

数据库系统原理和方法

(上册)

编者：周伯鑫讲

东南大学自动控制系

序 言

计算机应用始于数值计算。却在非数值计算中得到了广泛的应用。显示了它的强大生命力。在计算机应用领域中。数据处理方面的应用占整个计算机应用的70%以上。而数据库技术是数据处理的最新研究成果。使计算机应用渗透到工业、农业、商业、文教、卫生及军事等各领域。

当代社会是一个面向信息的社会。信息(数据)管理已成为组织活动中最重要的活动之一。如何组织和管理数据。使它最大限度发挥作用。已成为不可回避的重要问题。数据库技术就是在这种背景下迅速发展起来的。但是如何有效地使用数据库系统。以便充分发挥它的潜力。仍是人们关心的重要问题。

现在有关数据库技术的教材和参考书大都面向计算机专业人员的。随着数据库系统的应用渗透到各行各业。一些非计算机专业的学生和工作人员也需要具备数据库技术的基本知识。本书可供非计算机专业的研究生作为科书。也可供广大数据库实际工作专业人员自学用。使用本书时。读者只需具有较少的数据结构“和”操作系统的基本知识。

由于编者水平有限。书中错误。不妥之处。恳请读者不吝指教。

编 者

1989. 12月

曹玲芝

目 前

第一章 数据库系统概述

§ 1.1 数据库技术发展概述	1
§ 1.2 基本概念	3
§ 1.3 数据库系统的特点	8
§ 1.4 数据库的逻辑结构	9
§ 1.5 用户访问数据库的过程	11
§ 1.6 数据库管理系统 (DBMS)	12

第二章 数据文件的物理组织

§ 2.1 外存贮器的基本知识	14
§ 2.1.1 磁带存贮技术	14
§ 2.1.2 磁盘存贮技术	18
§ 2.1.3 层贮器中存贮数据文件的最佳放置	19
§ 2.2 文件组织的基本概念	23
§ 2.2.1 OS的文件管理	23
§ 2.2.2 逻辑记录与物理记录	24
§ 2.2.3 地址与指针	26
§ 2.2.4 分页与系统缓冲区	27
§ 2.2.5 文件组织	28
§ 2.3 顺序文件	29
§ 2.3.1 串行处理文件	30
§ 2.3.2 顺序处理文件	34
§ 2.3.3 增补文件	40
§ 2.4 Hash 文件	41
§ 2.4.1 Hash 文件的概念	41
§ 2.4.2 溢出处理技术	43
§ 2.4.3 Hash 文件的设计	44
§ 2.5 索引文件	51
§ 2.5.1 稠密索引文件和非稠密索引文件	51
§ 2.5.2 树结构索引文件	51
§ 2.5.3 多分树和B树	58
§ 2.5.4 ISAM和VSAM	67
§ 2.5.5 非关键字索引文件(倒排文件)	74

第三章 数据模型

§ 3.1.1	什么是数据模型	79
§ 3.2	实体之间的联系	81
§ 3.3	层次模型	82
§ 3.3.1	层次模型的概念	82
§ 3.3.2	层次模型表示m—n关系的方法	84
§ 3.3.3	层次模型的优缺点	85
§ 3.4	网状数据模型	86
§ 3.4.1	网状模型的基本特征	86
§ 3.4.2	特殊的网络结构 循环和回路	89
§ 3.5	关系模型	89
§ 3.6	数据模型的相互关系	92

第四章 关系数据库系统

§ 4.1	关系模型的教学概念	93
§ 4.2	关系数据库上的操作	94
§ 4.3	关系代数	95
§ 4.4	关系演算	102
§ 4.4.1	元组演算	102
§ 4.4.2	关系演算的安全限制	104
§ 4.4.3	域关系演算	105
§ 4.5	关系查询语言	106
§ 4.5.1	QUEL	106
§ 4.5.2	SEQUEL查询语言	110
§ 4.5.3	查询语言QBE	116
§ 4.5.4	关系操作语言的比较	135
§ 4.6	关系数据库的模式和子模式	136
§ 4.6.1	原模式、目标模式及其映射	136
§ 4.6.2	子模式、目标模式及其映射	140
§ 4.7	查询优化	143

第五章 层次模型数据库系统

§ 5.1	基本概念	152
§ 5.1.1	层次系统的数据模型	152
§ 5.1.2	层次顺序和层次路径	154
§ 5.1.3	层次模型的模式和子模式	155
§ 5.2	IMS系统	156
§ 5.2.1	IMS概述	156

§5.3	IMS的数据存贮策略	-----	158
§5.4	PDB和LDB的构造和数据库描述	-----	165
§5.5	IMS的数据操作语言	-----	174
§5.5.1	数据语言 DL/1	-----	174
§5.5.2	DL/1 举例	-----	176

第一章 数据库系统概述

数据库技术是六十年代末七十年代的产物。它的出现使计算机广泛应用于工农业生产、商业、行政管理、科学研究、工程技术和国防等各个领域。成为不可缺少的有力工具。在计算机应用中,数据处理占着最大的比重。而数据库系统又是数据处理的核心部分。它的效能往往决定着整个计算机应用的经济效益。

