

# 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 数据库的物理设计

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



### ❖ 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，它依赖于选定的数据库管理系统
- 为一个给定的逻辑数据模型选取一个最适合应用环境的物理结构的过程，就是数据库的物理设计

# 数据库的物理设计（续）

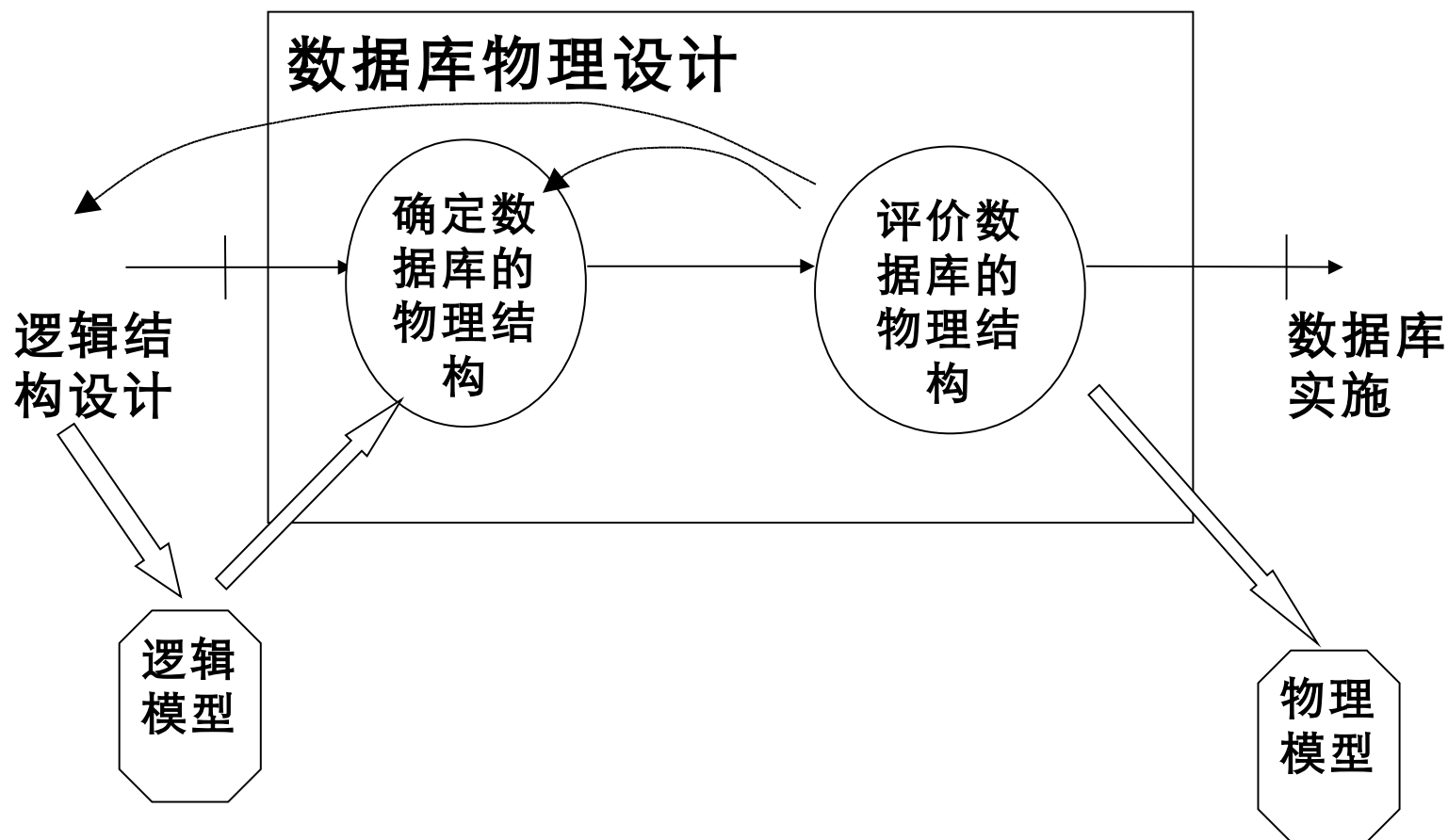


## ❖ 数据库物理设计的步骤

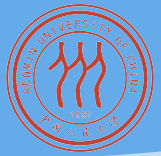
- 确定数据库的物理结构，在关系数据库中主要指存取方法和存储结构
- 对物理结构进行评价，评价的重点是时间和空间效率

