

新世纪高等学校计算机系列教材

数据库原理与技术

刘方鑫	主 编
罗昌隆 刘同明	副主编
孙志挥	主 审

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书是一本高等院校计算机专业或相关专业的数据库课程教学用书,它以关系数据库系统为核心系统,完整地论述了数据库系统的基本概念、基本原理和应用技术,力图使读者对数据库系统有一个全面、深入、系统的了解,为进一步从事数据库系统的研究、开发和应用奠定坚实的基础。

本书主要内容包括:数据库系统的产生与发展、数据库系统特点、数据库系统的数据模型、数据库系统结构、关系数据库方法、关系数据库标准语言 SQL、数据依赖和关系规范化理论、数据库保护、数据库设计、分布式数据库系统和数据库技术的新发展等。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

数据库原理与技术/刘方鑫主编.—北京:电子工业出版社,2002.2
(新世纪高等学校计算机系列教材)

ISBN 7-5053-7163-0

I. 数... II. 刘... III. 关系数据库—数据库系统—高等学校—教材— IV. TP311.13

中国版本图书馆 CIP 数据核字(2001)第 090182 号

责任编辑:李 影

印 刷:

出版发行:电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:20 字数:512 千字

版 次:2002 年 2 月第 1 版 2002 年 2 月第 1 次印刷

印 数:6 000 册 定价:25.00 元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。
联系电话:(010)68279077

前 言

数据库技术是 20 世纪 60 年代后期产生和发展起来的一项计算机数据管理技术,它的出现和发展使计算机应用渗透到人类社会的广阔领域。目前数据库的建设规模和性能、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

从 20 世纪 70 年代后期开始,国外许多大学把数据库原理与技术列为计算机专业的学习内容,我国高等院校从 80 年代开始也把数据库原理与技术作为计算机专业的主要课程之一。目前数据库技术已成为信息科学技术领域的重要基础。

本书是一本高等院校计算机专业或相关专业的数据库课程教学用书,它以关系数据库系统为核心系统,完整地论述了数据库系统的基本概念、基本原理和应用技术,力图使读者对数据库系统有一个全面、深入、系统的了解,为进一步从事数据库系统的研究、开发和应用奠定坚实的基础。

本书主要包括:数据库系统的产生与发展、数据库系统特点、数据库系统的数据模型、数据库系统结构、关系数据库方法、关系数据库标准语言 SQL、数据依赖和关系规范化理论、数据库保护、数据库设计、分布式数据库系统和数据库技术的新发展等。

本书主要特点:

(1)以关系数据库系统为核心。在系统论述数据库基本知识的基础上,着重讨论了关系数据库的原理与实现,其中对关系数据模型、关系数据库体系结构、关系规范化理论、数据库设计方法、数据库并发控制、查询优化等都有较详细、系统的说明,对层次数据库和网状数据库等传统数据库的内容进行了精简,仅对其数据模型作了简要介绍。

(2)力求反映当前数据库领域的新水平。根据数据库领域的新发展,加强了数据模型和数据表示、分布式数据库系统、面向对象数据库、主动数据库、并行数据库、工程数据库、数据仓库以及基于 Web 的数据库系统等内容,力求反映当前数据库发展的新技术、新水平和新趋势。

(3)注重理论联系实际,加强数据库应用技术。为适应信息技术发展的需要,对数据库设计方法、开发工具、数据库语言(SQL)等数据库应用技术进行了较全面的论述,为读者开发数据库应用系统或维护管理大中型数据库系统打下基础。

(4)在内容选取、章节安排、难易程度等方面充分考虑教学的需要,力求使教材概念准确、清晰、重点明确,内容广泛,便于取舍,每章均配有习题便于教学。

本书由刘方鑫、罗昌隆、刘同明、徐明、潘江波、李毅、吴士军编写。刘方鑫任主编,罗昌隆、刘同明任副主编。其中,刘方鑫编写第 1 章、第 2 章、第 3 章、第 10 章、第 12 章的 12.1~12.4 节;罗昌隆编写第 8 章、第 9 章、第 12 章的 12.5~12.10 节;刘同明编写第 4 章、第 5 章、第 6 章;徐明编写第 7 章;潘江波编写第 4、10 章的部分章节;李毅编写第 1 章、第 2 章、第 3 章的部分章节及负责全书的修改工作,最后由刘方鑫对全书进行了统稿。

