

# 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



### 3. 聚簇存取方法的选择

#### ❖ 什么是聚簇

- 为了提高某个属性（或属性组）的查询速度，把这个（或这些）属性上具有相同值的元组集中存放在连续的物理块中称为聚簇。
- 该属性（或属性组）称为聚簇码（cluster key）
- 许多RDBMS都提供了聚簇功能



# 建立聚簇的方法示例

## 1. 先创建一个聚簇

```
CREATE CLUSTER <聚簇名> (<聚簇码>) SIZE (<大小>);
```

## 2. 在聚簇上建立索引

```
CREATE INDEX <索引名> ON CLUSTER <聚簇名>;
```

[例] `CREATE CLUSTER emp_dept_cluster (deptno number(6) ) SIZE 1024;`  
`CREATE INDEX emp_dept_cluster_index ON CLUSTER emp_dept_cluster;`

各个RDBMS产品的关于聚簇的建立、维护的方法和语法不尽相同



# 聚簇存取方法的选择

## ❖ 聚簇的用途

### 1. 大大提高按聚簇属性进行查询的效率

[例] 假设要查询计算机系的所有学生。

- 学生数据表随机存放，计算机系的500名学生分散存储在500个不同的物理块上，则至少要执行500次 I/O 操作。
- 如果按照专业系名聚簇存放，将同一系的学生元组聚簇在一起存放，则可以显著地减少了访问磁盘的次数。计算机系的500名学生聚簇存储在50个不同的物理块上，只要执行50次 I/O 操作。



# 聚簇存取方法的选择（续）

## ❖ 聚簇的适用范围

➤ 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] 假设用户经常要按姓名查询学生成绩单。

```
SELECT sname, cno, grade from student, sc where student.sno=sc.sno
```

这一查询涉及学生关系和选修关系的连接操作，按学号连接这两个关系。

- 按照学号把学生表和选修表聚簇在一起。
- 相当于把多个关系按“预连接”的形式存放。
- 大大提高连接操作的效率。



# 聚簇存取方法的选择（续）

## ❖ 聚簇的适用范围

- 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，可以省去或减少对结果集的排序操作





# 聚簇存取方法的选择（续）

## ❖ 聚簇的局限性

- 在一个基本表上最多只能建立一个聚簇索引
- 聚簇只能提高某些特定应用的性能
- 建立与维护聚簇的开销相当大
- 对已有关系建立聚簇，将导致关系中元组的物理存储位置移动，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相应改变。

# 聚簇存取方法的选择（续）

## ❖ 聚簇索引的适用条件

- 很少对基表进行增删操作
- 很少对其中的变长列进行修改操作





# 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

### ❖ 确定数据的存储安排和存储结构

- 关系
- 索引
- 数据库缓冲区
- 日志
- 备份

- 内存/磁盘
- 行存储/列存储
- 集中/分散 存放
- 顺序/随机/聚簇 存放

### ❖ 确定系统参数配置

各个系统所能提供的对数据进行物理安排的手段、方法差异很大；  
设计人员应仔细了解给定的RDBMS提供的方法和参数，针对应用环境的要求，对数据进行适当的物理安排。

# 确定数据库的存储结构

## ❖ 影响数据存放位置和存储结构的因素

### ■ 硬件环境

### ■ 应用需求

- 存取时间
- 存储空间利用率
- 维护代价

必须进行权衡，选择一个折中方案

这三个方面常常是相互矛盾的



# 1. 确定数据的存放位置

## ❖ 基本原则

### ■ 根据应用情况将

- 易变部分与稳定部分分开存放
- 经常存取部分与存取频率较低部分分开存放

## ❖ [例]

- 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。

# 1. 确定数据的存放位置

## ❖ 基本原则：

根据应用情况和物理环境（磁盘或磁盘阵列的容量、内存的大小）

## ❖ 易变部分与稳定部分分开存放

## ❖ 经常存取部分与存取频率较低部分分开存放

## ❖ 将日志文件与数据库对象（表、索引等）分开存放

在海量数据和多用户环境下，把数据分布存放在不同的磁盘或磁盘阵列上，可以改进系统性能。

## 2. 确定系统配置

### ❖ 数据库管理系统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
- 缓冲区分配参数（使用的缓冲区长度、个数）
- 存储分配参数
- 物理块的大小
- 物理块装填因子
- 数据库的大小
- 锁的数目等





