的？

HAVING子句主要就是在聚合后对组记录进行筛选。类似where，但其与where又是不同的！

让我们还是通过具体的实例来理解GROUP BY 和 HAVING 子句

```
select userName ,userPassword, sum(loginTimes) from repeat
group by userName, userPassword
select userName ,userPassword, sum(loginTimes) from repeat
group by userName, userPassword having sum(loginTimes)>5
```

	userName	userPassword	(无列名)
1	a	a	10
2	a	b	5
3	b	b	5
4	fengyan	fengyan	27

	userName	userPassword	(无列名)
1	a	a	10
2	fengyan	fengyan	27

www.fwxt.cn

第二句，也即将登录总次数大于5的记录查出！

注意在这里，我们不能使用where,因为用where来限定条件，需要符合条件的列必须在表中存在！而这里的取出总登录次数大于5，表中本没

有这样一列，所以无法使用where来查询，而必须使

用having,了解这个，我们就应该知道where与having的差别了！

HAVING是对由sum或其它聚合函数运算结果的输出进行限制。

那么在了解了上面的知识后，我们如何再来解决开篇提到的问题，也即如何取出表中重复的记录？

我们要查出重复的记录，那么必然要用到分组查询，group by。分组的依据为表中各列。这样它会将各列都相同的分为一组，但同时我们要查询出重复的记录，因此必然需要相同的记录至少应该是2条以上。由此可知。我们需要对group by的结果进行限定，条件是记录数大于1的，因此我们联想到having。因此有了如下解决方案：

```
select userName as '用户名', userPassword as '密码', loginTimes as '登录次数', count(*) as '重复数'
from repeat group by userName, userPassword, loginTimes having count(*) > 1
```

	用户名	密码	登录次数	重复数
1	a	a	5	2
2	fengyan	fengyan	7	2

www.fwxt.cn

那么如果同时存在where及having是什么样的呢？看下图：

```
select userName ,userPassword, sum(loginTimes) from repeat
group by userName, userPassword having sum(loginTimes)>5

select userName ,userPassword, sum(loginTimes) from repeat
where userName='a' --注意where加的地方
group by userName, userPassword having sum(loginTimes)>5
```

	userName	userPassword	(无列名)
1	a	a	10
2	fengyan	fengyan	27

	userName	userPassword	(无列名)
1	a	a	10

www.fwxt.cn

还有另外一个相似的问题就是如何删除数据库中重复的记录！

这个问题应该有几种解决方案，比较常用的是使用临时表，如下：

—使用distinct关键字查出去重后的记录，并将结果存储在临时表中

```
select distinct * into #temp from repeat
```

—删除 repeat 表

```
delete repeat
```

—再从临时表中取出所有结果放回repeat表中

```
insert repeat select * from #temp
```

—删除临时表

```
drop table #temp
```

	userName	userPassword	loginTimes
1	a	a	5
2	a	b	5
3	b	b	5
4	fengyan	fengyan	5
5	fengyan	fengyan	www.fwxt.cn
6	fengyan	fengyan	8

操作之后的结果与上，去掉了重复的行我用高亮做了标记。也即使用distinct关键字去重，是去掉所有列都相同的，而对于userName及userPassword相同的，它没有处理出来，现在假设要将姓名与密码都相同的也去掉，如何处理。看下面：

```
select * from repeat a
```

```
where
```

```
(a.userName ) in (select userName from repeat group by userName,userPassword having count(*) > 1)
```

```
and
```

```
(a.userPassword ) in (select userPassword from repeat group by userName,userPassword having count(*) > 1)
```

	userName	userPassword	loginTimes
1	fengyan	fengyan	5
2	fengyan	fengyan	8
3	fengyan	fengyan	7
4	a	a	5
5	a	a	5
6	fengyan	fengyan	7

这样取出的是用户名与密码相同的记录！将select改为delete即可删除，当然，这样删除就全部删除了，如果想要留一条，还可在后面加限定条件，来决定留下哪一条！

悠久防伪一直专注于防伪溯源解决方案。

本文地址: <https://www.fwxt.cn/en/study-e13.html>