## GROUP BY、Having和where的作用

 (2008-12-15 00:04:05)

[IMG_256转载▼](http://blog.sina.com.cn/s/javascript:;)

|  |  |
| --- | --- |
| 标签： [杂谈](http://search.sina.com.cn/?c=blog&q=%D4%D3%CC%B8&by=tag" \t "http://blog.sina.com.cn/s/_blank) |  |

****SQL****是结构化查询语言(Structured Query Language)的缩写。这种语言允许我们对****数据库****进行复杂的操作。SQL语言的使用范围非常广泛。许多数据库产品都支持SQL语言，这意味着如果我们学会了SQL语言，我们可以把这种知识运用到MS Access 或 SQL Server, ****Oracle****, DB2以及非常多的其它数据库中。

　　SQL语言运用在关系型数据库中。一个关系型数据库把数据存储在表（也称关系）中。每个数据库的主要组成就是一组表。每个表又由一组记录组成－－每条记录在表中有相同的结构，包含固定数量的具有一定类型的字段。  
　　  
　　下面我们来看一个实际的数据库中的表。该表的表名为cia，包含250多条记录，每个记录代表一个国家。表由5个字段组成，字段的值有的是****字符串****类型，有的是数字类型。

name region area population gdp  
---- ------ ------ ---------- -----------  
Yemen Middle East 527970 14728474 23400000000  
Zaire Africa 2345410 44060636 18800000000  
Zambia Africa 752610 9445723 7900000000  
Zimbabwe Africa 390580 11139961 17400000000

　　下面我们可以用一些SQL语句来查询这个表中我们该兴趣的数据。

　　1. 中国的GDP是多少？

　　查询用的SQL语句为：

select gdp from cia where name='china'

　　查询结果为：

4800000000000

　　2. 给出每个地区的国家数和人口总数。并且按地区的人口数从多到少排序。

　　查询用的SQL语句为：

SELECT region, COUNT(name), SUM(population)  
FROM cia  
GROUP BY region  
ORDER BY 3 DESC

　　查询结果为：

region COUNT(name) SUM(population)  
------ ----------- ---------------  
Asia 14 2963031109  
Africa 59 793382933  
Europe 43 580590872  
....

　　怎么样，对SQL语言有了基本的了解了吧，同时对数据库，表，记录，字段等一系列在SQL语言中常用的感念也有大概的认识吧。如果不是很清楚也没关系，在接下来的内容中我们从SQL语言中最简单的内容逐步给大家作介绍，并提供丰富的练习让大家实际操作。相信学完本系列****教程****，你可以成为一个SQL语言的高手。

在介绍GROUP BY 和 HAVING 子句前，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。

SELECT SUM(population) FROM bbc

　　这里的SUM作用在所有返回记录的population字段上，结果就是该查询只返回一个结果，即所有国家的总人口数。

　 　通过使用GROUP BY 子句，可以让SUM 和 COUNT 这些函数对属于一组的数据起作用。当你指定 GROUP BY region 时， 属于同一个region（地区）的一组数据将只能返回一行值，也就是说，表中所有除region（地区）外的字段，只能通过 SUM, COUNT等聚合函数运算后返回一个值。

　　HAVING子句可以让我们筛选成组后的各组数据，WHERE子句在聚合前先筛选记录．也就是说作用在GROUP BY 子句和HAVING子句前．  
而 HAVING子句在聚合后对组记录进行筛选。

　　让我们还是通过具体的实例来理解GROUP BY 和 HAVING 子句，还采用第三节介绍的bbc表。

　　SQL实例：

　　一、显示每个地区的总人口数和总面积：

SELECT region, SUM(population), SUM(area)  
FROM bbc  
GROUP BY region

　　先以region把返回记录分成多个组，这就是GROUP BY的字面含义。分完组后，然后用聚合函数对每组中的不同字段（一或多条记录）作运算。

　　二、 显示每个地区的总人口数和总面积．仅显示那些面积超过1000000的地区。

SELECT region, SUM(population), SUM(area)  
FROM bbc  
GROUP BY region  
HAVING SUM(area)>1000000

　　在这里，我们不能用where来筛选超过1000000的地区，因为表中不存在这样一条记录。

　　相反，HAVING子句可以让我们筛选成组后的各组数据．

－－－－－－－－－－－－－－－－

**GROUP BY就是把某一列中，相同的字段进行分组**，

如  
一个客户买东西的记录表；  
  
客户，产品名称，价格，  
  
我们要计算出某一个客户的总销费额，就是不客他买什么东西，只计算他花多少钱，如  
  
客户，产品名称，价格  
张三，电视，2000  
李四，洗衣机，800  
张三，电冰箱，4500  
  
我们看到，张三有两个记录，李四有一个，我们可以这样得出上面的需求  
select 客户,sum(价格) as '消费额' from 表名称 group by 客户  
  
将返回  
客户，消费额  
张三，6500  
李四，800

－－－－－－－－－－－－－－－－－－－－－－  
  
**所谓元组，就是记录。WHERE先筛选记录，然后对记录进行分组；HAVING是先分组记录，然后对分组以后的每个组进行筛选。**  
  
比如教师（姓名（文本，8），职称（文本，8），性别（文本，2））这个表  
  
我可以使用WHERE 性别="男" GROUP by 职称    的子句列示出所有男性教师的职称分组情况。  
注意，这里的分组中的统计值只有男教师。  
当使用HAVING count（姓名）>8  group  by  职称    的子句，将会先对所有记录按照职称字段进行分组，然后显示出人数大于8个的那个组  
这个就是HAVING和WHERE的区别。一个是先分组，再对组进行筛选；一个是先筛选，然后对筛选后的记录进行分组。

select 系号, count(\*);  
from 表1 ;  
   group by 系号;  
   where 性别="女"  
  
select 系号, count(\*);  
from 表1 ;  
   group by 系号;  
    having 性别="女"  
这两个sql运行的结果不一样的.  
第一个是正确的.  
第二个是错误的.  
因为一个系有男有女的,你这个分组条件是永远不起作用.  
只要一个系中有一个女的,就会满足条件统计整个系的人数.  
同时统计人数时候,也包括男在内.  
having只不过是排除不满足having后的条件的某个分组.  
假如一个系中无女,那么这个系不统计人数.  
而where是把男生记录先排除了.在统计.  
不用where统计人数也包括在内,达不到你的目的.  
where是对整个表的数据进行筛选,  
haing是对分组筛选,只要满组条件该组会出现在结果的.