

图解sql面试题:如何查找重复数据?

### 【题目】

编写一个SQL查询,查找学生表中所有重复的学生名。

学	学生表	
学号	姓名	
0001	猴子	
0002	马云	
0003	王思聪	
0004	王思聪	
0005	猴子	

猴子

## 【解题思路】

1.看到"找重复"的关键字眼,首先要用分组函数(group by),再用聚合函数中的计数函数 count()给姓名列计数。

2. 分组汇总后,生成了一个如下的表。从这个表里选出计数大于1的姓名,就是重复的姓名。

计数 猴子 马云 王思聪 2

猴子

#### ・方法一

【解题步骤】

1) 创建一个辅助表,将姓名列进行行分组汇总

select 姓名, count(姓名) as 计数

from 学生表 group by 姓名;

## 计数

猴子	2
马云	1
王思聪	2

猴子

#### where 计数 > 1;

2) 选出辅助表中计数大于1的姓名

select 姓名 from 辅助表

#### 计数 姓名

辅助表

猴子	2
马云	1
王思聪	2

条件: 计数>1的

学生姓名

猴子

#### select 姓名, count(姓名) as 计数 from 学生表

3) 结合前两步,将"创建辅助表"的步骤放入子查询

select 姓名 from

group by 姓名 )as 辅助表 where 计数 > 1; 结果:

from 学生表

group by 姓名

where count(姓名) > 1;

零学会SQL: 汇总分析》)

姓名 猴子 马云 猴子 ・方法二 这时候有的同学可能会想,为什么要这么麻烦创建一个子查询,不能用这个语句(将count放到 where字句中)直接得出答案吗? select 姓名

如果我们运行下这个sql语句,会报下面的错误,问题出在哪里呢?

前面提到聚合函数(count),where字句无法与聚合函数一起使用。因为where子句的运行顺序

排在第二,运行到where时,表还没有被分组。(如果不清楚,可以系统看下我之前的课程《从

如果要对分组查询的结果进行筛选,可以使用having子句。所以,这道题的最优方法如下:

## 【本题考点】

select 姓名

from 学生表

group by 姓名

having count(姓名) > 1;

3)熟记SQL子句的书写顺序和运行顺序。

2) 考察对having语句的掌握,很多人会把聚合函数写到where子句中。

1) 考察思路,有两种解题方法,但是使用having语句的方法更高效。

【举一反三】

# 本题也可以拓展为:找出重复出现n次的数据。只需要改变having语句中的条件即可:

select 列名 from 表名

group by 列名 having count(列名) > n;

推荐:如何提升你的分析技能,实现升职加薪? 编辑于 2022-09-07·著作权归作者所有

▲ 赞同 43 ● 评论 0