如果评价结果满足原设计要求，则可进入到物理实施阶段，否则，就需要重新设计或修改物理结构，有时甚至要返回逻辑设计阶段修改数据模型

# 数据库的物理设计（续）



## 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

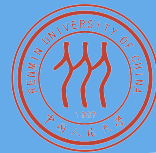
## 7.5.1 数据库物理设计的内容和方法



### ❖ 设计物理数据库结构的准备工作

- 对要运行的事务进行详细分析，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

# 数据库的物理设计的内容和方法（续）



## ❖ 选择物理数据库设计所需参数

- 数据库查询事务
  - 查询的关系
  - 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

# 数据库的物理设计的内容和方法（续）



## ❖ 选择物理数据库设计所需参数（续）

- 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



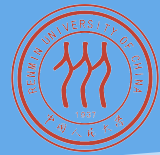
# 数据库的物理设计的内容和方法（续）



## ❖ 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）
- 设计关系、索引等数据库文件的物理存储结构

## 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**

## 7.5.2 关系模式存取方法选择



- ❖ 数据库系统是多用户共享的系统，对同一个关系要建立多条存取路径才能满足多用户的多种应用要求
- ❖ 物理设计的任务之一就是要确定选择哪些存取方法，即建立哪些存取路径

# 关系模式存取方法选择（续）



## ❖ DBMS 常用存取方法

- 索引方法
  - 目前主要是 B+ 树索引方法
  - 经典存取方法，使用最普遍
- 聚簇（Cluster）方法
- HASH 方法

# 一、索引存取方法的选择



## ❖ 根据应用要求确定

- 对哪些属性列建立索引
- 对哪些属性列建立组合索引
- 对哪些索引要设计为唯一索引

# 索引存取方法的选择（续）



## ❖ 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（或这组）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

## ❖ 关系上定义的索引数过多会带来较多的额外开销

- 维护索引的开销
- 查找索引的开销

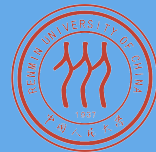
## 二、聚簇存取方法的选择



### ❖ 聚簇

- 为了提高某个属性（或属性组）的查询速度，把这个或这些属性（称为聚簇码）上具有相同值的元组集中存放在连续的物理块称为聚簇

# 聚簇存取方法的选择（续）



## ❖ 聚簇的用途

### ■ 1. 大大提高按聚簇码进行查询的效率

例：假设学生关系按所在系建有索引，现在要查询信息系的所有学生名单。

- 信息系的 500 名学生分布在 500 个不同的物理块上时，至少要执行 500 次 I/O 操作
- 如果将同一系的学生元组集中存放，则每读一个物理块可得到多个满足查询条件的元组，从而显著地减少了访问磁盘的次数



# 聚簇存取方法的选择（续）



## ■ 2. 节省存储空间

- 聚簇以后，聚簇码相同的元组集中在一起了，因而聚簇码值不必在每个元组中重复存储，只要在一组中存一次就行了

# 聚簇存取方法的选择（续）



## ❖ 聚簇的局限性

- 1. 聚簇只能提高某些特定应用的性能
- 2. 建立与维护聚簇的开销相当大
  - 对已有关系建立聚簇，将导致关系中元组移动其物理存储位置，并使此关系上原有的索引无效，必须重建
  - 当一个元组的聚簇码改变时，该元组的存储位置也要做相应移动

# 聚簇存取方法的选择（续）



## ❖ 聚簇的适用范围

- 1. 既适用于单个关系独立聚簇，也适用于多个关系组合聚簇

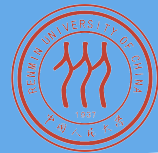
例：假设用户经常要按系别查询学生成绩单，这一查询涉及学生关系和选修关系的连接操作，即需要按学号连接这两个关系，为提高连接操作的效率，可以把具有相同学号值的学生元组和选修元组在物理上聚簇在一起。这就相当于把多个关系按“预连接”的形式存放，从而大大提高连接操作的效率。

# 聚簇存取方法的选择（续）



- 2. 当通过聚簇码进行访问或连接是该关系的主要应用，与聚簇码无关的其他访问很少或者是次要的时，可以使用聚簇。
  - 尤其当 SQL 语句中包含有与聚簇码有关的 ORDER BY，GROUP BY，UNION，DISTINCT 等子句或短语时，使用聚簇特别有利，可以省去对结果集的排序操作

