

数据库中group by和having语法使用详解

有个朋友问我一个返话费的问题，大概意思是这样的：只需把表deal中所有手机用户某天充值两次以上且总金额超过50的用户充值记录查询出来，至于怎么进行返话费那不是重点。

先看看group by的语法：

```
SELECT column1, SUM(column2) FROM "list-of-tables" GROUP BY "column-list";
```

GROUP BY子句将集中所有的行在一起，它包含了指定列的数据以及允许合计函数来计算一个或者多个列。

假设我们将从员工表employee表中搜索每个部门中工资最高的薪水，可以使用以下的SQL语句：

```
SELECT max(salary), dept FROM employee GROUP BY dept;
```

这条语句将在每一个单独的部门中选择工资最高的工资，结果将他们的salary和dept返回。

group by 顾名思义就是按照xxx进行分组，它必须有“聚合函数”来配合才能使用，使用时至少需要一个分组标识字段。

聚合函数有：sum()、count()、avg()等，使用group by目的就是要将数据分组进行汇总操作。

例如对员工表的操作：

```
select dept_id,count(emp_id),sum(salary) form employee group by dept_id;
```

这样的运行结果就是以“dept_id”为分类标志统计各单位的职工人数和工资总额。

再看看having的语法：

```
SELECT column1, SUM(column2) FROM "list-of-tables" GROUP BY "column-list"  
HAVING "condition";
```

这个HAVING子句的作用就是为每一个组指定条件，像where指定条件一样，也就是说，可以根据你指定的条件来选择行。如果你要使用HAVING子句的话，它必须处在GROUP BY子句之后。

例如还是对员工表的操作：

```
SELECT dept_id, avg(sal) FROM employee GROUP BY dept_id HAVING avg(salary)
>= 4000;
```

这样的运行结果就是以“dept_id”为分类标志统计各单位的职工人数和工资平均数且工资平均数大于4000。

下面开始我们的返话费查询功能的实现：

话费表deal字段有这些：sell_no:订单编号 name:用户名 phone:用户手机号 amount:充值金额 date:充值日期

上边就这些有效字段，假如数据（数据纯属虚构，如有*，纯是巧合）如下：

sell_no	name	phone	amount	date
00000000001	李晓红	15822533496	50	2011-10-23
08:09:23				
00000000002	李晓红	15822533496	60	2011-10-24
08:15:34				
00000000003	李晓红	15822533496	30	2011-10-24
12:20:56				
00000000004	杨 轩	18200000000	100	2011-10-24
07:59:43				
00000000005	杨 轩	18200000000	200	2011-10-24
10:11:11				
00000000006	柳梦璃	18211111111	50	2011-10-24
09:09:46				
00000000007	韩菱纱	18222222222	50	2011-10-24
08:09:45				
00000000008	云天河	18333333333	50	2011-10-24
08:09:25				

把以上数据当天(2011-10-24)交过两次话费，而且总金额大于50的数据取出来，要取的结果如下：

00000000002	李晓红	15822533496	60	2011-10-24
08:15:34				
00000000003	李晓红	15822533496	30	2011-10-24
12:20:56				
00000000004	杨 轩	18200000000	100	2011-10-24
07:59:43				
00000000005	杨 轩	18200000000	200	2011-10-24
10:11:11				

因为今天(2011-10-24)李晓红和杨轩交过两次以上话费，而且总金额大于50，所以有他们的数据，而柳梦璃，韩菱纱，云天河只交过一次，所以没他们的数据。

我的处理思路大概是这样的，先把当天日期的记录用group by进行手机号分组即一个手机号为一组，接着用having子句进行过滤，把交过两次话费且话费总金额大于50的手机号查出来，最后用手机号和日期条件组合查询就能完成数据的查询，具体如下。

注意日期处理细节，要查询的某一天(yyyy-MM-dd)的所有记录mysql是这样处理的：

```
SELECT date_format(date, '%Y-%m-%d') from deal;
```

查询出符合条件（交过两次以上话费，而且总金额大于50）的手机号：

```
select phone from deal where date_format(date, '%Y-%m-%d')="2011-10-24"
group by phone having count(phone)>1 and sum(amount)>50;
```

结合手机号和日期查询出最终记录：

```
select * from deal where date_format(date, '%Y-%m-%d')="2011-10-24" and
phone in
(select phone from deal where date_format(date, '%Y-%m-%d')="2011-10-24"
group by phone having count(phone)>1 and sum(amount)>50) order by phone;
```

里边嵌套了一个select语句，感觉效率低点了，谁有更高效的方法不？

附数据建库sql代码：

```
create database if not exists `phone_deal`;

USE `phone_deal`;

DROP TABLE IF EXISTS `deal`;

CREATE TABLE `deal` (
  `sell_no` varchar(100) NOT NULL,
  `name` varchar(100) default NULL,
  `phone` varchar(100) default NULL,
  `amount` decimal(10,0) default NULL,
  `date` datetime default NULL,
  PRIMARY KEY (`sell_no`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

insert into `deal`(`sell_no`,`name`,`phone`,`amount`,`date`) values
('00001','李晓红','15822533496','60','2011-10-23 08:09:23'),
('00002','李晓红','15822533496','50','2011-10-24 08:15:34'),
('00003','李晓红','15822533496','40','2011-10-24 12:20:56'),
('00004','杨轩','18210607179','100','2011-10-24 07:59:43');
```

```
('00005','杨轩','18210607179','50','2011-10-24 10:11:11'),  
('00006','柳梦璃','15822533492','1000','2011-10-24 09:09:46'),  
('00007','韩菱纱','15822533493','10000','2011-10-24 08:09:45'),  
('00008','云天河','15822533494','500','2011-10-24 08:09:25');
```