



7.5 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，它依赖于给定的计算机系统。

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。

👉 本节内容：

7.5.1 数据库物理设计的步骤

7.5.2 数据库的物理设计的内容和方法

7.5.3 关系模式存取方法选择

7.5.4 确定数据库的存储结构

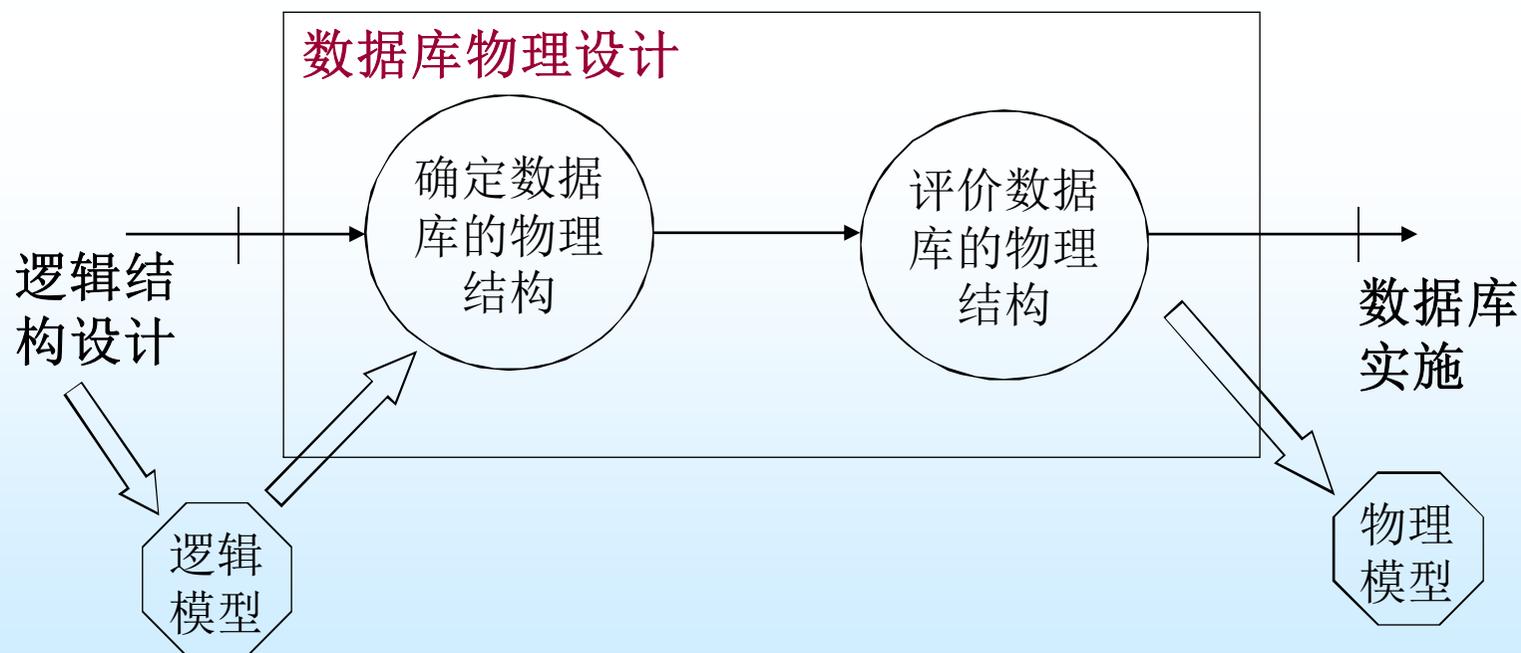
7.5.5 评价物理结构



7.5.1 数据库物理设计的步骤

数据库物理设计的步骤

- * 确定数据库的物理结构
- * 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。





7.5.2 物理设计的内容和方法

☞ 设计物理数据库结构的准备工作

1. 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
2. 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

☞ 选择物理数据库设计所需参数

(1) 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



7.5.2 物理设计的内容和方法

(2) 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

(3) 每个事务在各关系上运行的频率和性能要求

☞ 关系数据库物理设计的内容

1. 为关系模式选择**存取方法**(建立存取路径)
2. 设计关系、索引等数据库文件的**物理存储结构**



7.5.3 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。

物理设计的第一个任务就是要确定选择哪些存取方法，即建立哪些存取路径。

DBMS常用存取方法

- (1) 索引方法，目前主要是B+树索引方法
- (2) 聚簇 (Cluster) 方法
- (3) HASH方法



7.5.3 关系模式存取方法选择

一、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(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



7.5.3 关系模式存取方法选择

二、聚簇索引的选择

- * 建立聚簇索引后，基表中数据也需要按指定的聚簇属性值的升序或降序存放。也即聚簇索引的索引项顺序与表中元组的物理顺序一致。

☞ 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。

- * 2. 节省存储空间

- 聚簇以后，聚簇码相同的元组被集中在一起，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。



7.5.3 关系模式存取方法选择

☞ 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。

☞ 聚簇索引的选择原则

1. 经常进行更新操作的表不能进行聚簇索引
2. 对经常进行等值连接(自然连接)的表的连接项进行聚簇
3. 对重复率高的属性(组)进行聚簇索引，可以减少存储空间
4. 综合考虑聚簇代价与性能的关系，选择必要的聚簇



7.5.3 关系模式存取方法选择

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

(1) 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

(2) 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。



7.5.4 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
- * 2. 确定系统配置

☞ 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

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。



7.5.4 确定数据库的存储结构

☞ 确定数据库存储结构的基本原则

- * 根据应用情况将**易变**部分与**稳定**部分、**存取频率较高**部分与**存取频率较低**部分分开存放，以提高系统性能

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，可以考虑将表和索引分别放在不同的磁盘上，在查询时，由于两个磁盘驱动器分别在工作，因而可以保证物理读写速度比较快。
- * 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。



7.5.4 确定数据库的存储结构

☞ 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
- * 使用的缓冲区长度、个数
- * 时间片大小
- * 数据库的大小
- * 装填因子
- * 锁的数目
- * 等等

注：(1) 系统都为这些变量赋予了合理的缺省值。但是这些值不一定适合每一种应用环境，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。

(2) 在物理设计时对系统配置变量的调整只是初步的，在系统运行时还要根据系统实际运行情况做进一步的调整，以期切实改进系统性能。



7.5.5 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，选择出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