选用或设计高效能的数据库是一件十分慎重和细致的工作。国内外的经验告诉我们。盲目地选用不恰当的数据库会造成巨大的浪费。往往会在系统开发花费了大量劳动后发现系统不合应用要求。这时要想更改系统设计为时已晚。所以不论是设计数据库系统还是选用已有数据库系统。都必须掌握数据库系统的基本原理。详细了解它的设计特点和使用方法。数据库才能发挥最大的经济效益。

§ 1.1 数据库技术发展概述

随着计算机技术的发展。计算机系统从用于军事和科学计算扩展到应用数据处理各个领域。1945年。第一台电子计算机ENIAC的诞生。是出于军事目的。主要用来计算弹道轨迹。那时注重于计算速度。而对数据存储没有更多的要求。1950年由J. P. Eckert和J. W. Mauckley教授设计的UNIVAC-I电子计算机系统配置了磁带机这个快速大容量外部存储器。为计算机数据处理创造了条件。当时美国人口调查局用它进行人口调查数据处理。使计算机用于数据处理迈出了第一步。但50年代数据处理主要是独立设计的简单子系统。到60年代才出了数据库的概念。我们可以把20多年来数据库的发展发展分为几个阶段或时期。

数据库的萌芽时期

这一时期为60年代初到70年代初。是数据库概念的产生。探讨研究时期。最早从事数据库实践的是C. W Bachman。他在1962年设计了一个IDS (Integrated Data Store 当时尚未“数据库”一语) 系统。该系统于1963年运行。人们把它看作为数据库的原型。其设计基础是一个灵活的模式。支持把不同类型的记录链接在一起。目的是为了更有效地利用直接访问存储设备。它给数据库的早期研究工作以很大的影响。

其次。系统发展公司 (System Development Corporation)——是美国的一个军事机构) 发起于1963年6月在加利福尼亚州的圣巴巴拉 (Santa Monica) 举行题为“计算机中心数据库研究与管理” (Development and Management of a Computer-Centered Data Base) 的讨论会。会上提出了七篇论文中。有三篇的标题涉及到数据库 (Data Base) 一词。但人们理解并不十分确切。由于Base一词在军事术语中是基地的意思。所以人们把Data Base理解为“数据的提供基地”。该讨论会分四个工作会议。其中一个会议上提出了数据库系统

义是：(1)数据库是文件的集合；(2)文件是登记项的有序集合；(3)登记项有关键字或关键字和数据组成。这就是数据库的最早定义。当然，对数据库的看法与今天比较起来，还有很大的距离。1965年9月，系统发展公司等部门又联合举办了计算机中心数据库系统的第二次讨论会，讨论了数据库系统的描述和途径。从此，数据库这一术语开始流行起来，许多从事软件工作的人陆续转移到数据库方面来。

数据库的成长时期。

真正的数据库是在60年代后期才出现的。1968年美国的IBM公司研制成功并商品化的IMS系统是最早的数据库系统。该系统采用层次数据模型，即数据在逻辑上呈树结构。在以后几年中，IMS系统又做了许多改进，形成了几个新的版本。由于IBM公司的产品竞争力强，广销世界各地，故该系统应用比较广泛，对数据库的发展有很大的影响，是层次模型的典型代表。

1969年10月美国数据系统语言会议(The Conference On Data Systems Languages—简称CODASYL)的数据库任务组(The Data Base Task Group—简称DBTG)提出了网状数据模型的数据库系统规范，即DBTG报告，把数据库技术大大地向前推进了一步。以后又陆续地发表了改进版本，功能越来越强，并以广泛使用的COBOL语言作为主语言(COBOL语言是CODASYL的早期成果)。CODASYL关于数据库的一系列工作报告，澄清了许多概念，建立了若干权威性的观点，极大地推动了数据库的发展，为数据库走向成熟奠定了基础。伴随着CODASYL的工作进程，陆续出现了一些商用数据库系统。属于DBTG型的著名系统有：在IBM 360/370、UNINAC 7090、Siemens 4004、PDP-11/45上实现的IDMS系统以及Univac公司的DMS 1100系统和Honeywell公司的IDS/2系统等。

1970年8月，IBM公司的高级研究员E. F. Codd发表了题为“A Relational Model of Data for Large shared Data Banks”的论文，提出了关系数据库的概念。文章中解释了关系模型，定义了某些关系运算，研究了数据的函数相关性，定义了关系的第三范式，开创了数据库关系方法和数据规范化理论的研究，把数学方法引入到数据库的研究中，使数据库技术有了坚实的理论基础，把数据库技术推向新阶段。同时关系数据库有良好的用户接口，使用特别方便，因此，它引起了计算机界的兴趣，成为广泛研究的课题。E. F. Codd本人作为关系数据库的创始人和奠基人，获得了1981年ACM图灵奖，以表彰他在这一领域中的卓越贡献。1975年以后，出现了商用的关系数据库系统，IBM公司研制的System R系统是其中的代表。关系数据库系统的出现促进了数据库的小型化和普及化，目前微机上出现的dBASE II和dBASE III就是例证。

