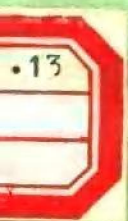


数据库系统原理

崔亨洙 编著



吉林大学出版社

数据库系统原理

崔亨洙 编著

数据库系统原理

崔亨洙 编著

责任编辑：杨继奎

封面设计：张沐沉

吉林大学出版社出版
(长春市东中华路 29 号)

吉林省新华书店发行
长春大学印刷厂印刷

开本：850×1168 毫米 1/32

1992 年 5 月第 1 版

印张：9.875

1992 年 5 月第 1 次印刷

字数：245 千字

印数：1—1900 册

ISBN 7-5601-1159-9/TP·18

定价：3.25 元

序 言

本书系统地叙述了数据库系统的基本概念、基本原理、基本理论和基本方法。本书的前三章是概论，着重阐述了一般性的概念、原理和方法。第四章介绍了网状模型数据库系统 DBTG 的基本概念和方法，第六章介绍了层次模型数据库系统 IMS，第七章介绍了关系数据库系统中的基本概念、原理和方法。第八章是关系数据库设计理论，第九章讨论了数据库的安全性、完整性、并发控制和数据库恢复问题。在第十章中提纲挈领地叙述了数据库结构设计的步骤和方法概要。因为数据库设计应当是一门独立的课程，在短短的几个学时内是不可能讨论很深入的，因此在本书中没有进行深入细致的讨论。至于第五章是为了帮助读者深入了解在实际运行的 DBTG 系统中的情况，它有助于理解 DBTG 的基本概念、原理和方法。

本书的初稿是为吉林大学计算机科学系软件专业开设的数据库系统基础课程而编写的讲义。在本书的编写过程中笔者试图贯彻着重阐述基本概念、基本原理、基本理论和基本方法的原则，因此，避开了某些方面的细节，例如，关于应用程序设计方面的讨论就较少。

由于编著者的水平所限、改写时间仓促，在本书中难免有缺点和不妥之处，恳请读者指教。

崔亨洙

1990年11月

于吉林大学

第一章 数据库和数据库系统

在科学、技术、经济、文化和军事等各个领域里，我们会遇到大量数据，这些数据是复杂的，它的数量在不断增长，而科学、文化、经济、军事等一切领域的发展都离不开数据，因此如何科学地管理数据是一个极为重要的课题。

数据库技术是使用计算机来管理数据的一门最新科学技术。经过二十多年的研究和实践，数据库技术已发展成一门完整的学科，它具有较系统的理论和很丰富的内容，根据这些理论和方法，人们研制出多种数据库管理系统并使用数据库系统科学地管理大量数据，它们正在为各领域的发展发挥重要的作用。

本章将介绍数据库和数据库系统的一些基本思想方法和基本概念。

1.1 数据及其管理

1.1.1 数据和数据管理

世界上存在着各种各样的现象(Phenomena)，我们把这些现象的描述叫做数据(Data)。数据对应于关于现象的记录好的事实(Fact)，而从这些现象中我们可以获得关于世界的信息。信息(Information)是日益增长着的知识(Knowledge)，这些知识是从数据得到的。数据(Datum，其复数形式为 Data)一词来自拉丁文，它的原意是事实。有时，我们也需要描述观念(Idea)，因此我们把数据概念进一步扩充，数据对应于任何现象或人们

值得形式化和记录的观念的描述。

通常，数据是通过某种通讯方法(比如图象、语言)在某种记录媒介上被记录下来的。信息和数据是互相联系的概念，但有一定的区别。如果我们把客观世界的某种现象或观念所反映的知识用一定方法描述出来，那么前者是信息而后者是数据。

因为信息和数据都是现象和观念所反映的知识，这是它们的共同点，因此当不需要严格区分时，我们把这两者不加区分地使用，如“数据处理”与“信息处理”是同义的，“数据资源”和“信息资源”是同义的。

信息是一种资源。我们常说能源是一种资源，材料是一种资源，同样信息是一种资源，而且是极为宝贵的资源。首先，信息在客观世界大量存在，当今世界信息来自各个方面，在生产活动、科研活动中，通过各种渠道，人们获得各种各样的大量的信息。其次，信息是非常丰富的。再者信息对人们的生产、科研等活动是非常有用的。信息作为一种资源，它随着时间的推移，越来越丰富，显得越来越重要。

