

数

据

库

系

统

概

论

马锦忠 陆正康
陈学德 李波 等编

南京大学出版社

7731113
M110

399253

数据库系统概论

马锦忠 陆正康 陈学德 李波 等编著

南京大学出版社

1995·南京

(苏)新登字 011 号

JS211/26
内 容 简 介

本教材以关系数据技术、理论的基本内容为主体，讲述了关系数据模型、关系数据库设计理论及当前著名的关系数据库系统 SYBASE 等内容，并介绍了分布式数据库系统和面向对象数据库系统的基本概念。可作为计算机专业本科生与研究生的数据库教材，也可作为计算机工作者的参考书。

数据库系统概论

马锦忠 陆正康 陈学德 李波 等编著

*

南京大学出版社出版

(南京大学校内 邮政编码:210093)

江苏省新华书店发行 句容市印刷厂印刷

*

开本 787×1092 毫米 1/16 印张 12 字数:300 千

1995 年 10 月第 1 版 1995 年 10 月第 1 次印刷

印数 1—5000 册

ISBN 7—305—02808—8/TP·128

定价: 16.00 元

(南大版图书若有印、装错误可向承印厂退换)

前 言

数据库技术从 60 年代开始,发展到今日,其内容已十分丰富,应用极为广泛。目前,对数据库技术与理论的学习,已不限于计算机专业的学生,而其它专业的学生和工程技术人员也广为需要。

本教材根据编者在教学实践中的体会,结合数据库技术的发展与应用,对现有教材内容作了一番精选:以关系数据库技术、理论的基本内容为主体,介绍了关系数据模型,关系数据库设计理论及当前最著名的关系数据库系统 SYBASE;对层次数据库与网状数据库技术的介绍作了精简;对当前数据库技术中具有影响的面向对象数据库的一些基本概念也作了介绍。以上这种选材处理 还是一种尝试,是否妥当,谨请专家与读者指教。

本书的编写主要参考了萨师煊、王珊所著的《数据库系统概论》和施伯乐等编著的《关系数据库的理论及应用》等教材。由于水平有限、时间匆忙,缺点与错误难免,恳请读者批评指正。

参加本书编写的有马锦忠(第一章,第二章,第十章),张艳红(第三章,第四章),李波(第五章),曹向阳(第六章),陆正康(第七章,第八章),陈学德(第九章),并由冯蓓雷副教授审校。在编写过程中,曾得到杜戎权等同志的很多协助与支持,在此一并表示感谢!

编 者

1995. 3.

目 录

第一章 概述	1
1.1 引言	1
1.2 什么是数据库	2
1.3 数据库结构	3
1.4 数据库管理系统	6
1.5 数据描述语言和数据操纵语言	7
1.5.1 数据描述语言	7
1.5.2 数据操纵语言	8
1.6 数据库系统	8
练习题	11
第二章 数据模型	12
2.1 什么是数据模型.....	12
2.2 E-R 数据模型	13
2.2.1 E-R 数据模型的主要概念	14
2.2.2 E-R 图	15
2.3 层次数据模型简介.....	19
2.3.1 层次模型的数据结构与约束	19
2.3.2 层次模型的数据操作	21
2.4 网状数据模型简介.....	23
2.4.1 网状模型的数据结构与约束	24
2.4.2 网状模型的数据操作	27
2.5 关系模型.....	28
2.6 小结.....	30
练习题	32
第三章 关系数据库	33
3.1 关系模型的基本概念.....	33
3.1.1 关系的数学定义	33
3.1.2 关系模式	35
3.1.3 关系数据库	36
3.1.4 关系模型的三要素	36
3.2 关系代数.....	38
3.2.1 关系代数中的五种基本运算	38