数据库的成熟时期

1981年以后，数据库系统逐渐成熟起来，商品化的实用数据库系统不断增多，各个领域普遍开始应用数据库系统。关系数据库系统作为主要应用的数据库系统开始流行，各种专用数据库系统(如CAD、CAM数据库系统等)广泛应用。目前，越来越多的部门开始使用数据库技术来管理它们的业务，结果，数据处理的成本大大下降，效率越来越高，可靠性不

断增加。从而推动了计算机应用的普及。

数据库系统的研究。目前主要集中在如下几个方面。

1. 数据库的设计方法论。数据库设计方法。长期未停留在经验和尝试阶段。工程规范程度不高。缺乏理论指导。设计的好坏在很大程度上取决于主要设计者的个人知识和实践经验。数据库设计方法论包括：①数据模型。包括用户模型和概念模型的设计方法。解决从现实世界到数据库的逻辑描述问题；②数据存储与存取方法的设计。即根据模型设计数据的物理存储结构。确定用户存取数据的方式；③数据库的管理和保护。

2. 数据库的规范理论的研究。研究数据的语言问题（即数据元素之间的关系）。构造规范数据模型。使存储数据能正确反映现实世界的联系。防止产生与客观世界矛盾的结果。

3. 数据库机器。用硬件构成数据库管理系统。提高系统的运行效率。目前这方面的工作仍处于实验阶段。

4. 分布式数据库系统。它是数据技术与计算机网络相结合的产物。分布式数据库系统的设计提示了许多新问题。如数据、数据定义、管理软件在网络各节点上的分布与配置；节点之间的通信协调；分布式查询处理；并发控制；数据完整性和完全性等问题。

5. 数据库本身的智能化。逻辑演绎和语义模型是人工智能的领域。由于人工智能的研究从根本上说是如何获取、表示和利用知识的问题。而数据库的研究是如何表示、检索和利用信息。两者十分类似。因此。人们就利用人工智能的语义网络模型思想来改进数据库的数据模型。使网络的概念不仅用于描述数据模型。而且用于说明查询。目前随着各种知识库的建立。人工智能和数据库这两个领域正在相互渗透、互相促进。共同发展。

§ 1.2 基本概念

I. 什么是数据库

数据库技术发展到今天仍然没有一个完全一致的定义。我们这里提供一个通俗的说法：“数据库可以看作是贮存过去的数据为未来使用”。从这个意义上来源。数据库不一定非要计算机不可。只要它能满足未来检索和处理的要求。实际上。一般所说的数据库意味着数据贮存和处理都是由计算机来实行的。因此又称计算机数据库系统。我们也可把上述定义修改为：“过去的数据存贮在计算机系统中为未来使用”。上面两个定义抓住了现代数据库的本质。计算机数据库更完善的定义为：把数据的集合和它的描述概括地集成在一起。采取有效地管理方法。满足各类不同用户的需要”。这个主义包含下述一些概念

- (1) 数据库是概括后的数据集合
- (2) 数据集成在一起减少数据的复制。
- (3) 数据集合中包含它的描述。这种描述称为模式。
- (4) 数据集合被有效地管理。满足不同用户的需要。

这里必须对第四点作进一步的说明：为了满足不同用户的需要。数据库必须使用柔性的数据结构和存贮结构。数据库的柔性愈大。数据库在长时期内对付各种用户需要的能力愈强。数据结构的柔性依靠数据按照自然关系组织而得到的。例如。考虑数据库中父记录和子记录。

每一个父记录与每一个子记录有关。每一个子记录与父记录也有关。此外同一个记录的子记录之间也有关系。这种数据库必须充分表示所有这种数据关系。同时满足私有性和完整性约束。即为了避免未授权的用户使用以及出现不希望的修改而给出的使用约束。为了有效地进行管理。需要支持和提供有效处理能力的各种存取路径和结构关键字索引。指针等)。在模式中给出数据和存贮结构的有关描述。为了满足各种程序使用。数据库在运行期间由独立的软件进行管理。这个软件叫做DBMS。即数据库管理系统。

II. 数据面临的问题

计算机数据处理系统由四个基本要素。它们是机器。程序。数据和人。机器和程序又称硬件和软件。

早期的数据处理方式是数据与使用它的程序以及存贮它的机器紧密相关。当时程序是人们关心的中心。数据仅是应用程序开发的副产品。这种紧密关系带来了一系列问题。例如各种应用的数据共享问题以及不同机器和系统之间数据传送问题。当新的系统安装代替原有的硬件和软件时。需要把原有程序和数据转换到新系统上去。但是在使用计算机的早期。人们集中力量优先解决程序和机器独立开来的问题。结果出现了高级语言。随着计算机使用的逐步推广。人们认识到数据转换问题比过去程序转换问题更广泛。新机器安装由于转换数据带来的开销愈来愈大。人们开始认识到数据与机器独立的重要意义。此外数据依附于应用程序带来不少弊病。具体表现为：

(1) 数据冗余性 (Redundancy)：多个应用程序共享数据必须重复复制。造成存贮空间的浪费。

(2) 数据不一致性 (Inconsistency)：由于数据的重复复制。当数据修改时。必须对复制的数据同时进行修改。如果某一处复制数据忘记修改。这就出现数据的不一致性。