信息资源的重要性是不容置疑的。但人们对信息资源的重要性的认识是经过很长时间的实践才逐步加深的，真正认识到信息资源的重要性还是近几十年的事情。随着科学技术的发展，生产力大大提高，在经济、文化、军事等各方面带来迅速发展，这样需要掌握大量的信息，研究和分析这些信息，并从中得出有用的结论，再把它应用到社会生产活动中，是必不可少的事情。电子计算机的问世和发展给人们提供了用计算机管理和处理信息的可能性。人们使用计算机管理和处理信息的同时进一步开发了信息资源，利用信息资源进一步推动生产发展和社会发展。人们通过这些活动加深了对信息资源重要性的认识。

来自各方面的信息是非常丰富的，迅速增长的。我们把这些信息，收集、加工处理。我们把数据的收集、组织、存贮、传

播、检索、分类、加工、计算、打印报表、输出等一系列活动总称信息处理或数据处理。数据处理的是对大量的、不易理解的数据，按一定方式加工和整理并从中抽取和推导出对某些方面有价值和有意义的信息，向人们提供有价值的部分，把它作为依据，作出决策。

在数据处理的一系列活动中，数据的收集、组织、存贮、传播、检索、分类等活动是基本环节，这些基本环节统称为数据管理或信息管理。数据处理中，对数据的加工、计算、打印报表等操作对不同的业务部门可以有不同的内容。

数据处理的历史已久。早在 19 世纪末已经使用机械方式处理数据，如使用卡片制表机编制人口普查表。20 世纪 40 年代中期电子计算机问世，就提供了用计算机进行自动化数据处理的可能性，以后随着计算机科学技术的迅速发展，人们对大量的数据进行高速的自动化数据处理。用高速的大容量的电子计算机处理数据可以办到用机械方式处理数据难以办到的事情。在高速的电子计算机上配备功能完善的软件，用它来进行数据处理，这是数据处理领域里的革命，而在数据处理领域里广泛采用数据库技术，数据库成为数据处理系统的核心部分。

1.1.2 计算机数据管理方法的演变过程

计算机数据管理方法的演变过程大致经过了三个阶段：人工管理阶段、文件系统阶段和数据库系统阶段。

1. 人工管理阶段 人工管理阶段是从本世纪 40 年代中期电子计算机问世到 50 年代中期这一阶段。这一阶段计算机主要用于科学计算。这一阶段计算机的硬件和软件还不够完善，速度低、内存小、外存还没有磁盘等直接存取存贮设备，没有操作系统，没有管理数据的软件。

在人工管理阶段，数据管理的特点是：

(a) 数据不保存 需要时把数据输入,用完就把数据撤走,数据不保存在计算机里.

(b) 没有数据管理的软件系统 程序中不仅规定数据的逻辑结构,还要设计物理结构、存取方法、输入输出方式等,这样数据和程序不具有独立性.

(c) 数据是面向应用的 一组数据对应于一个程序,程序之间可能有大量重复数据,一组数据面向一个程序.

2. 文件系统阶段 从50年代中期到60年代中期是文件系统阶段.这一时期计算机软硬件比过去有大的发展,已经有了磁盘、磁鼓等直接存取存贮设备.软件有操作系统,其中有专门管理数据的文件系统或信息管理功能.数据管理技术已有较大的发展.这一阶段的特点是:

(a) 数据可以长期保存在外存贮设备上 这样就可以多次对文件进行检索、修改、插入、删除等操作.

(b) 有数据管理软件 有文件系统这样的专门管理数据的软件.

(c) 文件的逻辑结构和物理结构之间有一定的区别.

这一阶段的主要缺点是:

(a) 文件还是基本上对应于一个或几个应用程序,数据还是面向应用的.

(b) 数据独立性不高.

(c) 冗余度大,浪费存贮空间.

