

第 1 题(本题2分): 如课程中提到的, 嵌套循环算法的I/O代价可以表示为 $B(R) + B(S) \cdot B(R)/M$, 其中 $B(R)$ 代表外循环表的页数, $B(S)$ 代表内循环表的页数, M 代表内存能容纳的页数。那么以下哪种方案无法提升该算法的效率?

- ☐ A: 增加内存容量 ✗
- ☒ B: 将小表放到内循环 (作为S), 大表放到外循环 (作为R) ✓
- ☐ C: 将大表放到内循环 (作为S), 小表放到外循环 (作为R) ✗
- ☐ D: 更换速度更快的硬盘 ✗

第 2 题(本题2分): 课程中介绍了两种连接操作执行算法, 嵌套循环和散列连接。以下哪种场景更适合使用嵌套循环算法?

- ☒ A: 一张表很大, 一张表很小。小的表几乎可以容纳到内存中。 ✓
- ☐ B: 两张表都比较大, 都无法容纳到内存中。 ✗
- ☐ C: 两张表都比较小, 都可以容纳到内存中。 ✗
- ☐ D: 几乎在所有情况下, 散列连接的效率都更高。 ✗

第 3 题(本题2分): 假设R表有1000行, S表有100行; 每页可以容纳10行数据 (无论R表或S表); 内存可以容纳3页; R表和S表均有x属性, 并且在x属性上均创建了索引; 假设x具备很高的辨识度并且分布均匀。那么, 理论上实现R和S在x上的等值连接的最佳算法是:

- ☐ A: 嵌套循环 ✗
- ☐ B: 散列连接 ✗
- ☒ C: 索引连接 ✓
- ☐ D: 都一样 ✗

第 1 题(本题2分): 如课程中提到的, 嵌套循环算法的I/O代价可以表示为 $B(R) + B(S) \cdot B(R)/M$, 其中 $B(R)$ 代表外循环表的页数, $B(S)$ 代表内循环表的页数, M 代表内存能容纳的页数。那么以下哪种方案无法提升该算法的效率?

- ☐ A: 增加内存容量 ✗ ✓
- ☒ B: 将小表放到内循环 (作为S), 大表放到外循环 (作为R) ✓ $B(R)$ 小点好
- ☐ C: 将大表放到内循环 (作为S), 小表放到外循环 (作为R) ✗
- ☐ D: 更换速度更快的硬盘 ✗ ✓

第 2 题(本题2分): 课程中介绍了两种连接操作执行算法, 嵌套循环和散列连接。以下哪种场景更适合使用嵌套循环算法?

- ☒ A: 一张表很大, 一张表很小。小的表几乎可以容纳到内存中。 ✓ 桶
- ☐ B: 两张表都比较大, 都无法容纳到内存中。 ✗
- ☐ C: 两张表都比较小, 都可以容纳到内存中。 ✗ 小表

2 题(本题2分): 课程中介绍了两种

哪种场景更适合使用嵌

☒ A: 一张表很大, 一张表很小。小的表几乎可以容纳到内存中。✓

☐ B: 两张表都比较大, 都无法容纳到内存中。✗

☐ C: 两张表都比较小, 都可以容纳到内存中。✗

☐ D: 几乎在所有情况下, 散列连接的效率都更高。✗

小表

大表

第 3 题(本题2分): 假设R表有1000行, S表有100行; 每页可以容纳10行数据 (无论R表或S表); 内存可以容纳3页; R表和S表均有x属性, 并且在x属性上均创建了索引; 假设x具备很高的辨识度并且分布均匀。那么, 理论上实现R和S在x上的等值连接的最佳算法是:

☐ A: 嵌套循环 ✗

☐ B: 散列连接 ✗

☒ C: 索引连接 ✓

☐ D: 都一样 ✗

100页 10页

index join