孙志挥教授审阅了全书,并提出许多宝贵意见,编者在此表示诚挚的谢意。

由于编者水平有限,书中疏漏谬误之处在所难免,殷切希望广大读者批评指正。

编 者
2001 年 9 月

目 录

第 1 章 数据库系统引论	(1)
1.1 数据库系统的产生与发展	(1)
1.1.1 数据和数据管理技术	(1)
1.1.2 数据库系统发展的三个里程碑	(3)
1.1.3 数据库系统的三个发展阶段	(4)
1.2 数据库与数据库系统	(5)
1.2.1 数据库	(5)
1.2.2 数据库系统	(6)
1.3 数据库系统的特点	(6)
1.3.1 实现数据的集中化控制	(6)
1.3.2 数据的冗余度小	(7)
1.3.3 采用一定的数据模型实现数据结构化	(7)
1.3.4 避免了数据的不一致性	(7)
1.3.5 实现数据共享	(8)
1.3.6 提供数据库保护	(8)
1.3.7 数据独立性	(8)
1.3.8 数据由 DBMS 统一管理和控制	(8)
习题 1	(9)
第 2 章 数据库系统的数据模型	(10)
2.1 数据模型概述	(10)
2.1.1 数据模型的基本组成	(10)
2.1.2 数据模型的发展	(11)
2.2 概念模型	(11)
2.2.1 E-R 数据模型的基本概念	(12)
2.2.2 E-R 图(E-R Diagram)	(14)
2.2.3 扩充的 E-R 数据模型	(15)
2.3 层次数据模型	(17)
2.3.1 层次数据模型的数据结构	(18)
2.3.2 层次数据模型的数据操作	(18)
2.3.3 层次数据模型的数据约束	(20)
2.3.4 层次数据模型的优缺点	(21)
2.4 网状数据模型	(21)
2.4.1 网状数据模型的数据结构	(21)
2.4.2 网状数据模型的数据操作	(23)
2.4.3 网状数据模型的约束	(25)

2.4.4	网状数据模型的优缺点	(25)
2.5	关系数据模型	(26)
2.5.1	关系数据模型的数据结构	(26)
2.5.2	关系数据模型的数据操作	(27)
2.5.3	关系数据模型的数据约束	(28)
2.5.4	关系数据模型的优缺点	(29)
2.6	面向对象数据模型	(29)
2.6.1	对象和对象标识符	(30)
2.6.2	属性和方法	(30)
2.6.3	封装和消息传递	(31)
2.6.4	类和实例	(31)
2.6.5	类层次结构和继承	(32)
2.6.6	持久性和版本	(33)
2.6.7	多态、重载、重定义与动态联编	(33)
2.6.8	面向对象数据模型与关系数据模型的简单比较	(34)
	习题 2	(34)
第 3 章	数据库系统结构	(36)
3.1	数据库系统的分级结构	(36)
3.1.1	数据库系统的三级模式结构	(36)
3.1.2	数据库模式间的映像与数据独立性	(38)
3.1.3	数据库的抽象层次	(39)
3.1.4	数据库的数据模式与数据模型的关系	(40)
3.2	数据库系统的组成	(40)
3.2.1	硬件支撑环境	(40)
3.2.2	软件系统	(41)
3.2.3	数据库	(41)
3.2.4	数据库管理员、系统分析员、程序设计员及用户	(42)
3.3	数据库管理系统的功能	(43)
3.4	DBMS 的系统结构	(44)
3.4.1	DBMS 的进程结构和多线索结构	(45)
3.4.2	DBMS 的组成	(45)
3.4.3	DBMS 的层次结构	(47)
3.5	DBMS 的语言翻译处理层	(48)
3.5.1	DDL、DCL、DML	(48)
3.5.2	语言翻译处理层工作过程	(50)
3.6	DBMS 的数据存取层	(51)
3.6.1	数据存取层的主要任务	(51)
3.6.2	数据库存取层的功能子系统	(51)
3.7	Client/Server 结构的 DBMS	(52)
3.7.1	Client/Server 结构的 DBMS 的功能划分	(52)

3.7.2	在 Client/Server 结构中数据存取的工作过程	(53)
3.7.3	Client/Server 结构 DBMS 中数据库服务器的优点	(54)
3.7.4	三层 Client/Server 结构	(54)
3.8	基于 Web 的数据库系统	(55)
3.8.1	基于 Web 的数据库系统的体系结构	(55)
3.8.2	CGI 和 API 技术	(56)
	习题 3	(58)
第 4 章	关系数据库方法	(59)
4.1	关系数据库的基本概念	(59)
4.1.1	关系的形式化定义	(59)
4.1.2	关系模式	(60)
4.1.3	关系数据库模式	(60)
4.2	关系数据库的用户接口和关系数据语言	(61)
4.3	关系代数	(62)
4.3.1	传统的集合运算	(63)
4.3.2	专门的关系运算	(64)
4.3.3	关系代数应用举例	(68)
4.4	关系演算	(68)
4.4.1	元组关系演算	(69)
*4.4.2	域关系演算	(72)
4.5	关系系统	(75)
4.5.1	关系系统	(75)
*4.5.2	全关系系统的十二条基本准则	(76)
4.6	关系系统的查询优化	(78)
4.6.1	查询优化的一般策略	(79)
4.6.2	代数优化	(79)
4.6.3	基于存取路径的规则优化	(85)
4.7	关系数据操纵语言的比较	(92)
	习题 4	(93)
第 5 章	关系数据库的标准语言 SQL	(99)
5.1	SQL 语言概述	(99)
5.2	关系模式的定义	(100)
5.2.1	基本表的定义	(100)
5.2.2	基本表的修改和删除	(104)
5.2.3	关于默认值	(104)
5.2.4	域	(105)
5.2.5	索引的建立和删除	(105)
5.3	数据查询	(106)
5.3.1	简单查询	(106)
5.3.2	复杂查询	(108)

