

## ● 第9章 存储管理

### 1. 试分析每个数据库对象对应一个操作系统文件和整个数据库对应一个或若干个文件，这两种存储关系数据库的策略各有什么优缺点

答：两种存储关系数据库的策略的优缺点如下：

- ① 每个 DB 对象对应一个操作系统文件。
  - 优点：易于管理和备份，可以利用操作系统的文件系统特性来进行管理，文件结构直接反映了数据库结构，便于理解和管理。
  - 缺点：文件数量庞大，数据库对象数量较多时，会产生大量的小文件，管理复杂性增加，并且可能对文件系统的性能造成影响。操作系统可能对单个文件的数量有一定限制，当文件数量过多时，查找和管理效率可能会下降。
- ② 整个 DB 对应一个或若干个文件（段页式）。
  - 优点：DBMS 控制这些大型文件，可以通过更好的磁盘访问模式和缓存管理优化其性能。由于数据库对象被集中存储，减少了操作系统需要管理的文件数量，从而简化了文件系统的负担。数据库可以更好地管理并发访问和锁定，因为它完全控制了所有数据的存储结构。
  - 缺点：复杂度增加，数据库内部的管理机制更加复杂，需要在大文件中处理碎片化、分配和回收空间等问题。备份和恢复复杂，由于数据集中存储，备份和恢复整个数据库时可能需要更大的时间和存储资源。这种方法不依赖操作系统的文件系统特性，可能需要数据库提供更多的管理功能，增加了 DBMS 的开发和维护成本。

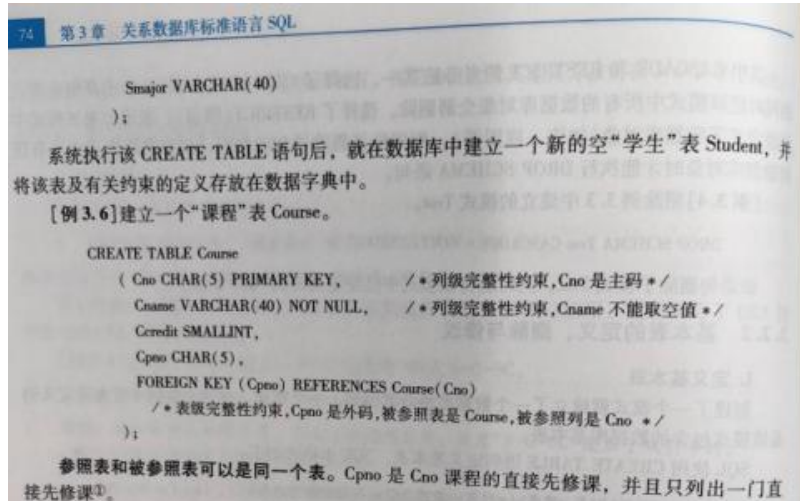
### 2. 假设 Course 表以定长记录方式存储。请描述 Course 表的记录存储，在以下情况下一条记录占多少个字节？

- 1) 字段可以在任何字节处开始
- 2) 字段必须在 4 的倍数的字节处开始
- 3) 字段必须在 8 的倍数的字节处开始

答：Course(Cno char(5), Cname char(40), Ccredit smallint, CPno char(5))

