# 第12章 索引与散列

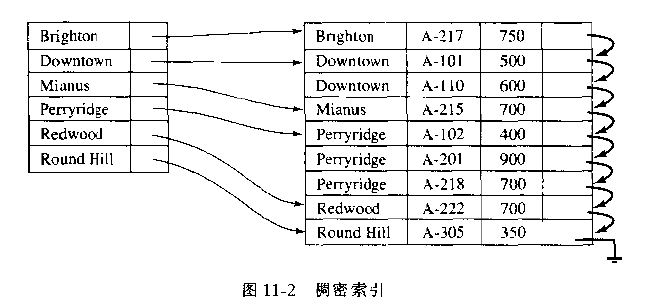
1. 基本索引类型：
2. **顺序索引（oedered index）**
3. **散列索引（hash index）**
4. 基本概念
5. **搜索码（search key）**：用于在文件中查找记录的属性或属性集称为搜索

码。

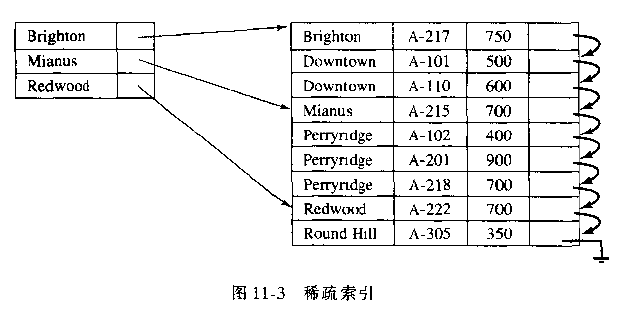
1. **索引结构：**每个索引结构与一个特定的搜索码相关联。



1. 评价索引技术的因素：
2. 访问类型（Access type）：能有效支持的访问类型。
3. 访问时间（Access time）：用该技术找到特定数据项所需的时间。
4. 插入时间（Insertion time）：插入一个新的数据项的时间。
5. 删除时间（Deletion time）：删除一个数据项所需的时间。
6. 空间开销（Spcae overhead）：一个索引结构占用的额外存储空间。
7. 顺序索引
8. 相关概念：
9. **聚集索引（clustering index）**：如果包含记录的文件按照某个搜索码指定的顺序排序，那么该搜索码对应的索引称为**聚集索引**，也称为**主索引（primary index）**。
10. **非聚集索引（nonclustering index）**：搜索码指定的顺序与文件中记录的物理顺序不同的索引称为**非聚集索引**，也称为**辅助（secondary）**索引。
11. **索引顺序文件：**称某个在搜索码上有聚集索引的文件为索引顺序文件。
12. 顺序索引类型：
13. 稠密索引（dense index）：文件中的每个搜索码值有一个索引记录。



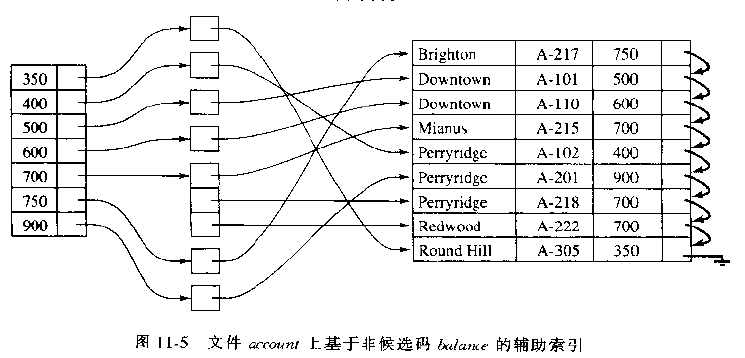
1. 稀疏索引（sparse index）：只为搜索码的某些值建立索引记录。



1. 辅助索引

（1）辅助索引必须是稠密索引，对于每个搜索码都有一个索引项，而且对文件中的每个记录都有一个指针；而聚集索引可以是稀疏索引，可以只存储部分搜索码值。

（2）如果辅助索引的搜索码不是一个候选码，仅仅指向具有每个搜索码值的第一条记录的指针是不够的，可以用一个附加的间接指针来实现非候选码的搜索码上的辅助索引：



1. 静态散列

（1）在对散列的描述中，我们使用术语**桶（bucket）**表示能存储一条或多条记录的存储单位。

（2）散列函数（hash function）特性：

1. 分布式均与的；
2. 分布式随机的。

（3）散列可以用于两个不同的目的：

1. **散列文件组织（hash file organization）**：通过计算所需记录搜索码值的函数直接获得包含该记录的磁盘块地址。
2. **散列索引组织（hash index organization）**：把搜索码以及与它们相关联的指针组织成一个散列文件结构。

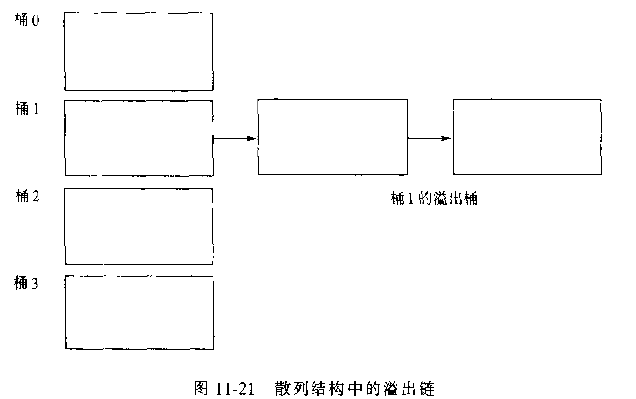
（4）**桶溢出（bucket overflow）**

**原因：**

1. 桶不足（insufficient bucket）
2. 偏斜（skew）：某些桶分配的记录比其他桶多。

解决方法：

1. **闭散列（closed hashing）**



1. **开散列（open hashing）**
2. 其他
3. 动态散列（dynamic hashing）：可扩充散列（extendable hashing）

当数据库增大或缩小时，可扩充散列可以通过桶的分裂或合并来适应数据库大小的变化。