5.4	聚集函数(Aggregation)	(113)
5.4.1	聚集函数的运算符	(113)
5.4.2	元组的分组	(113)
5.4.3	HAVING 子句	(114)
5.5	数据库更新	(114)
5.5.1	元组插入	(114)
5.5.2	元组删除	(115)
5.5.3	元组修改	(115)
5.6	视图	(116)
5.6.1	视图定义	(116)
5.6.2	视图查询	(117)
5.6.3	视图更新	(118)
5.6.4	视图删除	(119)
5.6.5	视图的作用	(119)
5.7	嵌入式 SQL	(120)
5.7.1	程序设计环境下的 SQL	(120)
5.7.2	单行选择语句	(122)
5.7.3	游标	(123)
* 5.8	动态 SQL	(126)
5.8.1	非查询类动态 SQL	(126)
5.8.2	查询类动态 SQL	(128)
	习题 5	(129)
第 6 章	数据依赖和关系规范化理论	(133)
6.1	为什么要规范化	(133)
6.2	函数依赖	(135)
6.2.1	函数依赖的定义	(135)
6.2.2	关系的关键字和超关键字	(137)
6.3	函数依赖理论	(138)
6.3.1	函数依赖的逻辑蕴涵	(138)
6.3.2	Armstrong 公理系统	(139)
6.3.3	属性集闭包	(141)
* 6.3.4	Armstrong 公理的正确性和完备性	(142)
6.4	函数依赖集的等价和覆盖	(144)
6.5	关系模式的分解	(146)
6.5.1	什么叫模式分解	(146)
6.5.2	分解的连接不失真性	(148)
6.5.3	依赖保持性	(151)
6.5.4	模式分解的级别——范式	(152)
6.5.5	模式分解算法	(154)
6.6	多值函数依赖与第四范式	(157)

6.6.1	BCNF 关系模式存在的问题	(157)
6.6.2	多值函数依赖的概念	(158)
6.6.3	第四范式	(158)
* 6.6.4	多值依赖公理系统和多值依赖中的模式分解	(159)
习题 6		(163)
第 7 章 数据库设计		(169)
7.1	数据库设计概述	(169)
7.1.1	数据库设计目标和设计方法	(169)
7.1.2	数据库设计的基本步骤	(171)
7.2	需求分析	(172)
7.2.1	需求分析的任务	(172)
7.2.2	需求分析的实现方法	(173)
7.2.3	需求分析实例	(174)
7.3	数据库概念结构设计	(177)
7.3.1	概念结构设计的引出	(177)
7.3.2	概念结构设计的要求及方法	(177)
7.3.3	局部概念结构设计	(179)
7.3.4	全局概念结构设计	(181)
7.3.5	概念结构设计实例	(184)
7.4	数据库逻辑结构设计及优化	(186)
7.4.1	逻辑结构设计任务与设计准则	(186)
7.4.2	概念模型向关系模型的转换规则	(188)
7.4.3	用关系规范化理论对关系数据模型进行优化	(190)
7.5	数据库的物理设计	(192)
7.5.1	数据库设计人员需掌握的物理设计知识	(192)
7.5.2	数据库物理设计的主要内容	(192)
7.6	数据库设计评价	(195)
7.6.1	数据库评价	(195)
7.6.2	数据库逻辑设计评价和物理设计评价	(196)
7.7	数据库的运行和维护	(197)
7.7.1	数据的载入和应用程序的设计与调试	(197)
7.7.2	数据库的试运行	(198)
7.7.3	数据库的运行和维护	(199)
习题 7		(199)
第 8 章 数据库安全性和完整性		(201)
8.1	数据库安全性概念	(201)
8.2	数据库安全性控制	(202)
8.2.1	用户标识与鉴定	(202)
8.2.2	存取控制	(202)
8.2.3	视图机制	(205)