(3) 增加了应用程序的维护费用：由于数据依附于应用程序。数据修改必然造成应用程序的相应修改。使应用程序不易维护。

人们从程序和程序语言与人建立相对紧密关系而与机器建立相对独立关系中获得启示。需要把数据与程序和机器的紧关系中解脱出来。建立程序和数据彼此更独立的关系。以及两者和机器的独立关系和人之间的紧密关系。

III. 数据库管理的目标

数据库管理能概括为组织内部数据资源的控制和应用。控制涉及数据库的维护和确保数据库的质量。限制授权人员对它的使用。追求数据库的完整性 (Integrity)。数据资源的应用指有效性的目标。包括现有资源的共享和增强未来的可用性。

(1) 共享性 (Sharability)

共享数据资源是数据库的基本目标。这意味着不同的人。不同的处理在同时使用相同的数据。

由于数据与组织内各方面有关。它属于整个组织而不是属于任何个人。因此数据库提供

共享的存取系统与单用户拥有的文件分时系统是完全不同的。在组织中，共享文件是一般情况。私有文件是例外情形。

共享数据库环境要求集中控制和协调数据集合包括数据的使用，并综合数据存贮，增加数据的一致性，减少冗余，便于数据维护。

共享性的目标包括：

- (a) 为各级不同类型的用户服务；
- (b) 处理相同存贮数据的不同用户视图；
- (c) 联结相互有关的数据；
- (d) 设置标准；
- (e) 控制并发修改，维护数据完整性；
- (f) 协调多用户之间的重启动和恢复操作。

由此可见，管理共享数据，提出了一些附加的问题。共享的中心内涵是调和用户之间的冲突的需求。（其中包括数据结构和存贮结构方面的要求）。

(2) 有效性 (Availability)

有效性意味着把组织内的数据带给该数据的用户。管理数据资源的系统对组织内的人应该易于使用的。即当什么时候什么地方需要时就按需要的方式和形式，提供合适的数据库。有效性既指数据又指提供数据的数据库管理系统。有效性目标包含两个方面。一个是功能，一个是形式。

功能的有效性使数据库适合用户定义和创造以及把数据存入数据库和从数据库取出数据。这些功能是由数据库管理系统直接实行的。

形式多样性包括系统结构方便地、及时地操作，在多样化的用户和多样化的操作环境²以经济的方式。应用多种语言，满足各种使用模式，存贮各种数据。

图 1-1 列出了数据库有效性的有关方面。

(3) 演变性 (Evolubility)

演变性指数据库能够随用户需要的增长和技术的发展作相应的改变。

技术的迅速发展对计算机结构（如平行处理，阵列处理，管道结构处理等）带来革命性变革。这就为索引技术和快速检索算法提供了可能性。数据库机的出现亦可能改变用户与数据库管理系统 (DBMS) 会话的方式。

计算机使用的更加普及，计算机网络的出现，计算机之间的数据交换愈来愈频繁。人们普遍关心数据库管理的基本功能，着手研究数据的可传递性³问题和解决数据交换中出现的问题，制定各种规范和标准。

众所周知，一旦系统装配，数据按它自己的规则定义和存贮后，要改变是非常困难的。数据库的设计应当使数据库重定义和重建的技术障碍、费用和破坏的风险减到最小。

除掉技术进步的因素外，用户的需求也是不断改变的。随着市场环境的变化，组织内人员和设备的变动，经营策略的改变，贮存数据内容和数量以及数据的处理方式也会随之改变。数据库的设计应当满足用户使用需求的变化和增长，否则数据库将失去自身的生命力。

