

数据库系统课程教案

第 19 单元

学时：2

教材内容	第 7 章 数据库设计 (7.5~7.6)
基本知识 点	数据库的实施、运行与维护，数据库物理设计的概念与步骤、关系模式存取方法的选择、数据库存储结构的确定和评价方法
教学重点	数据库物理设计的概念与步骤、关系模式存取方法的选择、数据库存储结构的确定和评价方法
教学难点	系模式存取方法的选择
要求掌握 内容	数据库的实施、运行与维护，数据库物理设计的概念与步骤、关系模式存取方法的选择、数据库存储结构的确定和评价方法
教学思路， 采用的教 学方法和 辅助手段， 板书设计， 重点如何 突出，难点 如何解决， 师生互动 等	<p>教学思路：</p> <p>一、复习旧课，巩固上次授课主要内容</p> <ol style="list-style-type: none"> 1、简述逻辑结构设计的任务。 2、简述 E-R 图向关系模型转换的原则。 <p>二、导入新课，明确本次授课的目的与要求</p> <ol style="list-style-type: none"> 1、掌握数据库物理设计的概念与步骤。 2、掌握关系模式存取方法的选择。 3、掌握数据库存储结构的确定和评价方法。 4、理解数据库的实施。 5、理解数据库的运行与维护。 <p>三、讲解本次授课的具体内容</p> <p>教学方法：整合教学内容，强化基础训练；努力营造生动活泼的课堂气氛，搭建师生间良好的沟通渠道；采用多媒体教学与传统的板书设计相结合的方式，教学手段灵活多变。</p> <p>辅助手段：通过 PPT 幻灯片演示结合板书设计和例题加以阐述。</p>
本章思考题 和作业	P241 第 11、12、13、14 题
主要教材 参考资料	<ol style="list-style-type: none"> 1. 《数据库系统概论》，萨师煊，王珊，高等教育出版社，2014.9 2. 《数据库系统概论学习指导与习题解答》，王珊，张俊，高等教育出版社，2015.7
备 注	

本次授课具体内容

第 7 章 数据库设计

7.5 物理结构设计

(一)数据库物理设计的概念：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。

(二)数据库物理设计的步骤

- 1、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。
- 2、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。

(三)数据库物理设计的内容和方法

1、准备工作

- (1)充分了解应用环境，详细分析要运行的事务，以获得选择数据库物理设计所需的参数。
- (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。

2、选择数据库物理设计所需的参数

- (1)数据查询事务：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、查询的属性等。
- (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、每个事务在各关系上运行的频率和性能要求。

3、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。

(四)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 常用的存取方法有：

1、索引(主要是 B+树索引)存取：是使用最普遍的存取方法，选择索引存取方法的一般规则是：

(1)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)。

(2)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。

(3)如果一个(或组)属性经常在连接操作的条件中出现，则考虑在这个(或这组)属性上建立索引。但关系上定义的索引数过多会增加维护索引和查找索引的开销。

2、聚簇存取：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。许多关系型 DBMS 都提供了聚簇功能。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，即聚簇索引的索引项顺序与表中元组的物理顺序一致。例 Create cluster index Stusname on Student(Sname)；在 Student 表的 Sname 列上建立一个聚簇索引，表中的记录将按 Sname 值的升序存放。一个数据库可以建立多个聚簇，但一个基本表上最多只能建立一个聚簇索引。

(1)聚簇的用途：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。另外，聚簇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，节省了存储空间。

(2)聚簇的局限性：聚簇只能提高某些特定应用的性能，而且建立与维护聚簇的开销是相当大的。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。

(3)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。例如，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。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是次要时，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。

③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组合聚簇；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，否则聚簇效果不明显。

④检查候选聚簇中的关系，取消其中不必要的关系：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；从聚簇中删除重复出现的关系，即当一个关系同时加入多个聚簇时，必须从这多聚簇方案中选择一个较优的，使该聚簇上运行各种事务的总代价最小。3、HASH 存取：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相等比较选择条件中，而且满足下列两个条件之一时，可以选择 HASH 存取方法。

(1)关系的大小可预知，而且不变。

(2)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。

(四)确定数据库的存储结构：主要是指确定数据的存放位置和存储结构，包括关系、索引、聚簇、日志、备份等的存储安排和存储结构；确定系统配置等。确定数据的存放位置和存储结构要综合考虑存取时间、空间、维护代价三方面的代价，这三个方面常常是相互矛盾的，如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，因此必须进行权衡，选择一个折中方案。

1、确定数据的存放位置：根据应用情况将数据的易变部分与稳定部分、经常存取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，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。

2、确定系统配置：DBMS 产品提供的存储分配参数有同时使用数据库的用户数和对象数、使用

的缓冲区长度和个数、时间片大小、数据库大小、装填因子、锁的数目等。系统都为这些变量赋了合理的缺省值，但这些值不一定适合各种应用环境，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。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。

(E)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。评价方法完全依赖于所选用的 DBMS，主要从定量估算各种方案的存储空间、存取时间和维护代价入手，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。