8.2.4	数据加密	(205)
8.2.5	审计	(205)
8.3	数据库安全性实例——Oracle 系统的安全性	(206)
8.4	数据库完整性	(208)
8.4.1	完整性约束条件的类型	(208)
8.4.2	完整性控制机制的功能	(209)
8.4.3	完整性约束的表达方式	(209)
8.4.4	Oracle 系统的完整性约束表示	(217)
	习题 8	(219)
第 9 章	数据库系统的恢复和并发控制技术	(221)
9.1	事务	(221)
9.2	数据库恢复	(222)
9.2.1	故障的类型	(222)
9.2.2	数据库恢复技术	(223)
9.2.3	数据库恢复策略	(225)
9.3	并发控制	(226)
9.3.1	并发操作引起的问题	(226)
9.3.2	调度的可串行性	(227)
9.4	基于封锁的并发控制技术	(229)
9.4.1	X 锁	(229)
9.4.2	(S,X) 锁	(230)
9.4.3	(S,U,X) 锁	(230)
9.4.4	两段锁协议	(231)
9.4.5	严格的 2PL 实施	(233)
9.5	死锁的检测、处理和防止	(233)
9.5.1	死锁预防	(233)
9.5.2	死锁的检测和处理	(234)
* 9.6	封锁的粒度和多粒度的封锁	(234)
9.6.1	封锁的粒度	(234)
9.6.2	多粒度封锁	(235)
* 9.7	基于时间标记的并发控制技术	(236)
	习题 9	(238)
第 10 章	分布式数据库系统	(239)
10.1	分布式数据库系统概述	(239)
10.1.1	什么是分布式数据库系统	(239)
10.1.2	分布式数据库系统的产生与发展	(240)
10.2	分布式数据库系统的特点	(243)
10.2.1	“全能”分布式数据库系统应符合的准则	(243)
10.2.2	分布式数据库系统的特点	(244)
10.3	分布式数据库管理系统(DDBMS)	(245)

10.3.1	DDBMS 的定义和功能	(245)
10.3.2	DDBMS 的组成	(246)
10.3.3	DDBMS 的分类	(247)
10.4	分布式数据库系统的体系结构	(249)
10.4.1	基于 ANSI/SPARC 数据模式的 DDBS 的体系结构	(249)
10.4.2	分布透明性	(252)
10.5	数据分段的规则和类型	(254)
10.5.1	数据分段的目的和规则	(254)
10.5.2	数据分段的类型	(254)
10.5.3	数据分布对数据管理的影响	(256)
* 10.6	分布式查询和优化	(257)
10.6.1	分布式查询处理	(257)
10.6.2	全局查询优化的目标	(258)
* 10.7	分布式事务管理	(258)
10.7.1	分布式事务的基本特征和管理目标	(259)
10.7.2	分布式事务的恢复技术	(259)
10.7.3	分布式数据库中的可串行性	(261)
* 10.8	分布式数据库系统中的并发控制	(262)
10.8.1	基于封锁的并发控制算法	(262)
10.8.2	基于时间戳的并发控制算法	(263)
10.9	异构数据库互连	(265)
10.9.1	异构型 DDBMS	(266)
10.9.2	SAG 规范和 IBM 的 DRDA	(266)
10.9.3	开放型数据库互连 ODBC 的结构	(267)
10.9.4	利用信关技术实现异构数据库互连	(269)
习题 10	(270)
第 11 章	数据库技术的新发展	(271)
11.1	传统数据库系统的局限性	(271)
11.2	新一代数据库系统	(272)
11.2.1	面向对象数据库系统宣言	(272)
11.2.2	第三代数据库系统宣言	(273)
11.2.3	数据库技术与相关技术结合形成新型的数据库系统	(274)
* 11.3	面向对象数据库系统	(275)
11.3.1	面向对象数据库系统概述	(275)
11.3.2	面向对象数据库管理系统 OODBMS	(275)
11.3.3	面向对象数据库系统的基本功能和实现途径	(276)
* 11.4	对象-关系数据库系统	(277)
11.4.1	对象-关系数据库系统的特点	(277)
11.4.2	对象-关系数据库系统实现的方法和途径	(278)
* 11.5	工程数据库系统	(278)

11.5.1	工程数据库系统的特点	(278)
11.5.2	工程数据库的数据模型	(279)
11.5.3	工程数据库的体系结构	(281)
* 11.6	并行数据库系统	(282)
11.6.1	并行数据库系统概述	(282)
11.6.2	并行数据库系统结构	(283)
11.6.3	数据划分	(285)
11.6.4	并行算法	(285)
11.6.5	查询优化方法	(287)
11.6.6	并行数据库系统与分布式数据库系统比较	(288)
* 11.7	主动数据库系统	(289)
11.7.1	主动数据库概述	(289)
11.7.2	主动数据库的模型	(289)
11.7.3	SQL3 中规则的表示和执行	(290)
11.7.4	主动数据库管理系统	(291)
* 11.8	数据仓库(Data Warehouse)	(293)
11.8.1	数据仓库的产生	(293)
11.8.2	数据仓库的概念和结构	(294)
11.8.3	联机分析处理(OLAP)	(295)
11.8.4	OLAP 实现技术	(298)
	习题 11	(301)
	参考文献	(302)

