

第七章 数据库设计

7.1 数据库设计概述

7.2 需求分析

7.3 概念结构设计

7.4 逻辑结构设计

7.5 物理结构设计

7.6 数据库的实施和维护

7.7 小结



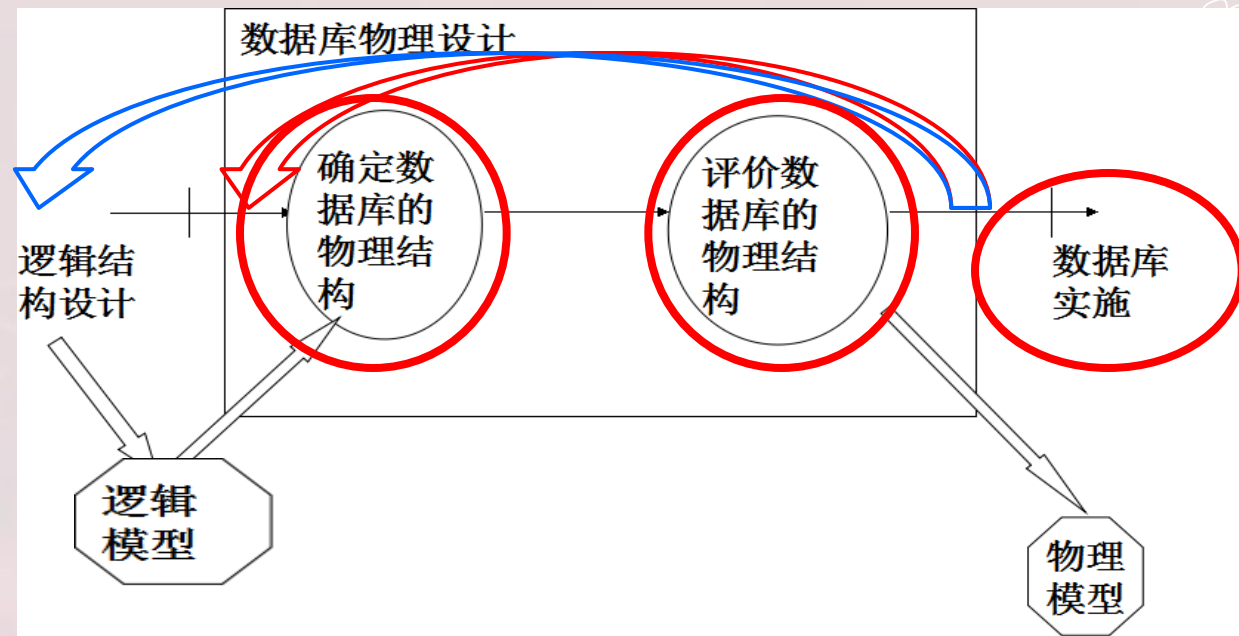
7.5 数据库的物理设计

❖ 什么是数据库的物理设计

- 为一个给定的逻辑数据模型选取一个**最适合应用要求的**物理结构的过程，就是数据库的物理设计。
- 数据库在物理设备上的存储结构与存取方法称为数据库的物理结构，它依赖于选定的DBMS。



数据库物理设计的步骤



◆ 确定数据库的物理结构

RDBMS中主要指存取方法和存储结构;

◆ 对物理结构进行评价

重点是时间和空间效率

IF 满足原设计要求

THEN 进入到物理实施阶段

ELSE

(重新设计

OR 修改物理结构

OR 返回逻辑设计阶段

修改数据模型)