7.6 数据库的实施和维护

(一)数据库实施：完成数据库物理设计后，设计人员要用 RDBMS 提供的 DDL 和其他实用程序将数据库逻辑设计和物理设计结果严格描述出来，成为 DBMS 可以接受的源代码，再经过调试产生目标模式，然后就可以组织数据入库，这就是数据库实施。数据的载入和应用程序的调试是实施阶段的两项重要工作

1、数据的载入

(1)人工方法：适用于小型系统，步骤是：

①筛选数据：需要装入数据库中的数据通常都分散在各个部门的数据文件中，首先必须把需要入库的数据筛选出来。

②转换数据格式：筛选出来的需要入库的数据，其格式往往不符合数据库要求，还需要进行转换。

③输入数据：将转换好的数据输入计算机中。

④校验数据：检查输入的数据是否有误。

(2)计算机辅助数据入库：适用于中大型系统，步骤是：

①筛选数据。

②输入数据：数据输入子系统应提供输入界面。

③校验数据：数据输入子系统采用多种检验技术检查输入数据的正确性。

④转换数据：数据输入子系统根据数据库系统的要求，从录入的数据中抽取有用成分，对其进行分类，然后转换数据格式。

⑤综合数据：数据输入子系统对转换好的数据根据系统的要求进一步综合成最终数据。

如果数据库是在老的文件系统或数据库系统的基础上设计的，则数据输入子系统只需要完成转换数据、综合数据两项工作，直接将老系统中的数据转换成新系统中需要的数据格式。

2、应用程序的编制与调试：数据库应用程序的设计应该与数据设计并行进行，因此在组织数据入库的同时还要调试应用程序。调试应用程序时由于数据入库尚未完成，可先使用模拟数据。

(二)数据库的试运行：应用程序调试完成且已有小部分数据入库后，就可以进行数据库的试运行。

1、试运行的任务：实际运行应用程序，执行对数据库的各种操作，测试应用程序的各种功能、性能是否满足设计要求。如果不满足，对应用程序部分则要修改、调试，直至达到设计要求。物理设计阶段在评价数据库结构估算时间、空间指标时作了许多简化和假设，忽略了许多次要因素，因此结果必然很粗糙，数据库试运行则是要实际测量系统的各种性能指标(不仅是时间、空间指标)，如果结果不符合设计目标，则需要返回物理设计阶段，调整物理结构，修改参数；有时甚至需要返回逻辑设计阶段，调整逻辑结构。

2、数据的分期入库：重新设计物理结构甚至逻辑结构，会导致数据重新入库。由于数据入库工作量实在太大，所以可以采用分期输入数据的方法先输入小批量数据供先期联合调试使用待试运行基本合格后再输入大批量数据逐步增加数据量，逐步完成运行评价。3、数据库的转储和恢复：在数据库试运行阶段，系统还不稳定，硬、软件故障随时都可能发生。系统的操作人员对新系统还不熟悉，误操作也不可避免。因此必须做好数据库的转储和恢复工作，尽量减少对数据库的破坏。

(三)数据库的运行和维护：由于应用环境在不断变化，数据库运行过程中物理存储也会不断变化，

对数据库设计进行评价、调整、修改等维护工作是一个长期的任务，也是设计工作的继续和提高。

1、数据库的转储和恢复：转储和恢复是系统正式运行后最重要的维护工作之一。DBA 要针对不同的应用要求制定不同的转储计划，以保证一旦发生故障能尽快将数据库恢复到某种一致性状态。

2、数据库的安全性、完整性控制：在数据库运行过程中，由于应用环境的变化，对安全性的要求也会发生变化，DBA 要根据实际情况修改原有安全性控制。由于应用环境的变化，数据库的完整性约束条件也会变化，也需 DBA 不断修正，以满足用户要求。

3、数据库性能的监督、分析和改进：在数据库运行过程中，DBA 必须监督系统运行，对监测数据进行分析，找出改进系统性能的方法。利用监测工具获取系统运行过程中一系列性能参数的值。通过仔细分析这些数据，判断当前系统是否处于最佳运行状态。如果不是，则需要通过调整某些参数来进一步改进数据库性能。4、数据库的重组和重构造

(1)重组：数据库运行一段时间后，由于记录的增、删、改，会使数据库的物理存储变坏，降低了数据的存取效率，数据库性能下降，这时 DBA 就要对数据库进行重组或部分重组(只对频繁增、删的表进行重组)。DBMS 一般都提供了供重组数据库使用的实用程序，帮助 DBA 重新组织数据库。在重组过程中，按原设计要求重新安排存储位置、回收垃圾、减少指针链，提高系统性能，但数据库的重组不会改变原设计的数据逻辑结构和物理结构。

(2)重构造：由于数据库应用环境发生变化，使原有的数据库设计不能很好地满足新的需求，需要调整数据库的模式和内模式。例如，在表中增加或删除某些数据项、改变数据项的类型、改变数据库的容量、增加或删除索引等。重构造数据库的程度是有限的，若应用变化太大，重构也无济于事，说明此数据库应用系统的生命周期已经结束，应该重新设计新的数据库系统。

本次授课小结

本次授课讲述了数据库的物理设计的概念与步骤、关系模式存取方法的选择、数据库存储结构的确定和评价方法、数据库的实施、数据库的运行与维护。

学生课后复习时应着重于其中的第 1、2、3 点内容，为进一步学习后续章节打好基础。

实验

无