第 1 章 数据库系统引论

数据库是数据文件以及用于处理这些数据文件的程序集合,是数据管理的最新技术。数据库技术已成为现代信息技术的重要组成部分,是现代计算机信息系统和计算机应用系统的基础和核心。数据库技术是 20 世纪 60 年代后期产生和发展起来的一项计算机数据管理技术,它的出现使计算机应用渗透到人类社会的广阔领域。目前数据库的建设规模和性能、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志,数据库技术也成为计算机科学技术学科的一个重要分支。

本章主要讨论数据库系统的产生与发展,以及数据库系统的基本概念和特点。

1.1 数据库系统的产生与发展

1.1.1 数据和数据管理技术

1. 数据

数据是载荷信息的媒体,它包括数值型数据和非数值型数据。数值型数据是以数字表示信息,而非数值型数据是以符号及其组合来表示信息。例如字符、文字、图表、图形、图像、声音等均属于非数值型数据。

数据是人类社会发展的一种重要的信息资源,如何有效地保存和科学地管理这些数据是人们长期以来十分关注的课题,从而促进了数据管理技术的发展。

2. 数据管理技术

数据管理是指对数据的分类、组织、编码、存储、查询和维护等活动,是数据处理的中心环节。数据管理技术可根据提供的数据独立性、数据冗余度、数据共享性、数据间相互联系、数据安全性、数据完整性和数据存取方式等水平的高低划分为三个不同的发展阶段。

(1) 人工管理阶段

20 世纪 50 年代中期以前,计算机主要用于科学计算,数据管理处于人工管理阶段,其特点是:

- 数据管理缺乏软件系统支持,依靠应用程序管理数据。在应用程序中程序员不仅要规定数据的逻辑结构,而且还要设计数据的物理结构,包括存储结构、存取方法、输入方式等。
- 数据是面向应用的,一组数据只能对应一个应用程序,数据不能共享。
- 应用程序完全依赖于数据,不具有数据独立性,一旦数据的逻辑结构或物理结构发生变化,应用程序就要做相应的修改。
- 数据一般不需要长期保存。

(2) 文件系统阶段

20 世纪 50 年代后期至 60 年代中期,数据管理进入文件系统阶段。在文件系统中,把数据按其内容、结构和用途组织成若干个相互独立的文件。用户通过操作系统对文件进行打开、读写、关闭等操作。文件系统管理数据具有如下特点:

- 文件系统利用“按文件名访问,按记录进行存取”的管理技术,可以对文件进行修改、插入和删除操作。数据存取直接由操作系统(OS)提供支持,即采用“应用程序—OS—数据文件”的存取方式。

- 数据可以长期保存在存储设备上供用户使用。

尽管用文件系统管理数据比人工管理阶段有了长足的进步,但面对数据量大且结构复杂的数据管理任务,文件系统仍不能适应,主要表现在如下几点:

- 数据独立性差。文件系统中文件逻辑结构的改变必须修改应用程序;由于语言环境的变化要求修改应用程序时,也将引起文件数据结构的改变,因此数据与程序间仍缺乏数据独立性。

- 数据冗余度大且易产生数据不一致性。在文件系统中,文件一般为某一用户或用户组所有,文件仍是面向应用的,因此数据共享性差,冗余度大。同时由于相同数据重复存储、各自管理,易产生数据的不一致性。

- 在文件系统中,尽管其记录内部有结构,但记录之间没有联系,故其整体是无结构的。

- 数据无集中管理,其安全性、完整性得不到可靠保证,并且在数据的结构、编码、输出格式等方面难以做到规范化和标准化。

- 文件系统一般不支持多个应用程序对同一文件的并发访问,故数据处理的效率较低。

- 使用方式不够灵活。每个已经建立的数据文件只限于一定的应用,且难于对它进行修改和扩充。

这些缺点在规模大、结构复杂的数据管理系统中尤为突出。例如,美国于 20 世纪 60 年代初制定了阿波罗登月计划。阿波罗飞船由二百万个部件组成,且这些部件是由分散在世界各地的若干厂家生产的。为了掌握工程进度和协调工程进展,阿波罗计划的主要合作者 Rockwell 公司,曾研制了一个基于磁带的零部件生产计算机文件管理系统。该文件管理系统的数据库冗余高达 60% 以上,且只能以批处理方式工作,系统维护也困难。这些问题一度成为实现阿波罗登月计划的障碍。