3.2.2 关系代数其它四种运算	39
3.3 关系演算	41
3.3.1 元组关系演算	41
3.3.2 域关系演算	44
练习题	45
第四章 关系数据库标准语言——SQL	48
4.1 SQL 语言概述	48
4.2 SQL 数据定义功能	49
4.2.1 基本表的定义和修改	49
4.2.2 索引的建立和删除	50
4.3 SQL 数据操纵功能	51
4.3.1 SQL 查询语句	51
4.3.1.1 简单查询	52
4.3.1.2 库函数查询	53
4.3.1.3 使用谓词 BETWEEN、IN 和 LIKE 的查询	54
4.3.1.4 连接查询	54
4.3.1.5 嵌套查询	55
4.3.1.6 SQL 语言中的并、交、差操作	57
4.3.2 SQL 更新语句	58
4.4 视图	61
4.4.1 视图的定义	61
4.4.2 视图的查询	62
4.4.3 视图的更新	62
4.5 SQL 数据控制功能	64
4.5.1 使用权的授予	64
4.5.2 使用权的收回	65
练习题	66
第五章 关系数据库系统——SYBASE	67
5.1 SYBASE 系统概述	67
5.1.1 SYBASE 的主要特点	67
5.1.2 SYBASE 软件的组成	68
5.1.2.1 SYBASE SQL Server	69
5.1.2.2 SYBASE SQL Toolset	71
5.1.2.3 接口软件	73
5.2 SYBASE 系统安全性、完整性、并发性控制	76
5.2.1 SQL Server 的结构特征	76
5.2.2 SYBASE 数据库的基本对象	77
5.2.3 SQL Server 安全管理	78
5.2.4 SQL Server 完整性控制	79
5.2.5 SQL Server 并发控制	80
第六章 关系数据理论	83

6.1	问题的提出	83
6.2	规范化	84
6.2.1	函数依赖	84
6.2.2	关键字	85
6.2.3	范式	86
6.2.4	2NF	87
6.2.5	3NF	88
6.2.6	BCNF	89
6.2.7	多值依赖	90
6.2.8	4NF	93
6.2.9	规范化小结	94
6.3	函数依赖公理	95
6.4	关系模式的分解	98
6.4.1	关系模式分解的三个定义	99
6.4.2	分解的无损连接性和保持函数依赖性	100
6.4.3	关系模式分解的算法	102
6.4.4	小结	104
	练习题	106
第七章	DBMS 及查询优化	107
7.1	DBMS 的结构	107
7.2	语言处理	108
7.3	模式转换	109
7.4	数据存取	111
7.5	数据库物理组织	112
7.6	一些主要的关系数据库管理系统	113
7.7	查询优化概述	114
	练习题	117
第八章	数据库设计	118
8.1	引言	118
8.2	用户需求分析	120
8.3	概念结构设计	123
8.4	逻辑结构设计	125
8.5	物理结构设计	129
8.6	数据库设计专家系统	129
	练习题	132
第九章	分布式数据库	133
9.1	概述	133
9.1.1	什么是分布式数据库	133
9.1.2	分布式数据库的开发动力	135
9.1.3	分布式数据库的技术问题	136

9.1.4	分布式数据库管理系统	138
9.1.4.1	分布式数据库管理系统的基本组成	138
9.1.4.2	分布式数据库管理系统的同构性与异构性	139
9.2	分布式数据库原理	140
9.2.1	分布式数据库的模式结构与分布透明性	140
9.2.1.1	分布式数据库的模式结构	140
9.2.1.2	数据分片	141
9.2.1.3	分布透明性	144
9.2.2	分布式查询优化	145
9.2.2.1	分布式查询优化问题	146
9.2.2.2	分布式查询优化的目标	146
9.2.2.3	连接查询的优化	147
9.2.2.4	分布式查询优化的重要性	148
9.2.3	分布式事务管理	151
9.2.3.1	分布式事务管理概述	151
9.2.3.2	并发控制	153
9.2.3.3	恢复机制	159
9.2.4	分布式数据库管理	160
9.2.4.1	分布式数据库的目录管理	161
9.2.4.2	授权与保护	161
9.3	分布式数据库实验系统与产品	163
9.3.1	实验系统与产品概述	163
9.3.2	Sybase System 10 的分布式数据库管理功能	168
第十章	数据库技术发展概述	176
10.1	新一代数据库技术的需求	176
10.1.1	新的应用领域	176
10.1.2	新的应用领域对数据库技术的要求	176
10.2	面向对象数据库(OODB)	178
10.2.1	面向对象数据库的主要概念	178
10.2.2	对象类之间的联系	179
10.2.3	面向对象数据库系统	180
10.3	关于新一代数据库系统的争论及发展趋势	182