3. 数据库系统阶段 到60年代后期,计算机的软硬件进一步得到发展,速度高容量大,已配备了大容量磁盘,各种软件系统进一步完善,而且需要管理的数据量急剧增加.人们在数据管理方面已积累了丰富经验,数据管理技术研究取得了很大进展,到60年代末期出现了数据库系统.它具有数据的集中统一管理、数据资源的共享、数据结构化、合理而复杂的数据模型的应用、数据冗余小、数据独立性高等等特点.关于数据

库技术的特点的详细讨论将在另一节里还要进行，这里不再详述。

1.2 数据单位

1.2.1 实体和属性

人类的认识来自客观世界，客观世界存在着的事物反映到人们的脑子里，形成概念。我们把实际存在着或能想到的客观的现实的事物叫做实体(Entity)。实体是客观事物的反映。

事物具有各种特性，它们反映到实体中，我们把它们叫做实体的属性(Attribute)。

房子、职工、学生等都是实体。学生的某些特性，如学生号、姓名、年龄、性别都是属性。学生这样的实体是由这些属性来刻划的，于是可把学生这个实体描述为这些属性的集合，这里学生号、姓名、年龄、性别等不是具体的，因此这样描述的实体是一个实体型。如果学生号为S10，姓名为张三，年龄为19岁，性别是男的，那么S10、张三、19、男的集合是相应于上述实体型的一个具体的例子。

属性值是取自一定范围内的，如性别只能为男或女，年龄取自1到200范围内。

计算机处理的是数据。实体反映为记录，记录有记录型和记录值之分。

1.2.2 数据单位

数据有不同的层次和各种不同的数据单位。下面我们来介绍一下数据的逻辑单位。数据单位叫法和含义在不同的文献和数据库系统中有些不一样，这里只是介绍较普遍使用的一些数据单位术语。

数据项(Data item) 数据项是命名的数据的最小单位,数据项是不可分割的数据单位.它用来描述事物的特征,因此它是描述属性的数据单位(这里我们把属性理解为不可分割的).例如学生的学生号S#、学生的姓名SN、学生的年龄SA、学生的性别SX都是数据项,它们分别对应于学生这个实体的属性:学生号、姓名、年龄和性别.在计算机系统里具体描述数据项时,我们用某种字符描述,如果允许汉字,那么也可以用汉字.

数据项有型和值之分,型是总的描述,值是一个具体的例子.例如S#,SN,SA,SX是型,而S10,张三,19,男是数据项的值.

数据项也叫做基本项(Elementary item)、字段等.

组项(Group item) 组项是数据项的命名的集合.在某些系统里组项中可以包括组项.比如“出生日期”,如果在这个名下再分成年、月、日,并把它们作为数据项(即基本项)来定义,那么“出生日期”就是一个组项.一个数据单位是基本项还是组项,那要看具体的环境和定义.如果在一个系统里,把出生日期看做是不可分割的,那么它是数据项,如果把它看做是年、月、日可以分开来处理的,并把年、月、日分别定义为数据项,那么出生日期便是一个组项.又比如学生的姓名SN,如果把姓和名定义为数据项,一个是姓,另一个是名,那么由应用环境的需要,可以定义一种组项,叫做“姓名”,姓名下分两个数据项.如果没有必要把姓和名分开,那么“姓名”便是一个数据项.

记录(Record) 记录是数据项或组项的命名集合.记录描述现实世界中的事物,因此它是对应于实体的一种概念.一个记录中的数据项或组项自然是跟这个记录相关的,不能随便把某些数据项安插进去.如在“学生”记录中,可以有学生号、姓名、年龄、性别、地址等等数据项,但不能有与它毫无相关的“数量”、“数量单位”等.只有在有组项这个概念的系统里,才

可以在记录中出现组项。

记录有记录型和记录值之分。记录型是数据项名(或组项名)的命名的集合, 而一个记录值是一个实例。

文件(File) 文件是同一个类型(即一个记录型)的记录值的命名集合。例如, 图 1.1 表示名为“学生文件”的一个文件例子。

学 生 文 件

S#	SN	SA	SX
S1	张三	19	男
S2	李四	20	男
S3	王五	18	女
S4	赵二	21	男
S5	李平	19	女

图 1.1

在这个文件中(S1, 张三, 19, 男), (S2, 李四, 20, 男)等每一行就是一个记录值。

1.3 数据库

数据库技术是科学地管理数据的一门科学技术, 自然这里指的是用计算机来管理数据。

要把大量的互相有联系的数据收集、整理并存贮在计算机外存如磁盘上面(如图 1.2), 这些数据是按一定的结构组织起来的, 对它可以随时检索、更新, 还可以维护。这些数据是共享的, 是为用户服务的, 应当为用户提供方便, 用户可以多种方式访问数据, 不同的用户可访问数据的范围可以不一样, 一个特定的用户以某种方式能访问与其有关的数据, 一个用户只能了解和他有关的那部分数据, 只能访问允许的那部分数据。