上述问题是文件系统本身难以解决的,造成数据处理效率低、成本高,并有许多潜在问题需要解决。这正是数据库系统产生的背景。

(3) 数据库系统阶段

20 世纪 60 年代后期以来,为了克服文件系统的弊病,适应日益迅速增长的数据处理的需求,人们开始探索新的数据管理方法与工具。这一时期,磁盘存储技术取得重要进展,大容量和快速存取的磁盘相继投入市场,给新型数据管理技术的研制提供了良好的物质基础。为了解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,数据库技术由此应运而生,出现了统一管理数据的专门软件系统——数据库管理系统(Data Base Management System—DBMS)。世界上最早推出的数据库系统当属美国通用电气公司(GE)C. W. Bachman 等人于 1963 年研制的 IDS(Integrated Data Store)系统。IDS 奠定了网状数据库的基础,它可使多个 COBOL 程序共享数据库中的数据。IDS 是数据库系统的先驱,为此,Bachman 于 1973 年获得了美国计算机协会(ACM)颁发的图灵(Turing)奖。

数据库(Data Base—DB)一词的提出可以追溯到 20 世纪 50 年代后期。据日本的铃木道夫说:“那时,美国军队为了作战和装备,想把各种数据集中在一个地方作为数据基地。于是就把各个部队分散的情报集中起来管理,根据统一取得的最新数据考虑战略部署。据说,这就是数据库术语的起源”。后来美国系统发展公司为海军基地研制数据库借用了 Base(基地)一

词,Data Base 形象地表示在信息的海洋中巡航所必需的基地。然而,具有实用价值的真正数据库系统是 20 世纪 60 年代后期出现的。这一时期数据管理技术有了突破性的进展,有三个重要事件标志着数据库管理技术进入数据库系统阶段,这三个事件被称为数据库系统发展的三个里程碑。

1.1.2 数据库系统发展的三个里程碑

1. IMS 系统

1969 年美国 IBM 公司和 Rockwell 公司合作,研制成世界上第一个实用的数据库系统 IMS(Information Management System),为阿波罗飞船于 1969 年顺利登月提供了重要保证。IMS 是一个 DB/DC(Data Base / Data Communication)系统。它采用了以层次数据结构为基础的数据模型,即数据组织在逻辑上成树型结构。在以后几年中,IBM 公司对 IMS 系统进行了改进,先后推出了 IMS-2、IMS/VS 版本。由于 IBM 公司具有极强的竞争力,从而使 IMS 系统获得广泛使用,并对数据库技术的发展产生了重要影响,成为层次模型数据库的典型代表。

2. DBTG 报告

1969 年 10 月美国数据系统语言委员会(CODASYL—Conference On Data System Language)下属的数据库任务组(DBTG—Data Base Task Group)提出了“数据库建议书”,建议用数据描述语言 DDL(Data Description Language)描述数据库,用数据操纵语言 DML(Data Manipulation Language)作为主语言的扩充对数据库的数据进行操作。该建议书经过广泛的讨论,提出了 180 多条修改意见,由 DBTG 于 1971 年 4 月进行修改并发表,通常称这个经过修改的数据库建议书为“1971 年 DBTG 报告”。DBTG 报告给出了网状数据库系统的方案,为建立网状数据库提供了完整的系统设计和语言规范。在数据库系统形成过程中 CODASYL 做出了重要贡献,它虽没有把当时世界上流行的事务处理语言完全统一,但 DBTG 方案确立了网状数据库的规范,使 20 世纪 70 年代中期的大部分计算机系统均采用 DBTG 方式的数据库系统。最具有 DBTG 方案典型风格的商用系统是 Cullinet 软件公司的 IDMS,它可以运行在 IBM 360/370,UNIVAC 7090,Siemens 4004,PDP-11/45 等机器上。

3. 关系数据库系统

关系数据库系统是以二维表的形式组织数据,以关系数学理论为基础的数据库系统。关系数据方法的雏形可见于 60 年代初,最早将关系方法用于数据管理的是 1962 年 CODASYL 发表的一篇论文“信息代数”(Information Algebra),之后 1968 年 D. L. Childs 关于集合论数据结构 STDS(Set-Theoretic Data Structure)上的 n 元关系的研究在 IBM7090 机上实现,这些研究工作为关系数据库的建立打下基础。然而系统完备的关系数据库理论的建立是从 70 年代开始的。1970 年 6 月,IBM 公司 San Jose 研究所的 E. F. Codd 在美国计算机协会会刊“Communication of the ACM”上发表了题为“大型共享数据库的数据关系模型”(A Relational Model of Data for Shared Data Banks)的著名论文。在论文中首次全面论述了关系数据库的概念,提出了关系模型,引进了关系代数,推导了关系演算,阐述了数据间存在的函数相关性,概括了关系规范,从而在计算机科学中开创了研究关系数据库理论与方法的新领域。以后 E. F. Codd 相继又发表了多篇关于关系模型的论文,定义了关系数据库的基本概念,引进了规范化理论,提出了数据子语言及其完备性问题,为关系数据库的全面开发奠定了坚实的理论基础。E. F. Codd 作为关系数据库的创始人和奠基人,1981 年 11 月在 ACM(美国计算机协会)洛杉矶年会上,由于他对数据库管理系统的理论与实践作出奠基性和持续性的贡献,荣获了计