第一章 概述

1.1 引言

数据库技术自产生以来,发展到今日已形成了坚实的理论基础和独特的数据处理技术,并获得了广泛的应用。

数据管理技术在进入数据库阶段以前,经历了人工管理阶段和文件管理阶段。在 50 年代中期之前,为人工管理阶段。这时计算机主要用于科学计算。从硬件上看,外存只有磁带、卡片、纸带,没有磁盘等直接存取设备;从软件上看,没有操作系统,没有管理数据的软件,数据处理方式是批处理。

50 年代后期到 60 年代中期,为文件系统阶段。计算机不仅用于科学计算,还大量用于管理,外存有了磁盘设备;在软件方面,操作系统中有专门管理数据的软件,一般称为文件系统;从处理方式讲,不仅有批处理方式,而且能联机实时处理方式。文件系统出现以后,计算机对数据处理前进了一大步,计算机能对外存贮器中的数据进行处理,可以进行查询、修改、删除等操作;能进行文件结构与存贮结构之间的转换,使程序和数据有了一定的独立性;顺序文件、索引文件、链接文件等文件技术提高了数据处理效率。

文件系统虽然有如上所述的进步,但仍然存在着数据冗余大,数据和程序之间缺乏独立性等方面的缺点。

数据管理技术进入数据库阶段是在 60 年代后期,其主要标志是如下三件大事:

(1) 1969 年美国 IBM 公司的数据库管理系统 IMS(Information Management System)问世。

(2) 同年美国 CODASYL 委员会公布了它的研究成果 DBTG 报告。

(3) 从 1970 年起,IBM 公司高级研究员 E. F. Cood 连续发表了一系列论文,奠定了关系数据库的基础。

1978 年,美国 ANSI/X3/SPARC(美国计算机与信息处理国家标准化委员会规划与需求委员会)DBMS 研究组的研究报告“ANSI/X3/SPARC DBMS Framework of the Study Group on DBMS”(简称 SPARC 报告)发表,它标志着数据库技术进入了成熟阶段。SPARC 报告提出了一个标准化的数据库系统模型,对数据库系统的总体结构、特征、各个组成部分以及相应接口,都作了较明确的规定。迄今为止,它仍然是数据库技术影响最大的重要文件。SPARC 报告中的一个重要思想是采用内部级、概念级、外部级这三级结构,以最大程度的数据独立性,提供了脱离具体数据库管理系统的数据库模型的概念模式,使用数据字典/目录以支持系统内部的数

据管理,以及映像功能的一些明确规定等,都在 DBMS 软件开发中起到了关键性作用。

数据库学科的研究范围十分广泛,概括地讲可以分为三个主要领域。

(1) 数据库管理系统软件的研制

数据库管理系统(DBMS)是数据库系统的基础,研制 DBMS 的基本目标是扩大功能、提高性能、增强可用性。研制以 DBMS 为核心的一组相互联系的软件系统已成为当前数据库软件产品的方向,这些在 DBMS 基础上运行的软件系统有:数据通讯软件、表格软件、数据字典、报表书写、图形系统等。

由于数据库应用领域的不断扩大,数据库不仅广泛应用于管理,而且开始应用到工程设计、图形图像和声音多介质处理、自动控制和计算机辅助设计等新的领域,这些新应用领域中所处理的数据与管理领域中的数据有很大的区别,如声音、图像,我们称之为非格式化数据。处理要求也有不同,因而研究这些新的应用领域中的数据库技术是一个新课题。这不仅涉及应用系统的设计方法,而且涉及到数据库系统模型,以及实现技术等各种新问题。面向对象的数据系统,多媒体数据库等研究方向的兴起,就是基于这些需求而产生的。

(2) 数据库设计

在数据库管理系统的支持下,按照某企业某组织的需求,设计一个结构良好,使用方便,效率较高的数据库及其应用系统,是数据库设计的主要任务。在这个领域内,主要研究课题是数据库设计方法学;数据模型和数据建模方法的研究;数据库设计计算机辅助方法的研究;数据库设计规范和研究,以及数据库设计工具的开发等。

(3) 数据库理论

数据库理论研究主要集中于关系的规范化理论、关系数据库理论、数据库与人工智能结合、数据库与逻辑、逻辑演绎和知识推理等方面,此外,演绎数据库、面向对象数据库、知识库系统的研究等也都是研究方向。