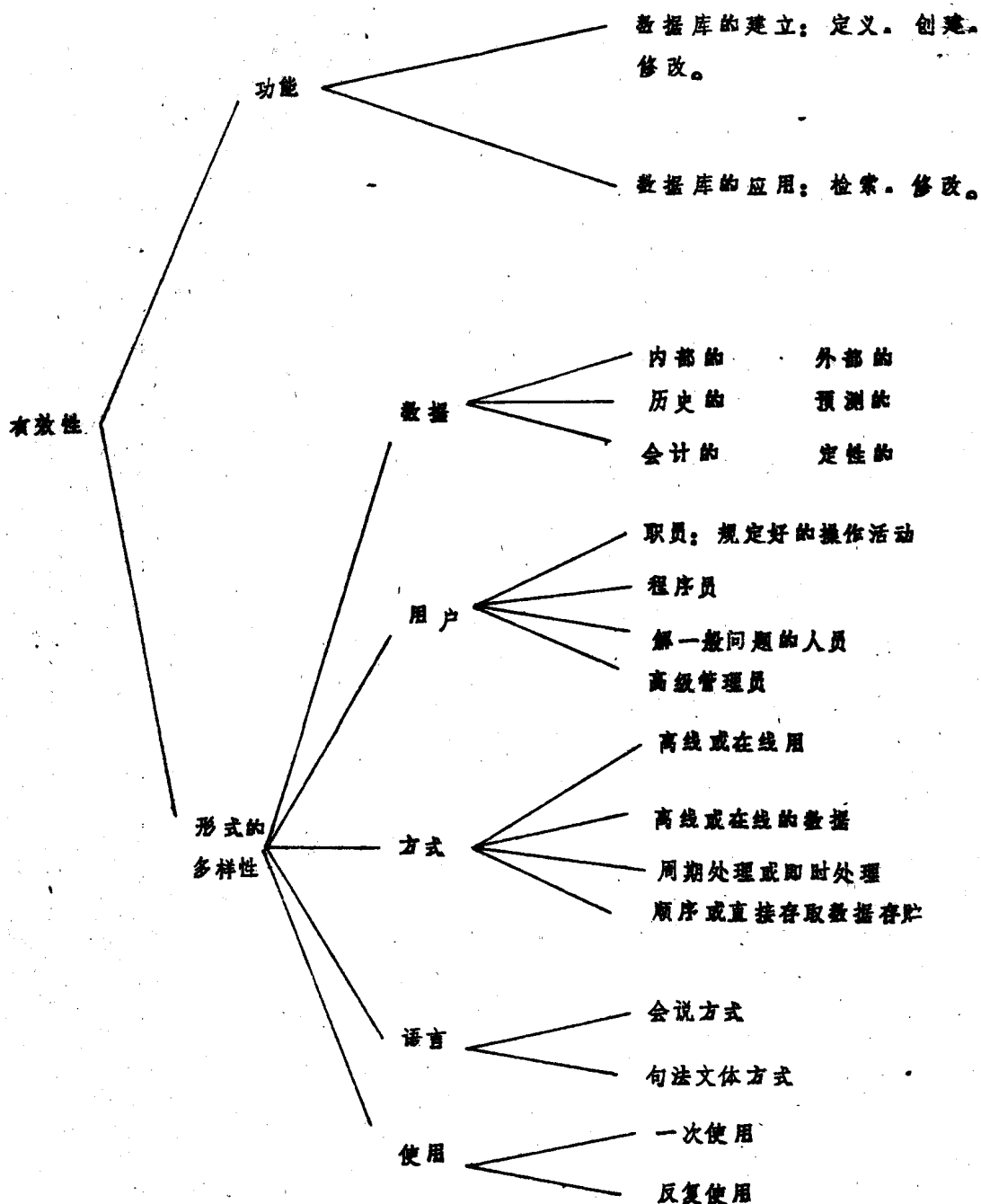


图 1-1 数据库有效性的诸方面

(4) 数据库的完整性 (Integrity)

数据库完整性有三个方面的内容:

- (a) 保护现有数据库的存在;
- (b) 维护数据库的质量

(c) 确保数据库数据的私有性 (Privacy)