计算机科学的最高荣誉——图灵奖。他在接受图灵奖时发表了题为“关系数据库：提高生产率的实践基础”的演讲，指出关系方法使用关系数据结构，提供了高度的数据独立性，采用非过程化的数据语言使程序员完全摆脱了数据管理，数据库系统自动完成诸如存储、查询和更新等操作的具体实现，从而使用户可以专心于他的应用问题，这就大大提高了生产率。

从数据库系统发展的三个里程碑可以看出：描述客观世界的实体及其相互联系的方法可采用不同的数据模型，即用树型结构描述的层次模型，用网状结构描述的网状模型，以及用表结构描述的关系模型，与这些数据模型相对应的数据库分别称为层次型数据库、网状型数据库和关系型数据库。

1.1.3 数据库系统的三个发展阶段

自 20 世纪 60 年代末一些数据库系统投入运行以来，随着计算机技术的飞速发展和应用需求的不断增加，数据库系统的发展也经历了不同的发展阶段。

1. 第一代数据库系统

20 世纪 70 年代，数据库系统以广为流行的网状型数据库和层次型数据库为代表。这一时期投入实际应用的数据库系统基本上是网状型和层次型数据库系统，它们已实现数据管理中的“集中控制与数据共享”这一基本目标。它们是 20 世纪 60 年代技术条件下的合理产物，为数据库技术的发展和應用奠定了基础，至今在一些大型系统中仍有一些网状型和层次型数据库在运行。

20 世纪 70 年代初关系数据模型的提出受到了人们的高度重视。然而，当时也有一些人认为关系模型仅仅是一种理想化的数据模型，用它来实现具有高查询效率的数据库管理系统是困难的。1974 年，数据库界开展了一场分别以 E. F. Codd 和 C. W. Bachman 为首的支持与反对关系数据库的大辩论。这场论战促进了关系数据库的发展，吸引更多的公司和研究机构对关系数据库原型进行研究，并不断推出一批研究成果。在这些关系数据库原型中，功能最强、技术上最有代表性的是 1976 年 IBM 公司宣布的 System R(1974~1979)和美国加州大学 Berkeley 分校的 Ingres 关系数据库系统。这两个数据库原型系统提供了比较成熟的关系数据库技术，为开发商品化的关系数据库软件创造了有利的条件。IBM 公司在 System R 的基础上先后推出了 SQL/DS(1982 年)和 DB2(1985 年)两个商品化关系数据库系统。商品化的 Ingres 关系数据库软件也于 1981 年由 INGRES 公司完成。与此同时，1979 年美国 ORACLE 公司推出了用于 VAX 小型机上的关系数据库软件 Oracle(v2.0)，这被认为是第一次实现了使用 SQL 语言的商品化关系数据库软件。

上述事实说明：数据库理论和技术在 20 世纪 70 年代有了长足的进步。因此，计算机界一些专家称 20 世纪 70 年代是数据库的年代。

2. 第二代数据库系统

20 世纪 80 年代出现了以关系数据库为代表的第二代数据库系统。

进入 20 世纪 80 年代，计算机技术的快速发展为数据库技术的发展和广泛应用提供了有力的支持。关系数据库由于具有坚实的理论基础，结构简单，操作方便，且在提高查询效率和改善关系数据库性能方面取得突破性进展，因而在这一时期得到迅猛发展。关系数据库已逐渐成为数据库发展的主流，几乎所有新推出的数据库系统都是关系型数据库系统。20 世纪 80 年代是数据库技术逐渐走上成熟的时期，一批性能不断改善，版本不断更新的商品化关系数据库软件相继投入运行，如 Oracle, Sybase, Informix, Ingres 等关系数据库系统已广泛应用于大

型信息管理系统。这一时期数据库技术得到全面的发展,其中如下几点值得一提:

(1) 关系型数据库系统的发展促进了数据库系统的小型化。20 世纪 80 年代推出了一批适于在微机环境下运行的关系数据库系统,在迅速普及和推广使用数据库技术方面发挥了极其重要的作用。1981 年美国 Ashton-Tate 公司首先推出微机关系数据库软件 dBASE II, 1984 年版本升为 dBASE III,1986 年又推出 dBASE III Plus。1987 年美国 FOX SOFTWARE 公司推出与 dBASE III 兼容的性能更好的 FoxBASE+(V2.0),之后又推出 Foxpro 等软件。一些生产大型关系数据库管理系统的公司也不断推出微机版数据库软件,如 1984 年 ORACLE 公司研制成 Professional Oracle 版微机数据库软件。这些微机版的数据库系统的大量运行,极大地拓宽了数据库应用领域,加速了社会信息化的进程。