本课程对上述三个方面的内容将作有选择有重点的介绍。

1.2 什么是数据库

数据库技术发展到今天,可说是较为成熟,但究竟什么是数据库还未有一个被普遍接受的定义。这说明数据库的内涵与外延还在不断地丰富与扩充,因而很难严格而简明地形式化定义它。这里,我们就人们从不同角度给出的定义或描述,列举几例:

- 数据库是团体用户的团体数据。
- 数据库是组织、存贮、管理数据的仓库。
- 数据库可以被定义为一个基于某种数据模型存贮起来的、为某个特定组织的多种应用服务、并具有尽可能小的冗余度的相互关联的数据集合。其数据结构独立于使用数据的程序,对数据的增添、修改和检索,由系统进行统一的控制,系统提供的数据库模型也有利于将来应用的扩展。

将最后一种数据库的定义概括起来,数据库具有以下一些特征。

(1) 数据共享性。数据共享的意义是指多种应用、多种语言互相覆盖地共享数据集合。在数据库里,数据不再分属于各个应用程序,而是集中存贮在一起,这样就大大提高了数据的使

用价值。

由于各种用户可以用不同的语言使用同一个数据库,因此,这就方便了用户。

(2) 数据一体化和结构化。在数据库中,数据是按照某种数据模型组织起来的,不仅文件内部数据之间彼此是相关的,而且文件之间在结构上也有有机地联系在一起,整个数据库浑然一体,这一特征通常称为“数据一体化”。

由于数据库是为多个应用目的服务的,数据库的结构不再面向特定的应用,而是面向整个用户集团或面向数据本身,因而具有较大的弹性,易于维护与扩充,应用数据灵活方便。

(3) 较少的数据冗余度。由于数据一体化,数据的冗余度大大减少,除了必要的副本以外,存贮数据的冗余度保持在尽可能小的程度。

(4) 数据独立性。在数据库系统中,系统提供映射功能,确保应用程序对数据结构和存取方法有较高的独立性,当应用发生变化或为了改进数据库性能而改变存贮介质或数据结构时,只需要对映像部分作相应地改变,一般不必修改应用程序。

通常,数据库系统提供两个层次的数据独立性。

一个是物理独立性,即由系统提供存贮结构与逻辑结构之间的映射,这样当改动存贮结构时,逻辑结构可以不改变,从而避免使应用程序“感知”。

另一个是逻辑独立性。一个数据库有它的总体逻辑结构(一般称为模式),而部门或某应用的需要有各种局部的逻辑结构(称子模式或外模式),数据库系统同样提供映射,实现局部逻辑与总体逻辑之间的转换,允许局部逻辑有独立于总体逻辑结构的表示方法。这样,当改变数据项之间的联系,增加新的数据类型或新的应用时,将引起数据库的重新构造,物理结构和总体逻辑结构发生改变,但局部逻辑结构仍可以保持不变,或所受影响很小,从而根据局部逻辑编写的程序,就不必改变或改变很小。

(5) 对数据实行集中统一的控制。系统提供统一的数据定义、删除、检索以及更新手段,并统一控制数据的安全性、完整性、保密性和并发性,使得对数据的应用更加有效和可靠。

当然,数据库还有其它一些特征,如可调整性等,但主要是以上五个方面的特征。

1.3 数据库结构

数据独立性是数据库技术最重要的特征之一,而数据独立性是通过数据库结构的层次化来实现的。

数据库结构分为三级,由外模式、模式和内模式组成,如图 1-1 所示。

- 外模式,亦称子模式或用户模式,是局部数据逻辑结构的描述,是数据库用户所见的数据视图。

- 模式,亦称逻辑模式,是数据库中全体数据的逻辑结构和特征的描述,是所有用户的公共数据视图。

- 内模式,亦称存贮模式,是数据的物理结构和存贮方式的描述,是数据在数据库系统中的内部表示。

数据库结构的三级模式是对数据的三个抽象级别,它把数据的具体组织留给了数据库管理系统(DBMS)去管理,使用户能逻辑地抽象地处理数据,而不必关心数据在计算机中具体的

