

groupby分组的应用 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/137/2021_2022_groupby_E5_88_86_c97_137602.htm 第五节 group by分组的应用首先先说说分组是怎样工作的。举个例子：表：temp1有以下记录

：bhslaaa1aaa4ccc2bbb5aaa9bbb7现在要统计一下temp1有中多少种编号，各种编号的总数量又是多少。很明显，这是使用分组。sele bh,sum(sl) as total_sl . from temp1 . grou by bh . orde by bh . into curs temp2命令的运行过程就象投票选举中的点票过程一样。它中间到底是采用什么样的技术来储存中间结果我不知道，但不这重要，我现在假设它使用临时表(实际上我想不大可能)，这样对结果没影响，但又容易理解。1、是逐个扫描记录。每遇到一个新的编号，就为这个编号建立一个临时表(再次重申，使用临时表只是我的假设布局，真正的处理方法我不知道)，然后把这个记录的内容放进对应的临时表中。如果不是新编号，就直接把记录内容放进对应的临时表中。2、全部记录扫描完了，各种编号对应的临时表也产生了，现在就是对每一个临时表采用sum()进行统计了。比如编号为"aaa"的临时表是cursor1，那它的内容就是

：Aaa1Aaa4Aaa9现在进行sum()统计了，它的效果就相当于sele bh,sum(sl) as total_sl from cursor1。每一个编号的临时表都要进行这样的sum()统计，然后把每个编号的统计结果联合起来，就想使用union一样，最后得出的结果就是

：aaa14ccc2bbb123、然后再给bh排序，结果就是

：aaa14bbb12ccc2看了上面那个例子，再看下面这条命令，想必也可以理解了吧：sele

cpk.cpbh, cpk.cpmc, sum(iif(isnull(thdmx.thsl), 0, thdmx.thsl) as total_thsl . from cpk left join thdmx . on cpk.cpbh=thdmx.cpbh . grou by cpk.cpbh . orde by cpk.cpbh . into curs temp1 产品表左联接提货明细表，这样没有提过的产品，在联接后的临时表中，对应的提货数量就是个null值。到了分组，第一步完成后，生成的临时表中也仍然有null值，所以在第二步进行sum命令时，就要使用iif(isnull())来过滤了。过滤的结果就是把null值改为0，其它的不变，然后再由sum进行汇总。最后第三步再排序，结果就出来了。以上是分组的最基本用法。现在说说分组的一些古怪用法。一、使用recn()。例子一：有种情况，一大段本来是应该连续的号码，但中间地漏掉了一些号码，现在要把这些漏了的号码找出来。恐怕很多人，一来就是使用scan endscan，再加上一些中间变量以作为判断依据，最后写成的代码有一大段，运行起来慢得要死。但如果采用以下这种方法，我想即使它不是最好，但也是很好的了。数据如下：1、6、3、5、2、7、9、10、15、16在这些数据当中，要找出漏掉了的4、8、11、12、13、14。1、首先排序sele bh from temp1 orde by bh into curs temp1结果是：1、2、3、5、6、7、9、10、15、16如果源表中的物理记录顺序和号码的顺序一样就不用做这一步了。2、取temp1中各个记录的recn()，即记录号，然后把编号减去记录号sele bh, val(bh-recn()) as aa from temp1 into curs temp2结果就是

：Bhaa102030516171921021561663、对aa使用分组，取每组中最小的bh和最大的bhsele min(bh) as minbh, max(bh) as maxbh from temp2 grou by aa into curs temp3结果如下

：minbhmaxbh135791015164、现在结果很明显了，某个记录

的minbh跟它上面那个记录的maxbh中间相差的，就是漏掉的号码。现在才使用scan endsfan就容易了。以上方法适用于数据量大、第三步的结果比源表少很多记录的情况下才会发挥效果。因为scan的速度不能和Select - SQL相比，何况它还要作条件判断。虽然这种方法使用了二条全遍历的SQL命令，但这是没过滤条件的SQL命令，速度是很快的，只是分组要多点时间而已。同时只适用于编号没有重复的情况。如果编号有重复，那在第一步的时候，就要使用：sele bh from temp1 goru by bh orde by bh into curs temp1把多余的编号去掉。如果了解那些编号是重复的，就可以使用：sele bh,coun(bh) as aa from temp1 grou by bh orde by bh havi aa>1 into curs temp1结果就是重复的编号，对应的aa字段就是重复的次数。在第二步，好象可以不用二条命令就可以得出结果了，只是暂时还没想到，不知各位对些有什么看法。上面这种方法有个缺点：就是使用了sum()函数，而它是只能统计数值型的数据，其它字符型、日期型等是不能统计的。所以如果想把一个表中所有记录的产品名称都串起来，变成一个字符串，那是不能的。但对于日期型，变通一下，有时还是可以使用的。二、日期相减的结果是数值，从而可以使用sum()函数。例子二编号名称 入库日期
1网卡2000.09.02
1网卡2000.09.03
1网卡2000.09.05
2vfp2000.09.05
2vfp2000.09.06
说明：把编号和名称相同，入库日期相间为1的记录合并为一条，并加入另一个表里，如结果：编号 名称 入库日期（表2和表1结构一样）
1网卡2000.09.03
1网卡2000.09.05
2vfp2000.09.06
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