数据库完整性的诸方面如图 1-2 所示。

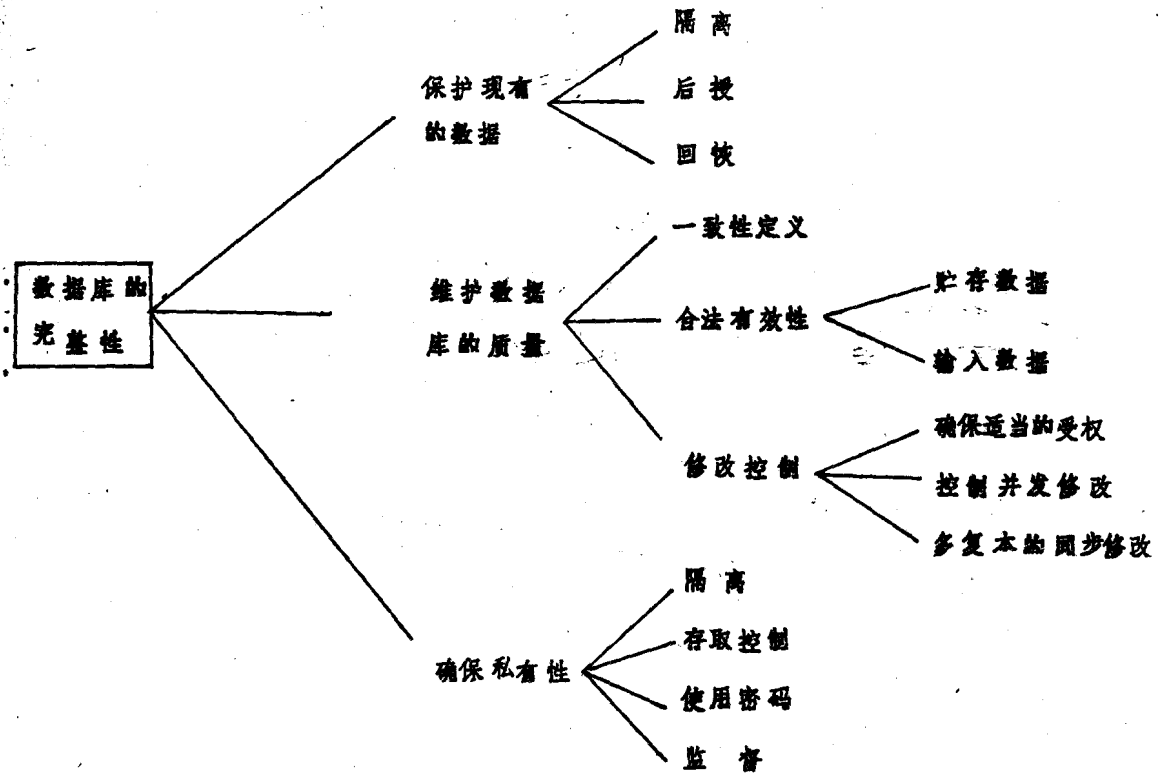


图 1-2 数据库完整性的有关方面

维护数据库完整性涉及数据库管理系统 (DBMS) 和数据库管理员两方面有关。在任务执行期间, DBMS 工作。对数据库内数据存取和操作的所有请求集中交给系统。当系统响应请求时, 它验证请求者的授权。确保贮存或者修改的。所有数据与它的定义一致。并进行审计和监督等。DBMS 是数据库管理员的“代理人”。它实行 DBA (数据库管理员) 预先规定的策略。

以往, 数据完整性的责任交给程序员或数据分析人员。系统对数据完整性不起什么作用。只是做些数据传输和贮存过程中的奇偶校验工作。

由于共享增加和数据价值的认识使人们认识到对数据清楚规定允许的操作具有重要意义。对数据资源的控制赋予授权负责人比程序员更好重要。

数据库绝对完整性是不可能做到的。DBMS 应当提供一些可达到的选择实现某程度的数据完整性。数据库完整性程度愈高, 代价也就愈高。数据的价值以及数据损失和误用造成的危害决定实行各种保护措施的范围。下节概括数据库系统的特点。

§ 1-3 数据库系统的特点

第一个特点：系统内有一个庞大的数据库。其中存放了企业或组织内各工作所需要全部数据。这些数据具有下列性质：

(a) 数据共享性 数据库中的数据可以同时为多个不同类型的用户或应用程序所使用。

数据集中管理的优点：

(1) 减少数据的重复（或冗余）

(b) 数据的独立性 应用程序不依赖于数据库中的数据组织方法和存放位置。它包含两方面的含义：①不同的应用程序均可以按自己所需的数据结构去询问数据库中的数据。②数据库中数据的内部组织发生变化时不需要重新编制或修改应用程序。

(c) 数据安全性和完整性 数据安全性指系统对数据库中数据存取权的控制。系统保证数据的机密和安全。防止无关用户对数据的非法存取及有意或无意的破坏。数据完整性是指系统保证数据库中数据的正确性、有效性和兼容性。系统提供各种保护手段（和数据转存、差错检查和修复），停止任何可能危害完整性情况的发生。

数据库第二个特点是：用户类型的多样性。按照使用数据库中数据的方式和所承担的责任，可分为四类用户。

数据库管理员 DBA (Administrator) 这是由负责数据库全面管理的人员组成。他们的职责包括：决定数据库的信息内容及其逻辑结构。决定物理存储结构和存取方法。规定用户权限及其检验策略。选择后备策略和恢复措施。监督系统运行状况。改进系统性能等。由此可见，数据库系统除了提供用于描述数据库整体数据结构的数据描述语言外，还应提供修改模式的语言。规定数据物理存储的语言。以及装入、监督、统计测量、转存、恢复和再组织等服务程序。

应用程序员 他们通过程序设计语言（如 COBOL、PL/1、FORTRAN）编写的程序与数据库打交道。为了便于这类用户按照自己所需的数据结构去描述和操纵数据库中的数据，数据库系统为他们提供了子模式描述语言 DDL (Data Descriptive Language) 和数据操纵语言 DML (Data Manipulation Language)。用户把这两种语言嵌套在高级程序语言中使用。因而 DDL 和 DML 又称为数据子语言。而普通程序设计语言称为主语言。

非程序员用户 一般为管理人员和领导人员。它们不熟悉程序编制方法，适合采用简单直接的方法（如填写表格或简易的终端查询命令等）与数据库系统打交道。

参数用户 这类用户利用预先登入系统的应用程序去与数据库系统打交道。用户只要从终端或其它输入装置送入相应参数，即可要求系统完成用户所需的工作。

数据库的第三个特点是：系统要配置一个数据库管理系统 (Data base Management System, 简称 DBMS)。它构成系统运行的核心。作为用户询问数据库的接口。DBMS 接收、分析、并解释用户提供的命令请求。然后转到相应处理程序去操纵数据库中的数据。（检索、存贮、删除、更新等）。

§ 1.4 数据库的逻辑结构

§ 1.4.1 数据库的三级抽象描述

“关于计算机和信息处理的美国国家标准委员会”及其“标准计划和需求委员会”ANSI/X3/SPARC(American National standards Commsttes on Computer and Information Processing and standards Planning and Requirement Committee)在1975年发表研究报告中认为:一个数据库从逻辑上可分为三级:外模式(External schema)、概念模式(Conceptual schesma)和内模式(Internal schema)它们之间的关系如图1-3所示。当今数据库文献虽然对数据库有关的术语各不相同,但在结构上都承认这个关系。因此掌握数据库三级结构及其联系和映射,是深入了解数据库原理的必由之路。

