

单选题

第 1 题(本题2分): 对于查询 `Select Distinct sno, sname From Student, SC Where Student.sno=SC.sno And cno='C001'`, 请用常识判断以下哪个查询计划的执行效率应该最高? (假设 SC 在 cno 上有索引, Student 在 sno 上有索引。)

- ☐ A: (1) Student和SC在sno上做连接; (2) 在(1)的结果上做cno='C001'的选择; (3) 将(2)的结果在sno, sname上做投影。✗
- ☐ B: (1) 在SC上做cno='C001'的选择; (2) 将(1)的结果和Student在sno上做连接; (3) 将(2)的结果在sno, sname上做投影。✗
- ☐ C: (1) 在SC上做cno='C001'的选择; (2) 将Student在sno, sname上做投影, 将(1)的结果在sno上做投影; (3) 将(2)的两项结果在sno上做连接。✗
- ☒ D: (1) 在SC上做cno='C001'的选择; (2) 将(1)的结果在sno上做投影; (3) 将(1)的结果和Student在sno上做连接; (4) 将(3)的结果在sno, sname上做投影。✓

给定查询计划的选项是针对以下SQL查询的:

```
sql Copy code

Select Distinct sno, sname From Student, SC Where Student.sno=SC.sno And
```

我们假设`SC`在`cno`上有索引, `Student`在`sno`上有索引。现在, 我们来分析这些选项:

A. 先做连接操作, 然后选择, 最后投影。这个顺序不太高效, 因为它首先创建了一个可能很大的连接结果集, 然后才应用选择条件。

B. 先在`SC`上应用选择条件(利用索引), 这将显著减少连接操作的大小, 这是一个有效的步骤。然后做连接操作, 最后投影。

C. 先做选择操作, 然后投影两个表, 最后做连接。这个顺序有点不必要, 因为它在连接之前做了两次投影。

D. 先在`SC`上做选择, 然后在结果上做投影, 接着与`Student`做连接。最后, 在结果上做投影以获取最终的列。这个顺序避免了不必要的数据处理, 因为它在连接前应用了投影, 从而减少了连接操作处理的数据量。

在这些选项中, 选项D的效率最高。因为它最先对`SC`应用选择条件, 这可以立即减少数据集的大小, 利用了`cno`的索引。然后, 它在较小的数据集上进行`sno`的投影, 这进一步减少了与`Student`表连接时需要处理的数据量。最后, 它在连接的结果集上进行投影, 以产生去重后的最终结果。这个计划高效利用了索引并最小化了操作所涉及的数据量。

因此, 考虑到索引的使用和操作所涉及的数据量, 选项D应该是执行效率最高的查询计划。

第 2 题(本题2分): 对于一个关系代数表达的查询计划 $(\sigma(A) \bowtie \sigma(B)) \bowtie C$, 它有几种等价的执行方式? (σ 表示选择操作, \bowtie 表示连接操作, 这里A表和B表之间的连接属性与A表和B表之间的连接属性是不同的。)

- ☐ A: 2种 ✗
- ☐ B: 4种 ✗
- ☐ C: 10种 ✗
- ☒ D: 15种 ✓

第 3 题(本题2分): 给定一张表R(a,b,c,d), 假设一个查询为Select a,b,c From R Where a=10 OR b=100。以下说法正确的是?

- ☐ A: 索引对该查询无法提供帮助 ❌
- ☐ B: 在a或b上创建索引都可加速该查询的执行 ❌
- ☒ C: 在a和b上各创建一个索引可加速该查询的执行 ✓
- ☐ D: 在a,b上创建一个复合索引可加速该查询的执行 ❌

B选项说在`a`或`b`上创建索引都可以加速查询, 这是正确的, 但这个选项没有考虑到可以同时在这两个字段上创建索引的可能性, 这在处理带有OR条件的查询时可能更高效。

在考虑索引对查询性能的影响时, 需要了解数据库查询优化器如何使用索引。在给定的查询中:

```
sql Copy code  
  
Select a, b, c From R Where a=10 OR b=100
```

有几种索引可以提供帮助:

- **在`a`上创建索引**: 这可以快速找到所有`a=10`的行。
- **在`b`上创建索引**: 这可以快速找到所有`b=100`的行。
- **在`a`和`b`上各创建一个索引**: 查询优化器可以独立地使用每个索引来找到满足每个条件的行, 然后将结果合并。这在数据库可以有效地合并两个索引搜索的结果时是有用的。
- **在`a,b`上创建一个复合索引**: 这对于包含`AND`条件的查询很有用, 因为它可以在同一个索引结构中同时定位`a`和`b`的值。但对于`OR`条件, 复合索引的效果可能没有单独为每个字段创建索引来得好, 因为复合索引通常需要两个条件同时满足才能最有效。

对于选项C (在`a`和`b`上各创建一个索引可加速该查询的执行), 这种方法可以加速查询, 因为查询优化器可以分别快速访问`a=10`和`b=100`的行, 然后将这两组结果合并起来, 这通常比全表扫描要快。

因此, 选择C是正确的, 因为它提供了两种可能的索引, 查询优化器可以选择使用其中的任何一个或两个来加速查询。这是考虑到大多数现代数据库管理系统都能有效地合并使用多个索引的搜索结果。

第 4 题(本题2分): 对于两个等价的查询计划 $\sigma(A) \bowtie B$ 和 $\sigma(A \bowtie B)$, 要判断它们哪个执行效率更高, 以下哪些因素的作用最小。

- ☐ A: B上是否有索引 ❌
- ☐ B: $A \bowtie B$ 的结果集大小 ❌
- ☐ C: 系统内存空间的大小 ❌
- ☒ D: 系统硬盘空间的大小 ✓

选项D, “系统硬盘空间的大小”, 对于判断两个等价查询计划哪个执行效率更高的影响最小。

这两个查询计划是:

1. 对B进行连接操作后再进行A的选择操作: $\sigma_A(R \bowtie S)$
2. 先对R和S进行A条件的选择操作后再连接: $(\sigma_A(R)) \bowtie S$

我们分析一下每个选项的影响:

A. B上是否有索引: 如果B上有索引, 那么连接操作可能会更快, 因为索引可以加速连接条件的匹配过程。因此, B上的索引对性能有显著影响。

B. $A \bowtie B$ 的结果集大小: 结果集大小会影响处理的速度, 因为如果结果集很大, 那么操作将涉及更多的数据处理和可能的磁盘I/O操作。结果集的大小直接影响执行效率。

C. 系统内存空间的大小: 内存空间大小会影响数据库管理系统能够为查询计划保留多少内存缓冲区, 这可能会影响执行速度, 特别是如果结果集可以完全在内存中处理。

D. 系统硬盘空间的大小: 虽然硬盘空间大小可能会影响数据库的存储能力, 但它对查询执行的效率影响不大, 尤其是与索引的存在、结果集的大小和内存空间相比。查询的效率主要受到读写速度和数据处理能力的影响, 而不是硬盘的存储容量。只有在硬盘几乎满时, 可能因为找不到足够的连续空间来存储临时文件, 从而对性能产生负面影响, 但这种情况相对罕见。

因此, 硬盘空间大小对这两个查询计划的执行效率影响最小, 正确答案是D。