# 确定系统配置（续）

- ❖ 系统都为这些变量赋予了合理的缺省值。

在进行物理设计时需要根据应用环境确定这些参数值，以使系统性能最优。

- ❖ 在物理设计时对系统配置变量的调整只是初步的，要根据系统实际运行情况做进一步的调整，以切实改进系统性能



# 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

- ❖ 对数据库物理设计过程中产生的多种方案进行评价，从中选择一个较优的方案作为数据库的物理结构。
- ❖ 评价方法
  - 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
  - 返回用户 征求意见 修改设计



# 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**

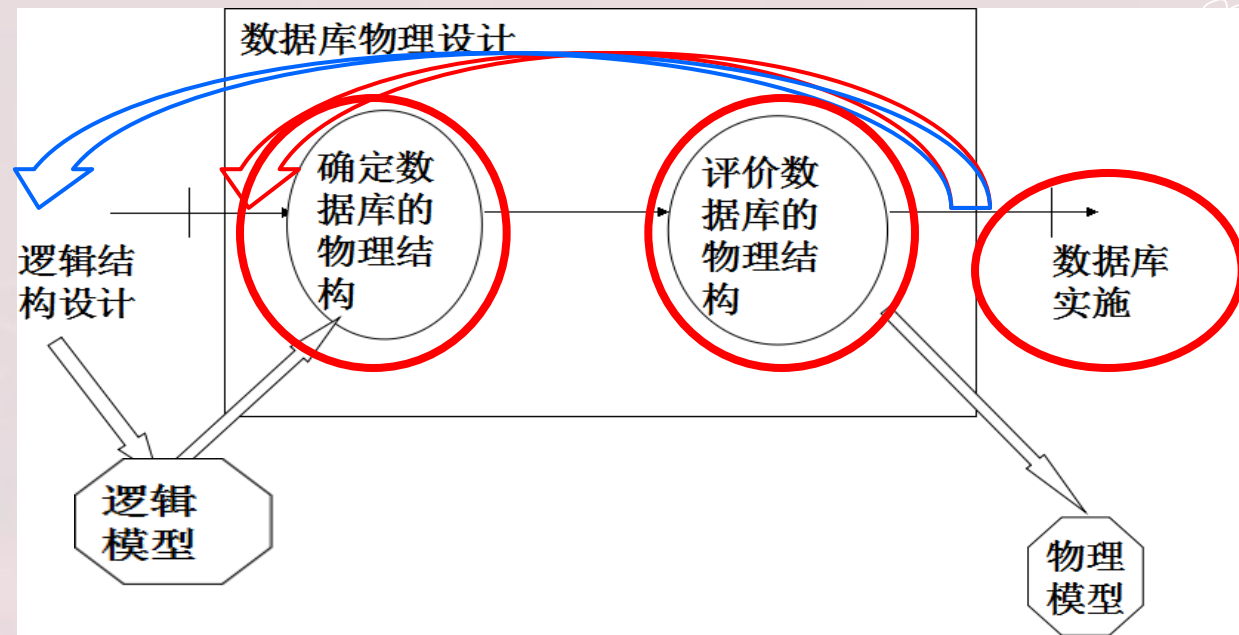
**7.5.2 关系模式存取方法选择**

**7.5.3 确定数据库的存储结构**

**7.5.4 评价物理结构**



# 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 评价结果满足原设计要求  
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)