(2) 随着数据库技术的发展和计算机网络的广泛应用,分布式处理系统,特别是分布式数据库系统在 80 年代得到很大发展。目前,几乎所有分布式数据库系统都是关系型的,而且大多数有影响的关系数据库软件都扩充为具有一定分布式数据库特征的数据库管理系统。如 1986 年 ORACLE 公司和 INGRES 公司先后推出 SQL* Star 和 Ingres/Star 的开放型分布式关系数据库系统。

(3) 随着计算机的广泛应用,特别是一些新的应用领域不断提出新的应用要求,关系型数据库、层次型数据库、网状型数据库都表现出不同程度的局限性。因此,20 世纪 80 年代后期人们又提出研制新一代数据库的设想。

由于早期的数据库主要用于商业或行政等事务处理,通常把处理常规数据和事务数据为中心的常规数据库称为传统数据库(Traditional Database),因此,有些文献上常称第一代、第二代数据库为传统数据库。

3. 新一代数据库系统

自 20 世纪 80 年代末,90 年代初以来,开发新一代数据库技术成为数据库技术研究的热点课题。

进入 20 世纪 90 年代,人们对数据库系统的功能提出许多新的期望和需求,除对常规数据进行处理外,还要求对图形、图像、声音等多媒体数据、时态数据、空间数据、知识信息以及各种复杂对象等非常规数据提供有效的数据处理功能。为适应这些应用的需求,人们提出许多新概念、新思想和新方法,以及一些新的数据模型和新数据库管理系统的体系结构。

1.2 数据库与数据库系统

1.2.1 数据库

数据库技术是一门新兴学科,它的概念、原理和方法仍在不断发展变化中。目前,还没有一个统一的、公认的数据库定义。因此,在教科书和文献中可以看到有关数据库定义的不同论述,下面给出数据库的一种描述性定义。

数据库是长期存储在计算机系统内的一个通用化的、综合性的、有结构的、可共享的数据集合,具有较小的数据冗余度和较高的数据独立性、安全性和完整性。数据库的创建、运行和维护是在数据库管理系统控制下实现的,并可为各种用户共享。

现就数据库的定义作如下几点说明:

1. 数据库是一个通用化、集成化的相关数据的集合

数据库是一个单位或一个应用领域的通用数据处理系统。它存储的是属于企业和事业部门、团体和个人的有关数据的集合。数据库中的数据结构按一定的数据模型组织、描述和储存,其结构基于数据间的自然联系,从而可提供一切必要的存取路径,且数据不再针对某一应用,而是面向全组织,具有整体的结构化。这是数据库和传统数据文件系统的基本区别之一。这种结构化的数据集合有较小的数据冗余度、较高的数据独立性,它是数据库的基本要素和管理对象。

2. 数据库应满足各种用户的不同需要

数据库用户通常可分为两类:一类是批处理用户,也称为应用程序用户。这类用户使用程序设计语言(如 COBOL、PL/SQL、C/C++等)编写应用程序,对数据库进行检索、插入、删除和修改等操作,并产生数据输出;另一类是联机用户,或称为终端用户。终端用户可使用简单的终端命令或查询语言对数据库进行存取操作,通常以查询性应用为主。

由于各种用户可以同时使用一个数据库,所以数据库应具有数据共享特性,并提供数据完整性控制、安全性控制以及并发控制功能。

3. 负责数据管理和维护的软件系统—数据库管理系统(DBMS—Data Base Management System)

DBMS 是一个数据库管理软件,它是数据库系统的核心。DBMS 为用户提供方便的用户接口,帮助和控制每个用户对数据库的各种操作,它还提供数据库的定义功能和管理功能。整个数据库的创建、运行和维护是在 DBMS 控制下实现的。

1.2.2 数据库系统

数据库系统(DBS—Database System)是实现有组织地、动态地存储大量相关的结构化数据,方便各类用户使用数据库的计算机软件/硬件资源的集合。DBS 由数据库、数据库管理系统及其开发工具、数据库管理员(DBA)、计算机硬件和软件、应用系统、应用程序用户和终端用户等组成。图 1-1 给出数据库系统示意图,它是存储介质、处理对象、管理系统和用户的集合。

1.3 数据库系统的特点

数据库系统不仅克服了文件管理系统存在的主要问题,且提供了强有力的数据管理功能,其主要特点如下:

1.3.1 实现数据的集中化控制

正如前面指出的,在文件系统中文件关于用户而存在,使数据处于一种分散的状态;而数据库系统则能集中地控制和管理数据。

1. 数据库的数据是集成式的

通常将一个部门所涉及到的全部数据都组织在一个数据库中。例如一个学校进行人事档案的管理、学生学籍的管理、科技情报的管理、教学管理等各种数据处理。我们可以利用数据库系统,把各种应用相关的数据集中在一个数据库中统一进行维护和管理,各职能部门随时可以从数据库中提取所需的数据。因此,数据库中的数据不是把程序用到的数据进行简单堆积,