参考教材：P74，例 3.6 的定义，smallint 是 2 个字节

- 1) 52 字节
- 2) 60 字节
- 3) 64 字节



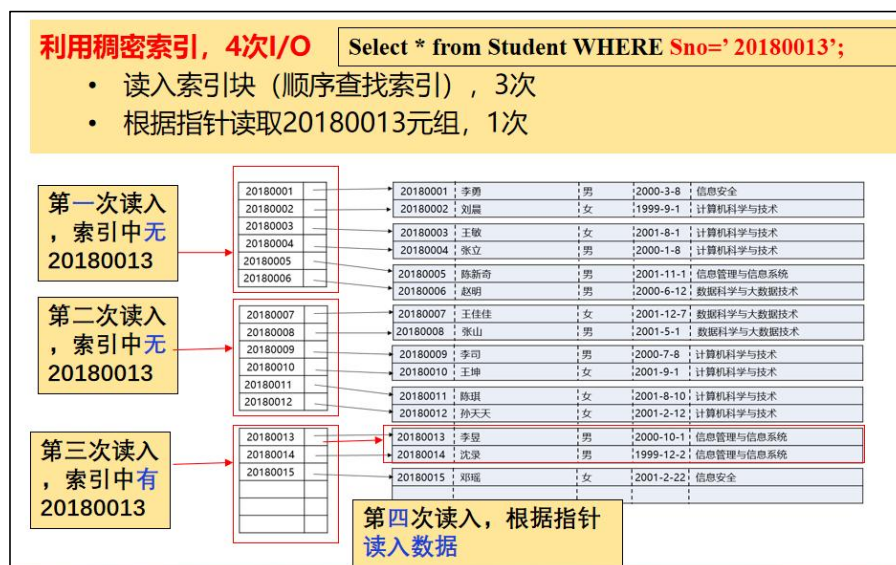
6. 分别描述利用图 9.12 的稠密索引、图 9.13 的稀疏索引、图 9.14 的多级索引、图 9.20 的 B+树索引、图 9.23 的哈希索引是如何对 student 表进行如下查询的。

- 1) 查询学号为 20180013 的记录。
- 2) 查询学号为 20180014 的记录。
- 3) 查询学号为 20180016 的记录。
- 4) 查询学号大于或等于 20180009 的记录。

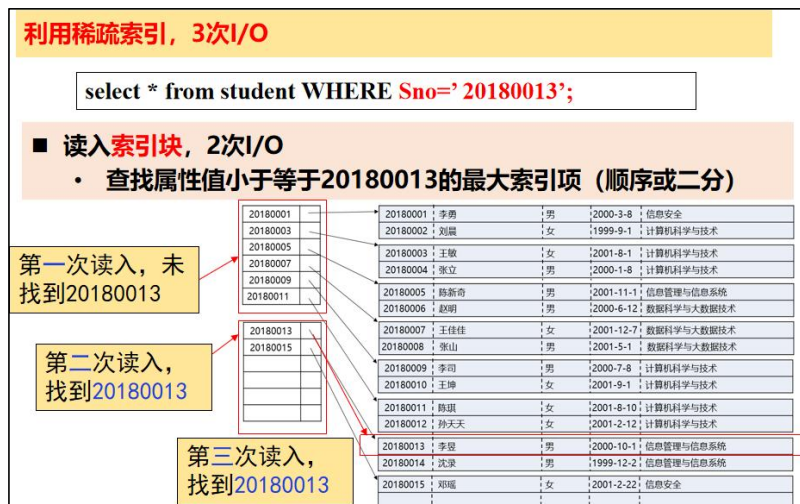
答:

- 1) 查询学号为 20180013 的记录。

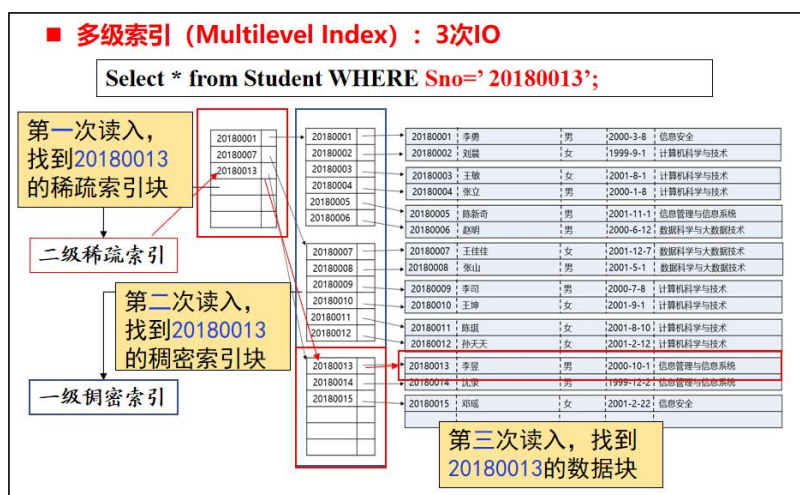
■ 图 9.12 稠密索引：4 次 I/O



■ 图 9.13 稀疏索引：共 3 次 I/O

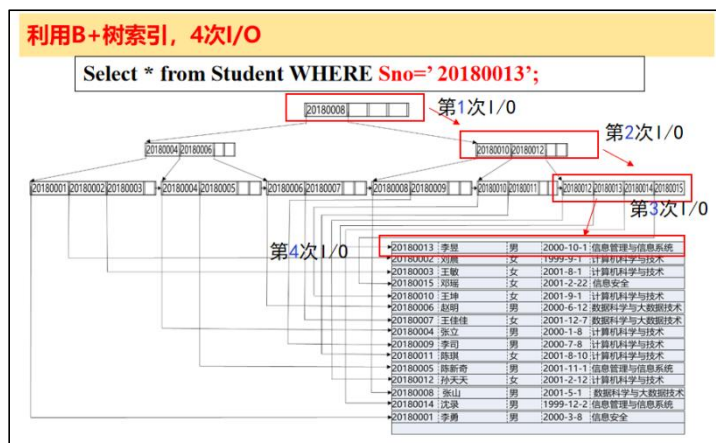


■ 图 9.14 多级索引：3 次 IO

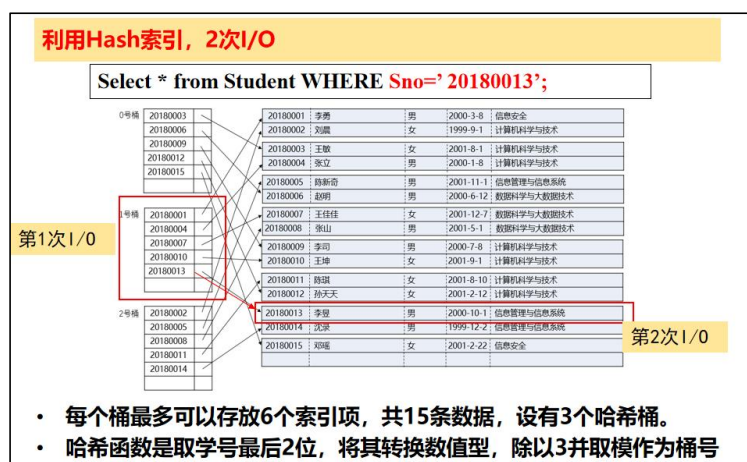


- ① 第一次 IO: 在二级稀疏索引中找到 20180013 所在的索引块
- ② 第二次 IO: 根据二级稀疏索引的 20180013 找到一级稠密索引的第三个索引块
- ③ 第三次 IO: 根据一级稠密索引的 20180013 找到数据存储块读到该行数据

■ 图 9.20 B+树索引：4 次 I/O



■ 图 9.23 哈希索引：2 次 I/O



2) 查询学号为 20180014 的记录。

■ 图 9.12 稠密索引：4 次 I/O

- ① 3 次 IO，3 次读稠密索引块，找到指向 20180014 的相关索引记录
- ② 根据稠密索引找到 20180014 数据块

■ 图 9.13 稀疏索引：共 3 次 IO

- ③ 第一次 IO：第一块稀疏索引中找 20180014 ( $>20180013$  且  $<20180015$ )，未找到
- ④ 第一次 IO：第二块稀疏索引中找 20180014 所在的索引块
- ⑤ 第三次 IO：根据稀疏索引找到 20180014 数据块

■ 图 9.14 多级索引：3 次 IO

- ⑥ 第一次 IO：在二级稀疏索引中找到 20180014 所在的索引块
- ⑦ 第二次 IO：根据二级稀疏索引的 20180014 找到一级稠密索引的第三个索引块
- ⑧ 第三次 IO：根据一级稠密索引的 20180014 找到数据块读到该行数据

■ 图 9.20 B+树索引：4 次 I/O

- ① 索引块 IO: 20180014 所在的索引块需要 3 次 IO (B+树的三层非叶节点)
- ② 数据块 IO: 读取 20180014 所在的数据库块 1 次 IO (B+树的叶节点)

