

转

关于group by的用法 原理

2017年02月08日 11:03:30

紫荆岁月Pawn

阅读数：31526

写在前面的话：用了好久group by，今天早上一觉醒来，突然感觉group by好陌生，总有个筋别不过来，为什么不能够select * from Table group by id，为什么一定需要一个列或者某个列的聚合函数，group by 多个字段可以怎么去很好的理解呢？不过最后还是转过来了，简单写写吧，大牛们直接略过吧。

=====正文开始=====

先来看下表1，表名为test：

22

24

<

>

id	name
1	aa
2	aa
3	bb
4	bb
5	cc
6	dd
7	ee
8	bb
9	cc

表1

执行如下SQL语句：

1	SELECT name FROM test
2	GROUP BY name

你应该很容易知道运行的结果，没错，就是下表2：

表2

可是为了能够更好的理解“group by”多个列“和”聚合函数“的应用，我建议在思考的过程中，由表1到表2的过程中，增加一个虚构的中间表：**虚拟表3**。下面说说如SQL语句执行情况：

- 1.FROM test：该句执行后，应该结果和表1一样，就是原来的表。
- 2.FROM test Group BY name：该句执行后，我们想象生成了**虚拟表3**，如下所图所示，生成过程是这样的：group by name，那么找name那一列，具有相同name值一行，如对于name值为aa的，那么<1 aa 2>与<2 aa 3>两行合并成1行，所有的id值和number值写到一个单元格里面。

id	name	number	id	name	number
1	aa	2	1	aa	2
2	aa	3	2	aa	3
3	bb	4	3	bb	4
4	bb	5	4	bb	5
5	cc	6	5	cc	6
6	dd	7	6	dd	7
7	ee	7	7	ee	7
8	bb	5			
9	cc	6			

表一

虚拟表3

👍
22

💬
24

🔖

📱

<

>

- 3.接下来就要针对**虚拟表3**执行Select语句了：
- (1) 如果执行select *的话，那么返回的结果应该是虚拟表3，可是id和number中有的单元格里面的内容是多个值的，而关系数据库就是基于关系的，单元格中是不允许以你看，执行select * 语句就报错了。
 - (2) 我们再看name列，每个单元格只有一个数据，所以我们select name的话，就没有问题了。为什么name列每个单元格只有一个值呢，因为我们就是用name列来gr
 - (3) 那么对于id和number里面的单元格有多个数据的情况怎么办呢？答案就是用**聚合函数**，**聚合函数就用来输入多个数据，输出一个数据的**。如cout(id)，sum(number) 合函数的输入就是每一个多数据的单元格。
 - (4) 例如我们执行select name,sum(number) from test group by name，那么sum就对虚拟表3的number列的每个单元格进行sum操作，例如对name为aa的那一行执行sum操作，即2+3，返回5，最后执行结果如下：

	na
1	aa
2	bb
3	cc
4	dd
5	ee

- (5) group by 多个字段该怎么理解呢：如group by name,number，我们可以把name和number 看成一个**整体字段**，以他们整体来进行分组的。如下图

id	name	number	id	name	r
1	aa	2	1	aa	
2	aa	3	2	aa	
3	bb	4	3	bb	
4	bb	5	4	bb	
5	cc	6	5	cc	
6	dd	7	6	dd	
7	ee	7	7	ee	
8	bb	5			
9	cc	6			

- (6) 接下来就可以配合select和聚合函数进行操作了。如执行select name,sum(id) from test group by name,number，结果如下图：

	na
1	aa
2	aa
3	bb
4	bb
5	cc
6	dd
7	ee

至此，我已经对我自己对如此简单的问题有如此天马行空的想法所折服，洗洗睡觉。

(完)