复习：7.2 需求分析中 数据字典

❖ 4. 数据存储

- 数据量
- 存取频度：每小时、每天或每周存取次数，每次存取的数据量等信息
- 存取方法：批处理 / 联机处理；检索 / 更新；顺序检索 / 随机检索

❖ 5. 处理过程

- 处理要求
处理频度要求，如单位时间里处理多少事务，多少数据量、响应时间要求等

物理设计的输入 及 性能评价的标准



7.5 数据库的物理设计

7.5.1 数据库物理设计的内容和方法

7.5.2 关系模式存取方法选择

7.5.3 确定数据库的存储结构

7.5.4 评价物理结构



7.5.1 数据库物理设计的内容和方法

❖ 关系数据库物理设计的内容

- 为关系模式选择存取方法（建立存取路径）
- 为关系、索引、日志、备份等数据库文件选择物理存储结构



7.5.1 数据库物理设计的内容和方法

不同的DBMS产品

➤ 物理环境、存取方法和存储结构有很大差别

➤ 能供设计人员使用的设计变量、参数范围很不相同

➤ 没有通用的物理设计方法，只能给出一般的设计内容和原则。

设计物理数据库结构的准备工作

❖ 充分了解应用环境，详细分析要运行的事务，以获得选择物理数据库设计所需参数。

❖ 充分了解所用RDBMS的内部特征，特别是系统提供的存取方法和存储结构。

➤ 有哪些索引（例如B+树索引，HASH索引，BITMAP索引等），如何建立索引；

➤ 有哪些存储结构（行存储，列存储，聚簇存储），如何选择。



7.5.1 数据库物理设计的内容和方法

❖ 物理数据库设计参数（部分列选）

■ 数据库查询事务

- 查询所涉及的关系
- 查询条件所涉及的属性
- 连接条件所涉及的属性
- 查询的投影属性

数据库上运行的事务会不断变化、增加或减少，要根据应用的变化，及时收集设计信息，调整数据库的物理结构。

■ 数据更新事务

- 被更新的关系
- 每个关系上的更新操作条件所涉及的属性
- 修改操作要改变的属性值

■ 每个事务在各关系上运行的频率和性能要求



7.5 数据库的物理设计

7.5.1 数据库物理设计的内容和方法

7.5.2 关系模式存取方法选择

索引方法、聚簇方法

7.5.3 确定数据库的存储结构

7.5.4 评价物理结构



索引存取方法的选择

❖ 为什么要建立索引

提高存取的效率——查询、插入、删除、更新的效率

❖ 如何选择索引存取方法

根据应用要求确定：

- 对哪些属性列建立索引
- 对哪些索引要设计为唯一索引、组合索引、
- 选择合适的索引方法



索引存取方法的选择

❖ 如何创建索引

CREATE [UNIQUE] INDEX 索引名字 ON 表名 [USING
索引方法] (列名1, 列名2, [, ...]);

CREATE UNIQUE INDEX studentname ON student
USING Hash (sname) ;

❖ RDBMS提供的索引方法:

- B-tree (B+树), hash (散列) R-tree、Bitmap等。
- 如果不指定, 缺省一般是B-tree。



索引存取方法的选择

❖ 选择索引存取方法的一般规则

- 如果一个（或一组）属性经常**在查询条件中出现**，则考虑在这个（这组）属性上建立索引（或组合索引）；
- 如果一个属性经常作为**最大值和最小值等聚集函数的参数**，则考虑在这个属性上建立索引；
- 如果一个（或一组）属性经常在**连接操作的连接条件中**出现，则考虑在这个（或这组）属性上建立索引