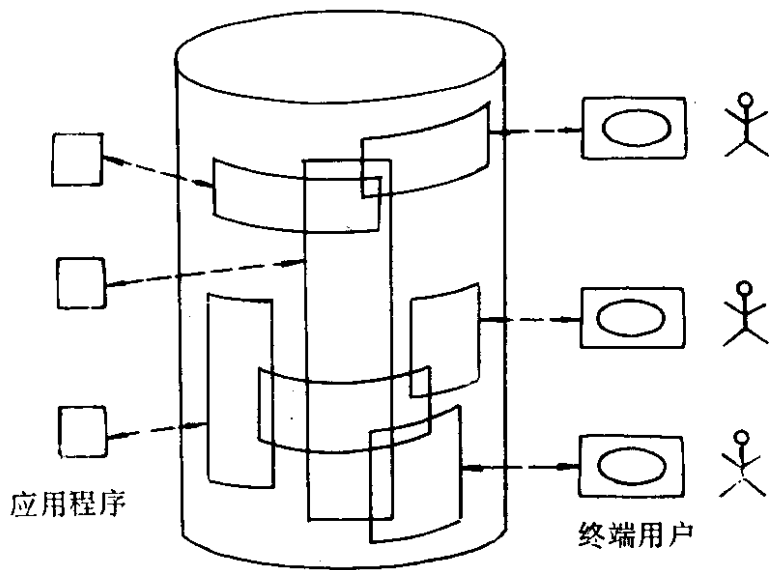


图 1.2

这好比是一个窗口，这个窗口就是用户的视图(View)。

数据库的数据是存贮在外存上的，数据的逻辑组织方式和物理组织方式应当适应查找方便，省空间，容易更新和维护。

数据库中的数据是由叫做数据库管理系统(Data Base Management System——DBMS)的软件来管理的，而DBMS是在操作系统支持下工作的。

数据库技术正在迅速发展，现在还没有大家公认的完整的严格的数据库这个概念的定义。但也有一些定义，例如，J. Martin 在《计算机数据库组织》一书中对数据库下了一个定义，我们觉得它反映了数据库这一概念的含义和特点。J. Martin 的定义如下：

可以把数据库定义为存贮在一起的相关数据的集合，这些数据无有害的或不必要的冗余，为多种应用服务；数据的存贮独立于使用它们的程序；对数据库插入新数据、修改和检索原有的数据均能按一种公用的可控制的方法进行；数据被结构化，为今后的应用研究提供基础。

1.4 数据库管理系统

数据库管理系统是管理数据库的软件，它处理数据库数据的定义、存取、管理、控制等等各方面。应用程序对数据库的操作当然在 DBMS 的支持和控制下进行。DBMS 是数据库系统的核心。

DBMS 的功能主要包括如下的几个方面：

1. 数据库的定义功能 它用来描述数据，包括整体逻辑结构的定义、局部逻辑结构的定义、物理结构即存贮结构的定义、保密规定等等。

2. 数据库管理功能 这一功能包括控制整个数据库系统的运行、数据检索及插入、修改、删除等更新管理、安全性和完整性以及并发控制等。

3. 数据库的建立和维护功能 这一功能包括数据库的建立、数据库更新、数据库再组织、数据库结构维护、数据库性能监视和数据库恢复等。

4. 通讯功能 它包括应用程序与 DBMS 之间的通讯等。

DBMS 是大型软件，现在研制的计算机系统几乎都配备 DBMS，所以可以把它看做是一种重要的系统软件。

DBMS 是在操作系统的支持下工作的。

1.5 数据库系统

1.5.1 数据库系统

我们知道，在数据库中存放的是数据，而这些数据是按一定方式组织起来并存放在计算机存贮器上的，这些数据是提供给用户使用的，可以检索也可以更新。于是首先应当有适当的

计算机硬件，还要有某些软件。一个数据库系统要包括四个主要部分：数据、硬件、软件 and 用户。因此，数据库系统是计算机硬件、软件、数据库和有关用户的总称。

硬件包括计算机主机、辅助存储器(磁盘、磁带等)、通道以及其它一些附属设备。计算机内存的容量应当足够大，以便存放操作系统、DBMS、应用程序等，而且能够开辟用户工作区和系统缓冲区等空间。对外存的要求是能够具备大容量的直接存取外存设备，以便存放数据库数据。此外还要求较高的通道能力。

