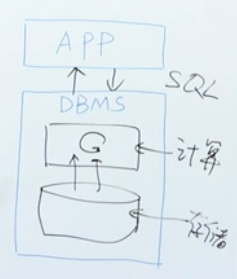
* 关系数据库的基本结构



* 1. 按页存储：使用表头串联（Inode/Index）
* SQL查询的执行过程
  1. 翻译Interpretation：SQL翻译成关系代数表达式
  2. 查询优化Query Optimization（预计）：寻找最优查询计划
  3. 查询执行 Query Evaluation
* 数据处理的性能问题

**第*1*题：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。√

**第*2*题：对于一个关系代数表达的查询计划 (σ(A) ⋈ σ(B)) ⋈ C，它有几种等价的执行方式？（σ表示选择操作，⋈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 √ **(A ⋈B) ⋈ C 3\*3种 A ⋈(B ⋈ C) 2\*3种**

**第*3*题：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

**第*4*题：对于两个等价的查询计划 σ(A) ⋈ B 和 σ(A ⋈ 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 ⋈ 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 √

作业

1-请根据你自己的理解找出以下查询的最佳关系代数查询计划。

Select sno, sname

From Student, SC

Where Student.sno = SC.sno

And grade<60

And cno in

(Select cno

From Course

Where cname = 'Math I');

2-对于关系表 R(a,b,c,d)，我们在属性b上创建了一个B树索引，那么投影操作 π​a,b​​(R)的执行是否可以利用该索引？如何利用该索引？

可以

3-课程中我使用了外部排序（External Sorting）的方式实现了投影操作的去重。请你设计另一种去重的方法，不使用排序，仍然可以达到很高的效率。

哈希