

第 1 题(本题2分): 考虑以下4个查询 Q1: Select a,b From T1, T2 Where T1.c=T2.c And T1.c=100; Q2: Select a,b From T1, T2 Where T1.c=100 And T2.c=100; Q3: Select a,b From T1, T2 Where T1.c=T2.c And T1.c>100; Q4: Select a,b From T1, T2 Where T1.c>100 And T2.c>100; 说法正确的是:

- ☒ A: Q1和Q2结果相同; Q3和Q4结果不同; ✓
- ☐ B: Q1和Q2结果不同; Q3和Q4结果相同; ✗
- ☐ C: Q1和Q2结果相同; Q3和Q4结果相同; ✗
- ☐ D: 4个查询结果均不同。 ✗

第 2 题(本题2分): 在学生表Student(s_no, s_name, birthday, gender)和学生选课表SC(s_no, c_no, grade)中求每个学生的姓名和平均成绩。哪个查询表达正确?

- ☐ A: Select s_name, AVG(grade) From Student, SC Where Student.s_no=SC.s_no Group By Student.s_no ✗
- ☐ B: Select s_name, AVG(grade) From Student, SC Where Student.s_no=SC.s_no Group By SC.s_no ✗
- ☐ C: Select s_name, AVG(grade) From Student, SC Group By SC.s_no ✗
- ☒ D: Select s_name, AVG(grade) From Student, SC Where Student.s_no=SC.s_no Group By SC.s_no, s_name ✓

第 3 题(本题2分): 以下哪些查询是等价的?

- ☐ A: Select Count(*) From T Where b>100 Group By a 和 Select Count(*) From T Group By a ✗
- ☐ B: Select Count(*) From T Where b>100 Group By a 和 Select Count(*) From T Group By a Having AVG(b)>100 ✗
- ☒ C: Select AVG(a) From T Where a<10 Group By a 和 Select AVG(a) From T Group By a Having AVG(a)<10 ✓
- ☐ D: Select a,AVG(b) From T Where b>100 Group By a 和 Select a,AVG(b) From T Group By a Having AVG(b)>10 ✗

选项C是正确的。这两个查询都会返回`a`值小于10的平均值，只是它们使用了不同的方法来实现这一点。

- 查询C1: `Select AVG(a) From T Where a<10 Group By a`
- 查询C2: `Select AVG(a) From T Group By a Having AVG(a)<10`

在查询C1中，首先通过`Where a<10`过滤出所有`a`值小于10的行，然后对每个不同的`a`值计算平均值。由于`Group By a`的存在，每个组只包含具有相同`a`值的行，而这个`a`值已经通过`Where`子句保证是小于10的。因此，每个组的平均`a`值实际上就是`a`值本身。

在查询C2中，没有初始过滤条件，所以会为表T中每个不同的`a`值计算一个平均值。然后，通过`Having AVG(a)<10`子句过滤掉所有组的平均`a`值不小于10的组。在这里，由于每个组都是由相同的`a`值组成，它们的平均值实际上也就是`a`值本身。

因此，两个查询都将返回相同的结果集，即所有`a`值小于10的`a`值的列表。

举个例子：

假设有表T如下：

a	b
5	200
5	150
9	120
9	110
15	90

- 对于查询C1：由于`Where`子句，我们只会考虑`a`值为5和9的行。由于`Group By a`，我们将得到两组（5和9），每组中的`a`值的平均数是`a`值本身（因为它们是相同的）。
- 对于查询C2：我们首先计算所有可能的`a`值的平均数，然后`Having`子句会排除掉所有平均值不小于10的`a`值的组。这将排除掉`a`值为15的行，所以我们只剩下`a`值为5和9的行，其结果与查询C1相同。

You
为什么b是错的，结合你的例子说明

ChatGPT
选项B是错误的，因为两个查询的过滤条件应用于不同的数据集。我们来仔细分析一下：

- 查询B1: ``Select Count(*) From T Where b>100 Group By a``
- 查询B2: ``Select Count(*) From T Group By a Having AVG(b)>100``

查询B1首先过滤掉所有`b`值小于或等于100的行，然后按`a`的值进行分组，并计算每个分组的行数。

查询B2首先对所有行按`a`的值进行分组，然后计算每个分组的`b`的平均值，最后只选择那些`b`的平均值大于100的分组。

在我的之前给出的例子中：

a	b
5	200
5	150
9	120
9	110
15	90

- 对于查询B1：我们先过滤掉`b`值小于或等于100的行，所以最后只考虑前四行。这样，`a`的值分别为5和9的行都会被计数，得到的结果是每个分组的计数，即：

a	Count
5	2
9	2

- 对于查询B2：首先对`a`的值进行分组，然后计算每个分组的`b`值的平均值：

a	AVG(b)
5	175
9	115
15	90

然后应用`Having AVG(b)>100`过滤条件，最终得到的结果将排除掉`a`值为15的分组，因为其平均值为90，不满足条件。所以结果是：

a	Count
5	2
9	2

虽然这个特定的例子在结果上看似相同，但这是因为每个分组的平均值都恰好大于100。如果我们考虑一个不同的例子，比如`b`的值有的大于100，有的小于100，但平均值恰好小于100，那么查询B2将不会包括那个`a`值的任何行，即使查询B1中可能有多行的`b`值大于100。

例如，如果我们在表中加入以下两行：

a	b
9	80
9	70

那么查询B1的结果不会改变，因为我们只考虑`b`值大于100的行。但查询B2的结果会改变，因为`a`值为9的行的平均`b`值现在是 $(120+110+80+70)/4 = 95$ ，这会导致`a`值为9的组不被包含在最终结果中。因此，这两个查询不是等价的。