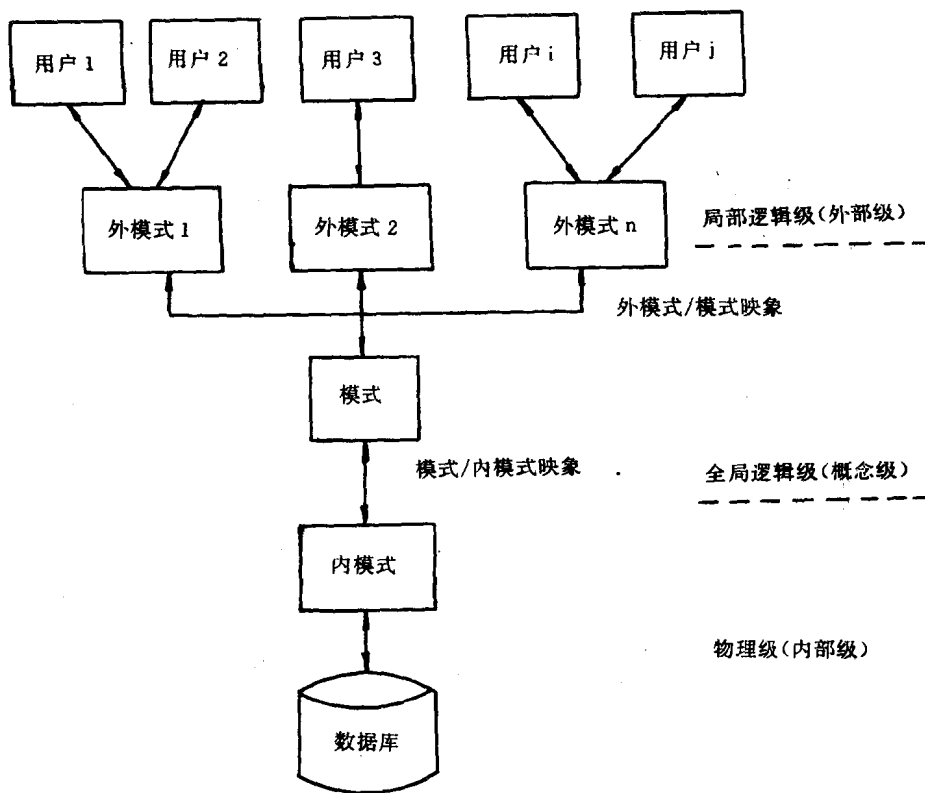


图 1-1 数据库的结构体系

表示和存贮。为实现这三个抽象层次的联系和转换,数据库系统在三级模式中提供两种映像:

- 外模式/模式映像
- 模式/内模式映像

下面对这些模式和映像作进一步的说明。

(1) 模式

模式是数据库中全部数据的逻辑表示或描述。所谓“逻辑”表示是指独立于存贮的关于数据类型以及它们之间联系的形式表示或描述。因此,模式既不同于内模式,比内模式抽象,不涉及数据的物理存贮细节和硬件环境;也不同于外模式,不涉及具体的应用程序和程序语言等方面。

模式是数据库数据在逻辑级上的视图,它通常以某种数据模型为工具来进行描述。模式不仅仅是数据的逻辑结构的定义(如一般高级程序设计语言中定义记录的名字、数据项的类型、长度、名字等),还定义与数据有关的安全性、完整性等要求,既要定义数据记录内部的结构,又要定义数据项之间的联系,以及记录之间联系等。

(2) 外模式

外模式是用户的数据视图,是与某一应用有关的数据的逻辑表示。不同用户的外模式可以相互覆盖。

外模式通常是模式的子集。这是各个用户的数据视图,不同的用户需求不同,看待数据的方式也可以不同,对数据保密的要求可能不同,使用的程序语言也可以不同。因此,不同用户的

外模式描述是不同的,即外模式中的结构、类型、长度、保密级别等都可以不同。

(3) 内模式

内模式是全体数据库数据的内部表示或者低层描述,用来定义数据的存贮方式和物理结构。例如记录是顺序存贮还是按照 B 树结构存贮,或者是按 hash 方法存贮,索引的组织方式是什么,数据是否压缩存贮,是否加密,数据记录结构规定等。

(4) 映象

外模式/模式的映象定义某一个外模式和模式之间的对应关系。这些定义通常包含在各自的外模式中。当模式改变时,外模式/模式的映象要作相应的改变(由 DBA 完成)以保证外模式不变。