在软件方面，主要包括操作系统、数据库管理系统、编译系统、应用程序和应用软件包。

数据是存储在辅助存储器上的。数据是以数据库的形式组织起来的。当然根据需求和可能性可建立一个或多个数据库。

用户有多种，我们考虑三类主要用户，第一类是应用程序员(Application programmer)，第二类是终端用户(End-user)，第三类是数据库管理员(DBA——Database Administrator)。

应用程序员是负责设计和编制应用程序的人。他们用某种高级程序设计语言和数据库管理系统所提供和支持的数据操作(纵)语言来编制应用程序。

终端用户是通过计算机终端，利用数据库去处理各种业务的人。比如图书管理员可以通过终端使用图书管理系统等去了解和管理图书，这时图书管理员便是一个终端用户。

数据库管理员(DBA——Database Administrator)负责全面的管理数据库系统的工作。主要包括：

1. 定义数据库 包括数据库的整体逻辑结构、局部逻辑结构和存储结构以及存取策略等。

2. 监督和控制数据库的运行和使用 DBA 定义合法权检验，当用户提出使用数据库时，负责安排，监督数据库的使用，协调用户等。

3. 数据库的维护和改进 DBA 定义后援和恢复策略, 遇到故障时负责数据库的恢复, 数据库运行后, 性能测试和性能分析, 根据应用环境的改变改进数据库设计, 数据库再组织, 通过改进和再组织提高时间效率和空间效率等.

1.5.2 数据库系统的体系结构

数据库系统中, 数据有三个不同的视图因而有三个不同的级别, 如图 1.3 所示. 这是某些数据库系统研究组织提出来的

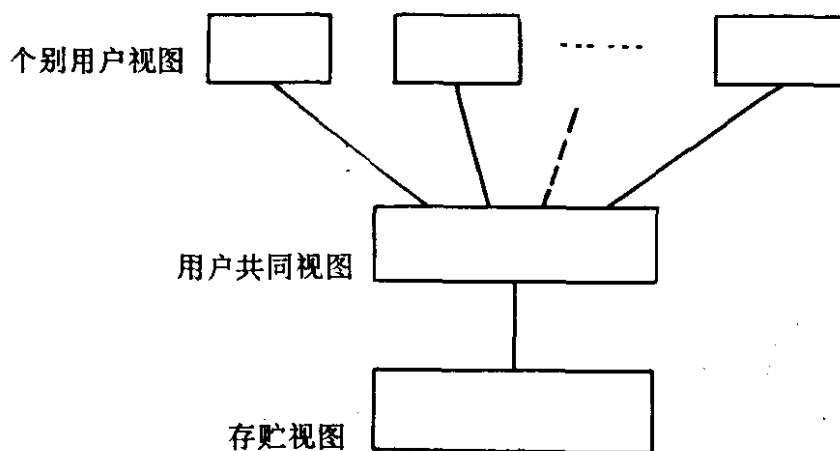


图 1.3

典型观点. ANSI/X3/SPARC (the Standards Planning and Requirements Committee of the American National Standards Committee on Computers and Information Processing) 数据库管理系统研究组提出的研究报告和 CODASYL (the Conference on Data Systems Languages) 提案都是这种观点.