数据库结构的三级抽象描述定义了数据库的三个层次,反映了看待数据库的三种不同角度。

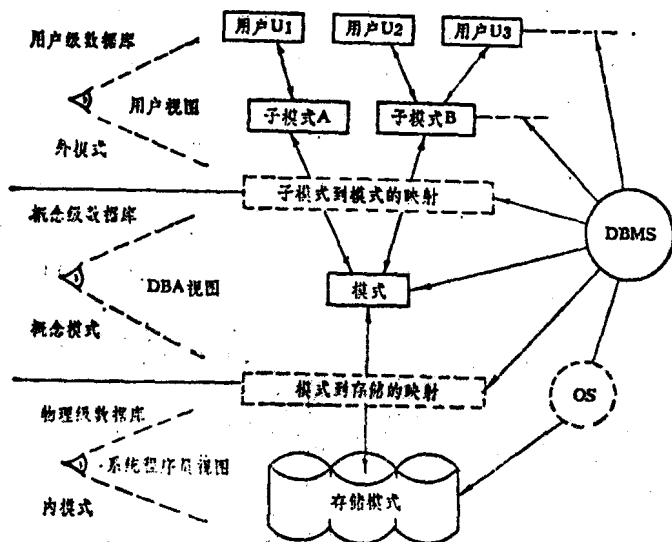


图 1-3 数据库的分级结构

1. 物理数据库:它是物理存储设备上实际存储的数据集合。它是数据库的最内层。这些存储着的数据是用户加工的对象。从机器的角度看,它们是指令操作处理的位串、字符和数字。从系统程序员角度看,这些数据是用一定的文件组织起来的一个物理文件(称为存储文件)。系统程序员编制专门的存、取程序,实现对文件中数据的存取。

所以物理数据库也称为系统程序员视图(或数据存贮结构)。它由数据的物理模式定义。它与物理设备的特性密切相关。

2. 概念数据库:它是数据库结构中的中间一层。它是数据库的整体逻辑表示。规定每个数据的逻辑定义及数据间的逻辑联系。

概念数据库并不是实际数据库。它不涉及实际数据库的存储。其目的是为了把所有用户

视图有机地结合成一个逻辑整体。它涉及的是数据库所有对象的逻辑关系。它是数据库管理员 (DBA) 概念下的数据库。故称为数据库管理员的视图。

3. 用户级数据库: 是数据库的最外层, 对应于外模式。是用户所看到和使用的数据库。因而称为用户视图。

它逻辑上表示一个或一些特定用户的数据集合, 即用户记录的集合。它由数据子模式定义。通常一个具体用户或一个特定的应用。只对数据库中某一部分的数据感兴趣。数据库管理员只能让他们访问那些与他们有关的数据。数据库中的其余数据对他们来说是透明的。也就是看不到的。这是对数据库实施安全保护的一项有力措施。

用户级数据库是概念数据库的一个子集。对于同一数据, 用户级数据库与概念数据库可以由不同的逻辑定义和组成。由数据库管理系统负责它们之间的自动转换。

例如在人事管理时, 人员的年龄在用户级数据库和概念数据库可以定义不同的域和组成概念。由于年龄时刻在变, 在概念数据库中不适合。宜放生日。而用户感兴趣的是年龄。因此, 可以把现在的日期减去概念数据库中生日成为用户感兴趣的年龄。这个转换由服务程序自动生成。

对于一个数据库系统来说, 实质上存在的只是物理数据库。概念数据库是物理数据库的一种抽象 (逻辑) 的描述。用户数据库只是用户需要的数据库子集的逻辑描述。是用户与数据库的接口。

用户通过子模式进行操作。通过子模式到模式的映射与概念数据库联系起来。又通过模式到存储的映射与物理数据库联系起来。所谓映射是一种对应的规则。规定映射双方如何转换。

数据库的三级结构就是靠映射联结和完成的。映射又是实行数据独立性的保证。

如果物理数据库发生变化, 只要相应地改变物理数据库与概念数据库之间的映射, 就可以保证概念数据库不变。那么用户级数据库可以保持不变。从而保证应用程序不发生变化。实现了数据的物理独立性。

至于数据的逻辑独立性, 由概念数据库与用户级数据库之间的映射保证。使应用程序可以按照自己的数据结构去询问数据库。只是各应用程序的用户数据库到概念数据库的映射不同而已。改变应用程序只要改变这一级的映射, 与物理数据库无关。

当然各级映射的维护是件困难而复杂的事情。但换来了用方的方便。数据库结构中的三级以及它们之间的映射都是由数据库管理员负责管理和维护的。

§ 1.4.2 模式 (schema) 和子模式 (subschema)

I. 模式

模式是数据的整体描述。体现了数据库所采用的数据模型。数据库系统描述还包含寻址方式、访问控制、保密定义、安全完整度等内容。

模式表示有图示与语言两种。前者画出数据库的模式图。非常直感。但不严格。可以忽略许多细节。通常只在模式设计的第一阶段使用。语言描述则严格、准确。可以陈述许多细

节。能为计算机接受。故通常在模式设计的第二阶段使用。用语言书写的模式称源模式机器不能直接使用。必须将源模式用机器代码表示。变为机器使用的模式。称目标模式。

II 子模式 子模式是用户与数据库的接口。因此每个用户均必须使用一个子模式。多个用户也可以使用同一个子模式。从模式可以用某种规则导出子模式。在关系方法中。通过关系运算就可以从模式导出子模式。子模式到模式之间的对应关系称为子模式到模式的映射。有的系统把这种映射也写在子模式中。但必须指出。子模式可以作出不同于模式的改变。子模式也有图示和语言描述两种形式。给用户以子模式有如下好处：