模式/内模式的映象定义数据逻辑结构和存贮结构之间的对应关系。例如,说明逻辑记录和字段在内部如何表示的。当数据库的存贮结构改变时,模式/内模式的映象也必须作相应的修改(由 DBA 完成),使得模式保持不变。

正是由于具有上述两种映象功能,使得数据库系统具有较高的逻辑独立性和物理独立性。

(5) 概念模式

SPARC 报告也提出了数据库结构应具有三级模式结构。这三级结构是概念模式、外模式和内模式,其中外模式和内模式与上面讲的大致相当,而概念模式与上面所讲的模式是有差别的。

SPARC 报告中的概念模式是指一个部门或企业所对应的现实世界的真实模型。概念模式仅描述实体和它们的性质,不涉及机器世界的概念。概念模式通常用概念数据模型(或语义数据模型)为工具来描述。

本书中将概念模式和模式加以区分。

- 概念模式是信息世界范畴内的信息结构,而模式是机器世界范畴内的逻辑表示。

- 概念模式独立于具体的计算机系统,是一个部门或企业“长期无限制”的模型。常用概念数据模型(如 E-R 数据模型)为工具来描述。而模式依赖于具体数据库管理系统。常用基本数据模型(如层次、网状、关系等数据模型)来描述。因此,概念模式将依据具体的基本数据模型进行转换,以形成数据库系统模式。

- 概念模式和模式都是描述信息或数据的整体结构的,它们是现实世界在不同层次上的抽象,概念模式离机器更远。从机器观点上看,概念模式比模式更抽象。

总之,数据库结构分层描述后,既提高数据独立性,又增强了数据库系统的灵活性。

至此,我们介绍了数据库结构的三个层次的各种模式描述,它们之间的逻辑关系如图 1-1 所示。需要强调的是,数据库模式只是数据库的一种结构描述,还不是数据库本身,它是一个可以按照给定的某种约束装入具体数据的框架,但它本身并不涉及任何具体数据值。我们常常提到“定义了一个数据库”,实际上只是指定义了一个数据库模式。只有按照该模式装入数据以后,这个数据库才算真正地建立起来。在许多书籍里,常常把一个数据库称为它的模式的一个实例(instance)。而把模式称做该数据库的“信息格式”(information format)。一般说来,数据库的内容是随时变化的,但模式则是相对稳定的。

最后需要说明一点,上述的数据库结构层次化是对于完善的数据库管理系统而言的,目前现有的数据库管理系统商品软件中,三级数据结构划分常常不清晰,有的系统仅支持一级数据结构。关系数据库系统在支持独立性方面一般要优于其它类型的数据库系统。

1.4 数据库管理系统

数据库管理系统(Data Base Management System, 简记为 DBMS)是指数据库系统中对数据进行管理的软件系统,它是数据库系统的核心组成部分,数据库系统的一切操作,包括查询、更新以及各种控制都是通过 DBMS 进行的。

DBMS 是基于某种数据模型的。因此,可以把它看成是某种数据模型在计算机系统上的具体实现。根据所采用的数据模型的不同,DBMS 可以分为层次型、网状型、关系型等若干类型,但在不同的计算机系统上,由于缺乏统一的标准,即使是同种类型的 DBMS,它们在用户接口、系统功能等方面也常常是不同的。

二十几年来,围绕数据库技术标准化的问题进行了大量的讨论,这些讨论大大促进了人们对 DBMS 的深入理解(在这一领域的一些最重要的结果包括美国 CODASYL 委员会 DBTG 任务组的 1971 报告及其后续机构的系列改进版本,ANSI/SPARC 任务组的 1978 年报告,E. F. Cood 的 1981 图灵奖演说及其在 1985 年前后发表的一系列论文等)。

这些讨论的结果是确认:一个 DBMS 的主要目标是使数据作为一种可管理的资源来处理,DBMS 应该使数据易于为各种不同的用户所共享,应该增进数据的安全性、完整性和可用性,并提供高度的数据独立性。具体来说,一个比较完善的 DBMS 至少应该具有以下功能:

(1) 数据库定义功能:定义数据库的结构。包括模式、存贮模式和子模式,以及每一个子模式与模式,模式与存贮模式之间的映像;定义数据的完整性约束,保密限制约束条件。这些定义通常由数据库管理员(DBA)或数据所有者按系统提供的数据库定义语言的源形式给出,由 DBMS 自动将其转换成内部目标形式存入数据词典,供以后进行数据操作或数据控制时查阅使用,某些定义也允许用户查阅。

(2) 数据库操纵功能:包括数据的初始装入、对数据的存取和更新操作、数据库结构的维护与重新组织、数据的转贮等等,系统提供统一的数据操纵语言,允许用户根据需要在授权的范围内自由地进行上述操作。

(3) 数据库控制功能:DBMS 对数据库的控制主要包括三个方面:数据安全性控制、数据完整性控制以及在多用户多任务环境下的并发控制。

数据安全性控制是对数据库的一种保护,它的作用是防止数据库中的数据被未经授权的人访问、并防止他们有意或无意中数据库造成的破坏性改变。

数据完整性控制是 DBMS 对数据库提供保护的另一个重要方面。我们知道,数据的价值在于它的正确性,在于它正确地表达了现实世界中客体的信息。而这些客体的各种信息常常是具有某些固有特性和联系的。这些特性和联系体现为数据本身的语义内涵和各种关联关系,与语义内涵相矛盾的数据显然是无意义的,因而是必须避免的。完整性控制的目的是保持进入数据库中的存贮数据的语义的正确性和有效性,防止任何操作对数据造成违反其语义的改变。

并发控制的策略包括封锁单位大小的确定,死锁的预防、检测和解除等。

DBMS 还具有系统缓冲区的管理以及数据存贮的某些自适应调节机制等其他控制功能。

(4) 数据库通信功能:提供与操作系统的联机处理,与分时系统以及远程作业输入的相应

接口, DBMS 与网络软件的通讯功能等。

1.5 数据描述语言和数据操纵语言

数据库的模式是用数据模型为工具来设计的,设计好的模式提交给系统(机器),是通过数据描述语言来实现的。对系统中数据库进行的访问,更新等操作,则是通过数据操纵语言来实现的。下面对数据描述语言和操纵语言的概念作一个介绍。

1.5.1 数据描述语言

数据描述语言(Data Description Language, 简记作 DDL)用以定义数据库的各级数据结构及它们之间的映像,定义各种完整性约束和保密限制条件。它可以进一步区分为模式描述语言(Schema DDL)、子模式描述语言(Subschema DDL)和数据存贮描述语言(DSDL-Data Storage Description language),分别用于描述数据库的模式、子模式和存贮模式。

模式 DDL 和子模式 DDL 是描述数据库逻辑结构的工具,它们的基本功能如下:

(1) 标识数据库的各种结构,对每一个子结构说明其数据结构类型,并给出唯一命名。数据子结构类型的划分是因系统而异的。在网状数据库系统中,数据子结构的类型包括数据项类型、组项类型、记录类型以及系型等,而在关系型数据库系统中,数据子结构则仅仅限于关系和属性两种数据类型。

(2) 描述各种数据子结构之间的结构和语义联系,它包括:说明一个较大的数据子结构(如记录型)由哪些较小的数据子结构(如组项和数据项),以什么样的方式(如顺序、可重复度等)组成,以及定义一个记录型的主关键字等。

(3) 描述每一个基本数据项的基本特性,如类型特性(字符型、整数型、浮点数值型或日期型等等)、字符长度、值域、是否允许空值等。

(4) 描述安全控制方式和完整性约束条件。

(5) 由于子模式是在模式的基础之上定义的,因此在子模式 DDL 中,还必须提供定义子模式与模式之间的映像的能力。

另一方面,作为描述数据库存贮结构的手段,CODASYL 的数据库管理工作组(DBAWG)引入了 DSDL 的概念,这是一种将数据库模式映象到物理存贮模式的新的语言,提出这种语言的基本思想是提供丰富的物理存贮方式,供系统管理人员根据实际应用的需要加以选择,以便提高时空效率,但同时又保持数据独立性,避免由于存贮结构的改变而导致程序维护任务的增加。不过,这一设计思想在商业上可得到的 DBMS 产品中实现的程度还很有限,大多数 DBMS 还只限于提供索引选择和其他有限的存贮选择方式(如 INGRES 提供堆方式、有序堆方式、索引顺序组织方式和 HASH 方式及每种方式的压缩存贮形式共八种选择),尚未形成一种完善的存贮模式定义语言。