■ 图 9.23 哈希索引: 2 次 I/O

- ① 第一次 IO: 读取按照哈希函数计算 20180014 所在的哈希桶
- ② 第二次 IO: 根据哈希桶的索引读取 20180014 所在数据块

3) 查询学号为 20180016 的记录。

- 图 9.12 稠密索引: 3 次 IO, 3 次读稠密索引块, 未找到相关记录, 查询结束
- 图 9.13 稀疏索引: 3 次 IO, 2 次读稀疏索引块, 1 次读数据块, 数据块内未找到记录
- 图 9.14 多级索引: 2 次 IO, 1 次读稀疏索引块, 1 次读稠密索引块, 未找到索引
- 图 9.20 B+树索引: 3 次 I/O, 读 3 次索引块(B+树的三层非叶节点), 未找到指向 20180016 的指针, 停止搜索
- 图 9.23 哈希索引: 1 次 I/O, 读取按照哈希函数计算 20180014 所在的哈希桶, 桶内未找到相关记录, 停止搜索

4) 查询学号大于或等于 20180009 的记录。

- 图 9.12 稠密索引: 6 次 IO, 2 次读稠密索引块 (第 1 块和第 2 块, 顺序存放无需读第 3 索引块), 4 次读数据块。
- 图 9.13 稀疏索引: 5 次 IO, 1 次读稀疏索引块 (第 1 块, 顺序存放无需读第 2 索引块), 4 次读数据块
- 图 9.14 多级索引: 6 次 IO, 1 次读稀疏索引块, 1 次读稠密索引块 (第 2 块, 顺序存放无需读第 3 索引块), 4 次读数据块
- 图 9.20 B+树索引: 12 次 I/O, 读 5 次索引块 (B+树的三层非叶节点), 7 个数据块
- 图 9.23 哈希索引: 假设条件哈希索引不支持范围查询, 走全表扫描;  
若数据顺序存储时, 二分查找一共 5 次 IO;  
若是随机存储, 一共 8 次 IO。

注意: 本题, 如果不按照 IO 次数描述, 仅写过程也可以。

- 稠密索引: 依次读入索引块, 由于稠密索引是有序索引, 因此可以在索引块中利用顺序查找或二分查找法, 查找 Sno 属性值为 20180009 的索引项。根据该索引项中的指针读入存储块, 取出满足条件的所有记录。由于文件是顺序存放的, 所以后面所有记录的 Sno 均大于或等于 20180009, 因此依次读

入之后的所有存储块并取出相应记录。

- 稀疏索引:依次读入索引块,找到小于或等于 20180009的最大索引项,即 20180009。根据该索引项中的指针读入存储块,取出满足条件的所有记录。由于文件是顺序存放的,所以后面所有记录的 Sno 均大于或等于 20180009,因此依次读入之后的所有存储块并取出相应记录。
- 多级索引:首先,读入二级索引(稀疏索引)的索引块,从中找到小于或等于 20180009的最大索引项,即 20180007。然后,根据二级索引项的指针,读取相应的一级索引(稠密索引)索引块,找到索引块中 Sno 属性值大于或等于 20180009的索引项,再根据索引项的指针读取存储块,从存储块中取出第一条满足条件的记录。由于关系表是顺序存放的,可知该记录后的所有记录均满足条件,因此依次遍历并取出该存储块中的记录。遍历完当前存储块后,依次读取其后的存储块及其中的记录,直到关系表的最后。
- B+树索引:首先,从根节点开始沿相应的父子结点指针逐层向下搜索,直到在叶结点中找到属性值为 20180009的索引项,可知其后所有叶结点中的索引项属性值均大于或等于20180009。根据当前叶结点中的指针读入存储块,读取存储块中满足条件的记录。遍历完当前叶结点后,沿指向下一叶结点的指针继续遍历,直到最后一个叶结点。
- 哈希索引:哈希索引本身并不能保证记录的顺序,因此不能利用哈希索引进行范围查询。通过依次读取每个存储块,遍历每条记录并取出所有 Sno>=20180009的记录。