(1) 接口简单。使用方便。用户只需按子模式编写应用程序或在终端键入操作命令。无需了解数据的存储结构。

(2) 保证数据的独立性。应用脱离了数据的物理组织。使数据的物理组织方式改变或引入新的存储设备无需修改或重编应用程序。

(3) 提供数据共享性。用同一模式产生出不同的子模式。减少了数据冗余度。

(4) 孤立数据。安全保密。用户通过程序只能操作其子模式范围内的数据。从而使程序错误传播的范围最小。同时保证了其它数据的安全性。

§ 1.5 用户访问数据库的过程 (图 1-4)

现以用户通过应用程序存取一个记录为例。说明用户访问数据库数据过程中的主要步骤。

(1) 应用程序A向DBMS发出读一个记录的命令。程序给出记录类名及欲读记录的关键字值。

(2) 检查A的存取权限。决定是否执行A的命令。

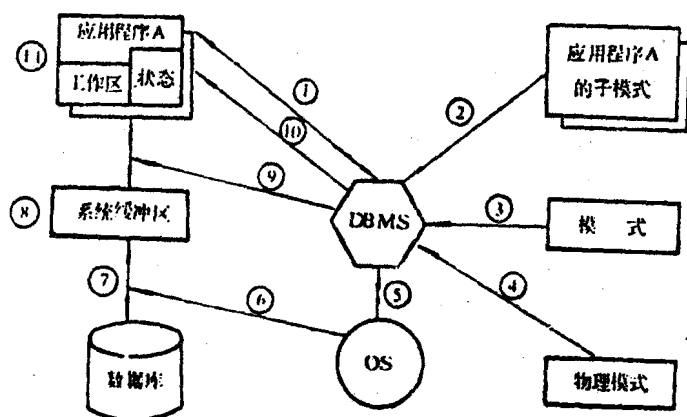


图 1-4 访问数据库的过程

(3) 在决定执行A的命令后。DBMS调用模式。根据子模式 $\longleftrightarrow$ 模式变换规则。决定读入那些模式记录；

- (4) DBMS 查看物理模式。确定应该读入哪个或哪些物理记录;
- (5) DBMS 的访问程序找到有关的数据块的地址。向 OS 发出读进块命令;
- (6) OS 执行读的命令。
- (7) 把所要求的块 (或页面) 从外存贮器送到系统缓冲区。
- (8) DBMS 比较模式和子模式。导出用户程序需要的记录形式。
- (9) DBMS 把数据从系统缓冲区送到用户程序 A 的工作区。
- (10) DBMS 向用户程序 A 送命令执行情况的状态信息。
- (11) 应用程序 A 使用工作区的数据。

如果要进行其它操作。其过程与读记录类似。若修改数据。首先把要改的记录读到用户工作区。由用户程序在工作区中进行修改。然后向 DBMS 发出写记录的命令。其过程与读相反。

§ 1.6 数据库管理系统 (DBMS)

DBMS 是一个非常复杂的系统软件。它的职能就是有效地实现数据库三级之间的转换。它的工作归纳为以下三方面:

- (1) 描述数据库模式。把用数据描述语言所描述的数据库模式和子模式从原形式转换成目标形式。存放在数据库中供系统查阅 (也称为数据库字典)。
- (2) 管理数据库。包括控制整个数据库系统的运行; 控制用户的并发访问; 执行对数据的安全、保密、完整检验; 实施对数据检索、插入、删除、修改等操作。
- (3) 维护数据库。其功能包括: 初始时装入数据库; 运行时记录工作日志。监视数据库性能; 当性能变坏时重新组织数据库; 在用户要求或系统设备变化时修改和更新数据库; 在系统软、硬件发生故障时恢复数据库等。
- (4) 数据通讯。负责处理数据流动。这些数据可能来自应用程序、计算机终端或其它系统。也可能由系统内运行的进程所产生。它们可能被送到调用队列缓冲区。终端或正在执行的某个进程中。这部分工作通常与操作系统协同完成。

从程序的角度看。DBMS 实际上是系统程序所组成的一个集合。每一个程序完成一定的功能。其主要程序为:

- (1) 语言处理方面有模式 DDL 翻译程序 (把模式 DDL 源形式翻译成目标形式)。子模式翻译程序 (把子模式 DDL 源形式翻译成目标形式)。终端询问解释程序。控制命令解释程序等。
- (2) 系统运行方面有系统控制程序 (控制、协调 DBMS 各个程序活动。使其有条不紊地运行)。访问控制程序 (通过核对标准、口令、授权、检验访问的合法性)。并发控制程序 (在多用户同时访问时。协调各用户访问)。保密控制程序 (在执行操作前核对保安规定)。数据完整性控制程序 (核对数据库完整性约束条件)。数据访问程序和通信控制程序等。
- (3) 系统建立、维护方面有数据装入程序、工作日志程序。性能监督程序。系统重组程序和系统恢复程序等。