目前,在不同的 DBMS 中,模式 DDL 和子模式 DDL 实现的方式差异很大。在 DBTG 系统中,模式 DDL 被设计为一个独立的描述性语言,用它写出的数据模式很像一个英文说明书,极易阅读,子模式 DDL 则规定嵌入某种普通程序设计语言(通常称为主语言)中使用,设计成具

有与主语言相兼容的文法形式,因而不同的主语言有不同的子模式 DDL。在关系数据库 DBMS 中,则通常将各种 DDL 和 DML 组合成一种统一风格的数据子语言,称为关系数据子语言。关系数据子语言通常同时支持两种使用方式,既可嵌入主语言中使用(称作嵌入方式),又可以独立地以交互方式使用(称作自含方式),这是关系数据模型较之其它方式的灵活之处。SQL 语言是关系数据库标准定义语言。有关 SQL 语言中的数据定义功能等,将在 SQL 语言这章详细介绍。

1.5.2 数据操纵语言

数据操纵语言(Data Manipulation Language,简记作 DML)提供用户或应用程序访问数据库系统的接口。普通程序设计语言中通常也都含有对数据文件操作的语句,但其共同特点是每次只能处理一个记录。DML 语句则一般是集合型的操作,因此 DML 是一种更高级的数据处理语言。

DML 是因数据库系统而异的。一种典型的关系数据子语言是美国 IBM 公司开发的 SQL,该语言中的 DML 语句主要有:

查询语句: SELECT [UNIQUE] <选择表达式>

[INTO <目标表>]

FROM <关系名表>

[WHERE <条件表达式>]

[GROUP BY <字段名表> [HAVING <条件表达式>]];

插入语句: INSERT INTO <表名>:<元组或查询表达式>

删除语句: DELETE <表名> [WHERE <条件表达式>]

修改语句: UPDATE <表名> SET <字段名>=<表达式> {,<字段名>=<表达式> }

应用这些语句,可以对数据库中的数据记录进行各种插入、删除、修改和查询处理,我们将在 SQL 语言这一章中详细介绍这些语句的应用。

1.6 数据库系统

前面我们对数据库技术的一些重要特征:数据库的结构化、一体化、数据独立性等作了介绍,数据库技术的这些优点,使它成为数据管理技术的主流。在各种应用中,都是以数据库系统作为支撑基础的。

什么是数据库系统呢?

所谓数据库系统,它的含义已经不仅仅是一组对数据进行管理的软件(通常称为数据库管理系统),也不仅仅是一个数据库。一个数据库系统是一个实际可运行的、按照数据库方式存贮、维护和向应用系统提供数据或信息支持的系统。它是存贮介质、处理对象和管理系统的集合体。

按照这一观点,一个数据库系统至少由以下五个部分组成:

(1) 计算机基本系统：包括中央处理机、主存贮器、外部存贮设备、数据通道等各种存贮、处理和传输数据的硬件设备。操作系统、语言软件和各种实用程序等必要的软件。它们构成数据库系统必不可少的软硬环境。计算机系统一般是从市场上选购的，但对于数据库系统来说，要特别关注主、外存存贮容量，I/O 存取速度，可支持终端数和性能稳定性等指标，在许多应用中还要考虑系统支持联网的能力和配备必要的后备存贮设备等。

(2) 数据库管理系统(DBMS)：基于某种数据模型对数据进行管理的软件系统，我们在后文中将进行更深入的讨论。

(3) 数据库：与一个特定组织的各种应用相关的全部数据的汇集。通常由两大部分组成：一部分是有关应用所需要的工作数据的集合，称作物理数据库，它是数据库的主体；另一部分是关于各级数据结构的描述数据，称作描述数据库，通常由一个数据词典系统管理。

(4) 数据库管理员(Data Base Administrator, 简记为 DBA)：一组熟悉计算机数据处理业务、负责设计和维护数据库的技术人员。由于在数据库系统运行的整个生命期中，数据库管理员的干预自始至终都是必不可少的。因此，通常把 DBA 也视为数据库系统的一个必要组成部分。

(5) 系统运行所必需的各类文档和说明书。

图 1-2 给出了一个数据库系统主要组成部分及其相互关系的示意图。