# 聚簇存取方法的选择（续）



## ❖ 设计候选聚簇

- 对经常在一起进行连接操作的关系可以建立聚簇
- 如果一个关系的一组属性经常出现在相等比较条件中，则该单个关系可建立聚簇
- 如果一个关系的一个（或一组）属性上的值重复率很高，则此单个关系可建立聚簇。即对应每个聚簇码值的平均元组数不太少。太少了，聚簇的效果不明显

# 聚簇存取方法的选择（续）



## ❖ 优化聚簇设计

- 从聚簇中删除经常进行全表扫描的关系；
- 从聚簇中删除更新操作远多于连接操作的关系；
- 不同的聚簇中可能包含相同的表，一个表可以在某一个聚簇中，但不能同时加入多个聚簇
  - 从这多个聚簇方案（包括不建立聚簇）中选择一个较优的，即在这个聚簇上运行各种事务的总代价最小

# 三、HASH 存取方法的选择



## ❖ 选择 HASH 存取方法的规则

- 当一个关系满足下列两个条件时，可以选择 HASH 存取方法
  - 该关系的属性主要出现在等值连接条件中或主要出现在相等比较选择条件中
  - 该关系的大小可预知，而且不变；  
或  
该关系的大小动态改变，但所选用的 DBMS 提供了动态 HASH 存取方法

## 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**



## 7.5.3 确定数据库的存储结构



### ❖ 确定数据库物理结构的内容

- 1. 确定数据的存放位置和存储结构
  - 关系
  - 索引
  - 聚簇
  - 日志
  - 备份
- 2. 确定系统配置

# 1. 确定数据的存放位置



## ❖ 确定数据存放位置和存储结构的因素

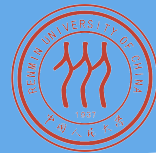
- 存取时间
- 存储空间利用率
- 维护代价

这三个方面常常是相互矛盾的

例：消除一切冗余数据虽能够节约存储空间和减少维护代价，但往往会导致检索代价的增加

必须进行权衡，选择一个折中方案

# 确定数据的存放位置（续）



## ❖ 基本原则

- 根据应用情况将
  - 易变部分与稳定部分分开存放
  - 存取频率较高部分与存取频率较低部分，分开存放

# 确定数据的存放位置（续）



例：

- 数据库数据备份、日志文件备份等由于只在故障恢复时才使用，而且数据量很大，可以考虑存放在磁带上
- 如果计算机有多个磁盘或磁盘阵列，可以考虑将表和索引分别放在不同的磁盘上，在查询时，由于磁盘驱动器并行工作，可以提高物理 I/O 读写的效率

# 确定数据的存放位置（续）



例（续）：

- 可以将比较大的表分别放在两个磁盘上，以加快存取速度，这在多用户环境下特别有效
- 可以将日志文件与数据库对象（表、索引等）放在不同的磁盘以改进系统的性能

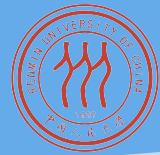
## 2. 确定系统配置



❖ DBMS 产品一般都提供了一些存储分配参数

- 同时使用数据库的用户数
- 同时打开的数据库对象数
- 内存分配参数
- 使用的缓冲区长度、个数
- 存储分配参数
- .....

## 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**

## 7.5.4 评价物理结构



### ❖ 评价内容

- 对数据库物理设计过程中产生的多种方案进行细致的评价，从中选择一个较优的方案作为数据库的物理结构



# 评价物理结构（续）



## ❖ 评价方法（完全依赖于所选用的 DBMS ）

- 定量估算各种方案
  - 存储空间
  - 存取时间
  - 维护代价
- 对估算结果进行权衡、比较，选择一个较优的合理的物理结构
- 如果该结构不符合用户需求，则需要修改设计

## 小结（续）



- 在逻辑设计阶段将 E-R 图转换成具体的数据库产品支持的数据模型如关系模型，形成数据库逻辑模式。然后根据用户处理的要求，安全性的考虑，在基本表的基础上再建立必要的视图（VIEW）形成数据的外模式
- 在物理设计阶段根据 DBMS 特点和处理的需要，进行物理存储安排，设计索引，形成数据库内模式