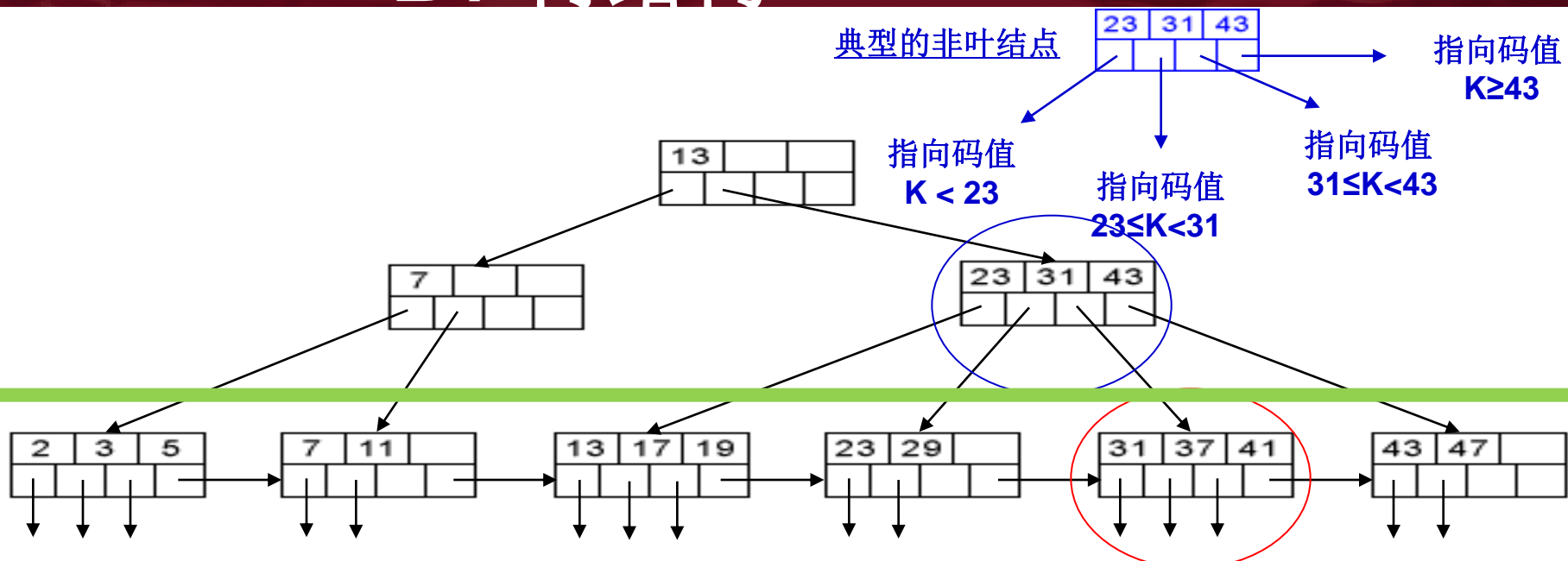
1 B+树索引简介

❖ B+树索引的特点:

- 多分平衡树，存取效率高
- 既能随机查找、又能顺序查找
- 增删改操作，保持平衡



B+ 树结构



典型的叶结点

31	37	41

指向顺序集上的
下一个叶结点

指向码值为
31的记录

指向码值为
37的记录

指向码值为
41的记录

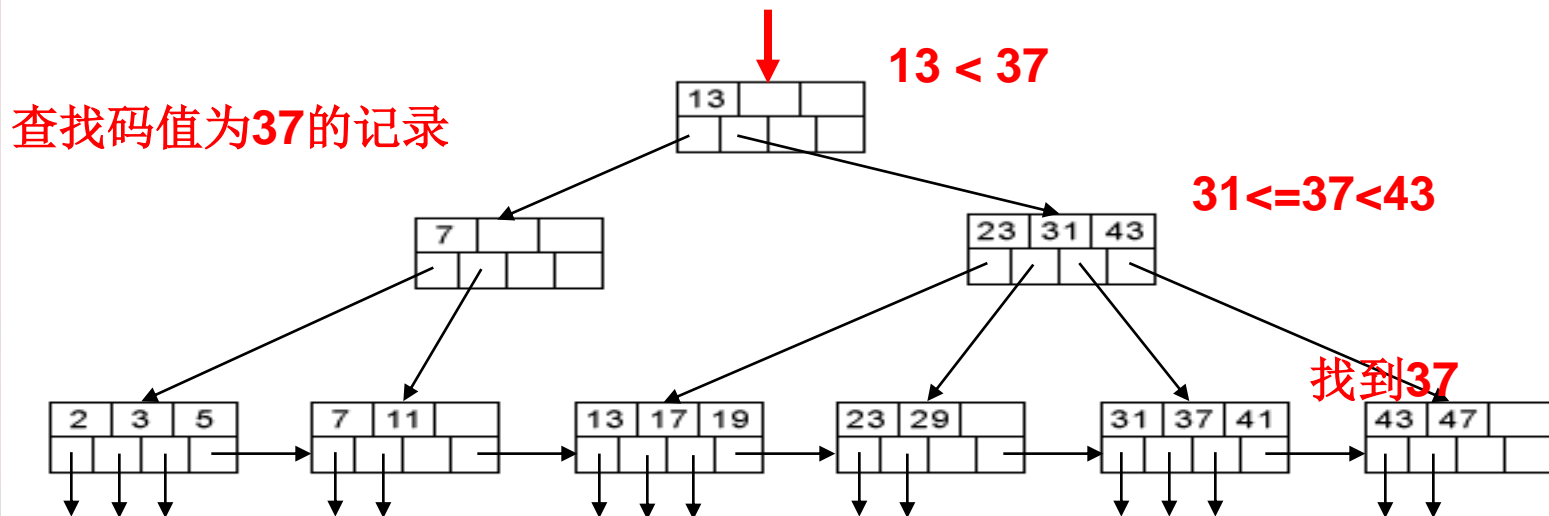
秩 $n=3$:

每个块存放码值的最大个数:3

最小码数: m

最小指针数: p

B+树中的随机查找



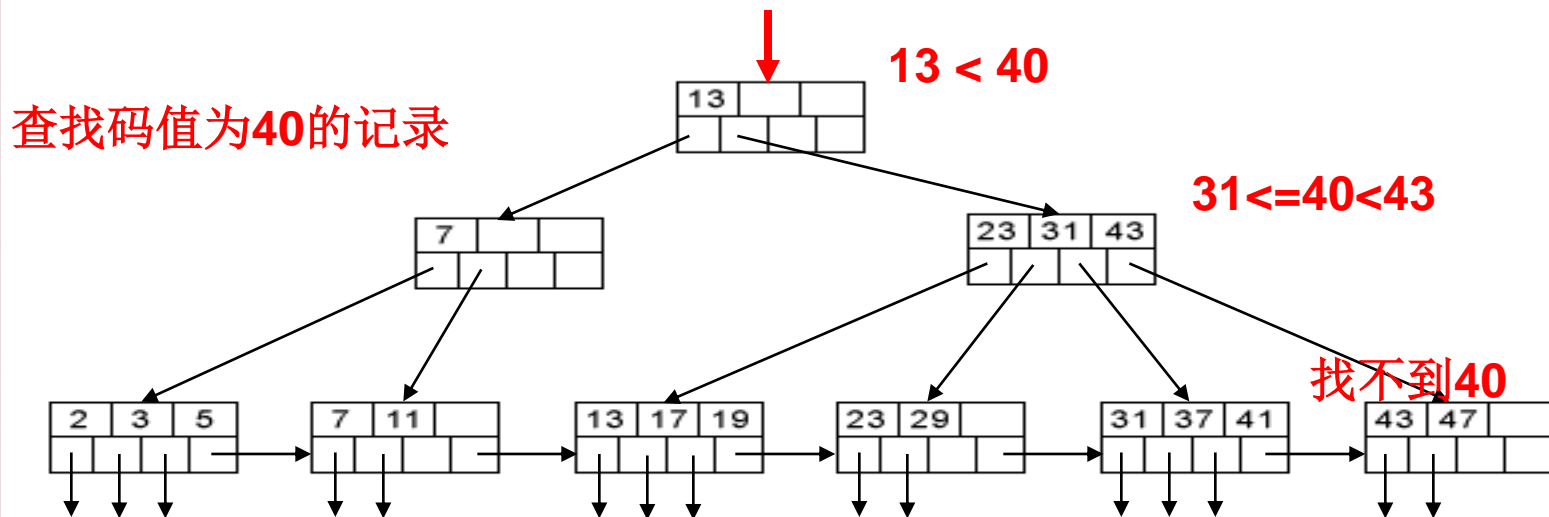
IO数: 4

若把第一、第二层结点保存在缓冲区

IO数: 2

码值为37的记录

B+树中的随机查找



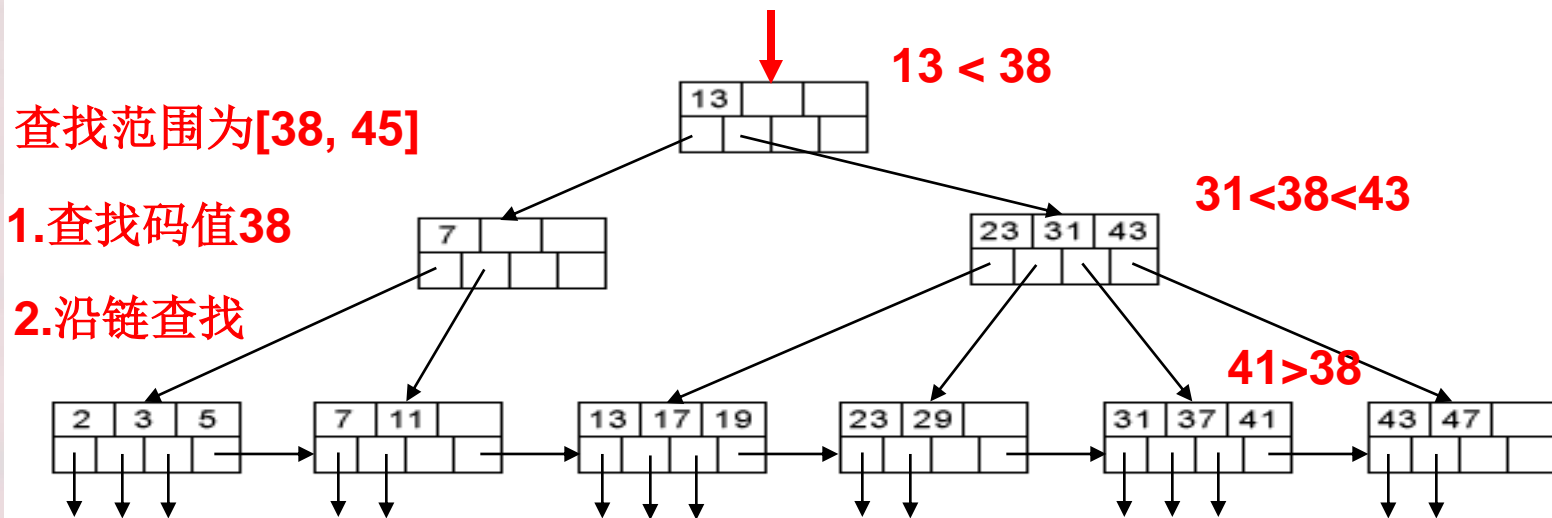
IO数: 3

若把第一、第二层结点保存在缓冲区

IO数: 1

没有码值为40的记录

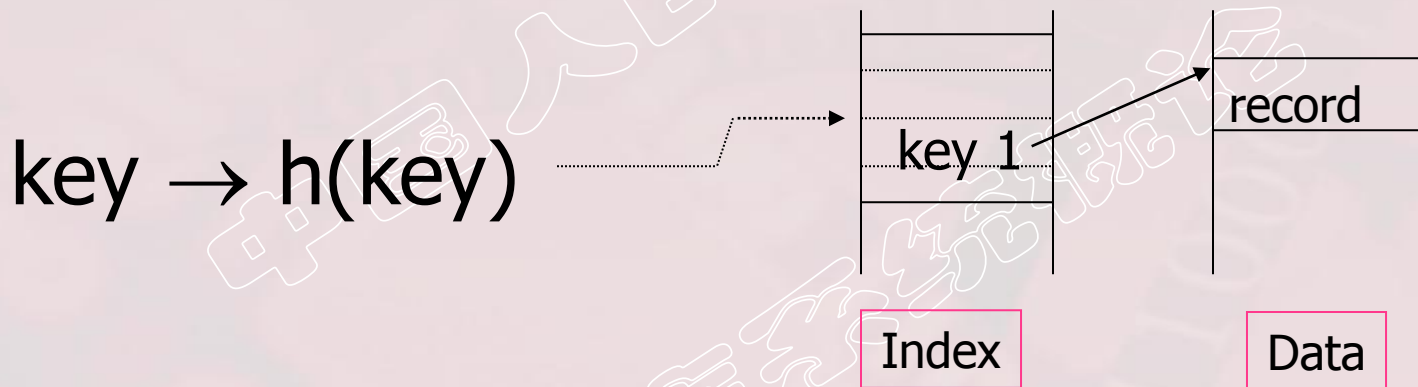
B+树中的范围查找



码值为41、43的记录

2. HASH存取方法的选择

❖ 为什么要使用Hash存取方法



2. HASH存取方法的选择

❖ 选择Hash存取方法的规则

如果一个关系的属性主要出现在**等值连接**条件中或主要出现在**等值比较**选择条件中，而且满足下列两个条件之一：

- 该关系的大小可预知，而且不变；
- 该关系的大小动态改变，但所选用的数据库管理系统提供了动态Hash存取方法。



索引存取方法的选择（续）

❖ 索引带来的额外开销

- 维护索引的开销
- 查找索引的开销
- 存储索引的开销

❖ 确定是否需要建立索引，选择哪种索引



7.5 数据库的物理设计

7.5.1 数据库物理设计的内容和方法

7.5.2 关系模式存取方法选择

索引方法、聚簇方法

7.5.3 确定数据库的存储结构

7.5.4 评价物理结构



索引存取方法的选择

❖ 为什么要建立索引

提高存取的效率

❖ 选择索引存取方法的一般规则

❖ B+树的结构、B+树中的随机查找和顺序查找