在图 1.3 中可以看到, 这三种数据视图是: 个别用户视图、用户共同视图、存贮视图. 存贮视图是最接近实际物理存贮的一层, 即与实际存贮数据的方式有关的一层, 存贮视图只有一个; 个别用户视图是最接近用户的一层, 即各单个用户看待数据的方式有关的一层, 个别用户一般只对整个数据库的某一部分感兴趣, 从而一般只对某一部分有关, 因此存在着多个个别

用户视图；用户共同视图是介于存贮视图与个别用户视图之间的一层，它是与整个数据库数据有关的一层，用户共同视图只有一个。

在描述或定义数据库数据的时候，对应于上述三种视图都要给出相应的描述。在用户共同视图定义中包括数据库数据的整体逻辑结构的定义，在个别用户视图的定义中包括和个别用户有关的那一部分数据的局部逻辑结构的定义，在存贮视图定义中包括数据库数据的存贮结构的定义。关于各种视图的定义，将在下一节中给出更详细的叙述。

目前，在数据库领域里所使用的术语以及它们的含义的解释并不完全一致。在 ANSI/X3/SPARC 报告中数据库体系结构分三级(层)；外部级(External level, 有的译为外层)、概念级(Conceptual level, 有的译为概念层)、内部级(Internal level, 有的译为内层)，它们分别对应于外视图(External view)、概念视图(Conceptual view)、内视图(Internal view)，这些大致相当于个别用户视图、用户共同视图、存贮视图。

在当前运行的数据库系统中，并不都按上述方法来描述数据库的体系结构。这里给出的三级结构以及在图 1.4 中给出的数据库系统的体系结构是一般认为可用来叙述数据库系统的较典型的参考结构，它较好地体现了数据库系的体系结构的主要思想，它符合许多系统。

用户共同视图用模式(Schema)来定义。模式是数据库内容和结构方式的完整的表示，包括数据库数据的整体逻辑结构的描述，它包括各种记录型的定义、记录型之间的联系、数据项的定义，有的还可能包括完整性约束的定义、权限检查等其它一些内容。

个别用户视图用子模式(Subschema)来定义。当然对一个(或一类)个别用户要定义一个子模式。子模式是与一个个别用户视图有关的那部分数据的逻辑结构的描述，也就是局部逻辑

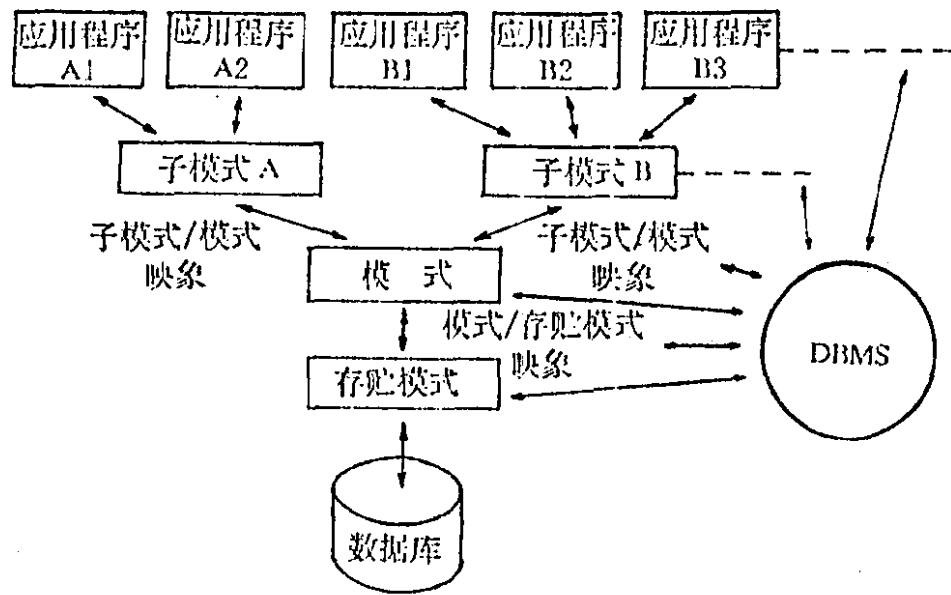


图 1.4

结构的定义，其中包括本视图中的记录型的定义、记录型之间的联系、数据项的定义。把一个子模式可以看做是一个模式的逻辑子集，但它不一定是一般意义上的子集。子模式与模式之间通过子模式/模式映象进行转换。

存储视图用存储模式(Storage schema)来定义。在存储模式中定义数据的存储结构即数据的物理结构，包括各种存储记录型的定义，还可能包括存储方法、存取策略的某些特征的定义。模式与存储模式之间通过模式/存储模式映象进行转换。

在图 1.4 中可以看到，按照这种结构，一个数据库系统中数据要按三种级别来定义(描述)，其中模式只有一个，存储模式也只有一个，而子模式可以有多个。最上边是应用程序，它