



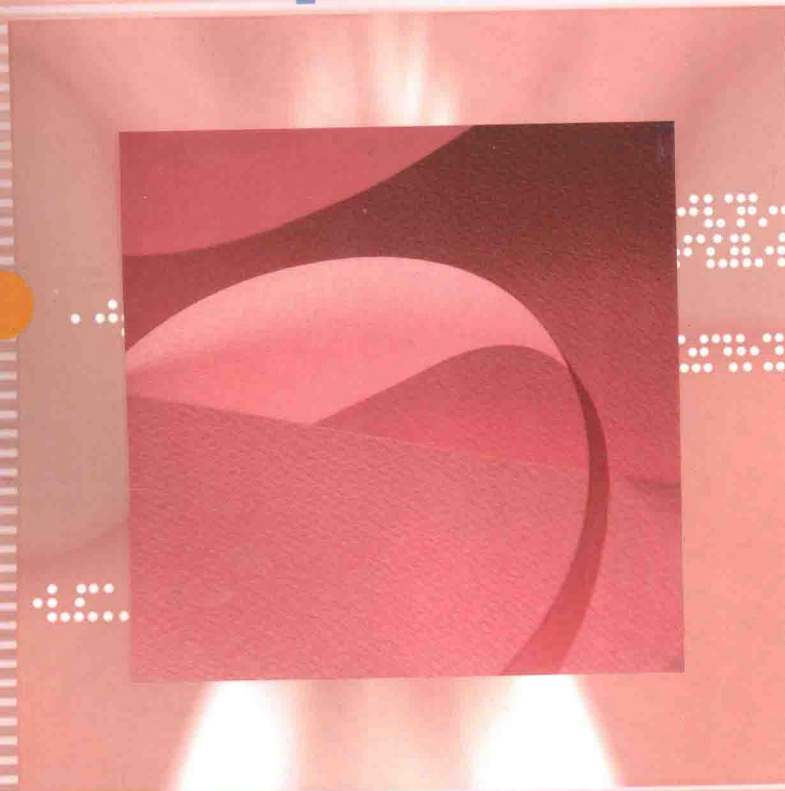
普通高等教育“十一五”国家级规划教材

高级数据库技术与应用

(第2版)

Advanced Database Technology and Application
(Second Edition)

汤庸 叶小平 陈洁敏 汤娜 编著



高等教育出版社

普通高等教育“十一五”国家级规划教材

高级数据库技术与应用

Gaoji Shujuku Jishu yu Yingyong

(第2版)

汤 庸 叶小平 陈洁敏 汤 娜 编著

高等教育出版社·北京

内容提要

本书是一部关于高级数据库技术与应用的基本教程和参考书。

全书共 6 章和 1 个附录。第 1 章绪论, 简述数据库技术发展和高级数据库技术的内涵与外延。第 2 章关系数据库基础, 简要回顾关系数据库经典内容。第 3 章高级数据模型, 主要讨论第三代数据库中的对象关系与面向对象数据模型, 以及经典模型在时间和空间两个基本维度上的扩展。第 4 章智能数据处理, 主要讨论知识库、主动数据库、决策支持系统、数据仓库与数据挖掘等智能数据分析与推理技术。第 5 章网络环境数据管理, 主要讨论分布式数据库、移动数据库、Web 数据库和 XML 数据库技术。第 6 章高级数据库应用实践, 主要包括时态数据库管理系统 TempDB、数据库技术综合应用实例 TempKB 和大数据技术应用实例——学者网。附录包括数据库学术资源、TempDB 2.0 中 ATSQL2 的 BNF 定义和“高级数据库技术”课程网站三个部分。

本书可作为高等学校计算机及相关专业高年级和研究生相关课程教材, 也可作为数据库学习、研究者的参考书。

图书在版编目(CIP)数据

高级数据库技术与应用 / 汤庸等编著. -- 2 版. --
北京: 高等教育出版社, 2015. 8
ISBN 978-7-04-043673-0

I. ①高… II. ①汤… III. ①数据库系统 - 高等学校 - 教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2015) 第 172091 号

策划编辑 倪文慧	责任编辑 韩 飞	封面设计 张 志	版式设计 杜微言
插图绘制 杜晓丹	责任校对 陈 杨	责任印制 田 甜	

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京嘉实印刷有限公司
开 本 787mm × 1092mm 1/16
印 张 19.75
字 数 420 千字
购书热线 010 - 58581118
咨询电话 400 - 810 - 0598

网 址	http://www.hep.edu.cn http://www.hep.com.cn
网上订购	http://www.landaco.com http://www.landaco.com.cn
版 次	2008 年 2 月第 1 版 2015 年 8 月第 2 版
印 次	2015 年 8 月第 1 次印刷
定 价	31.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 43673-00

前 言

数据库技术自 20 世纪 60 年代诞生以来,经历了多个重要发展阶段,产生了 C. W. Bachman、E. F. Codd 和 J. Gray 三位图灵奖获得者,形成了一个不断发展的庞大软件产业。进入 21 世纪后,在计算机网络技术推动之下,数据库技术更加呈现出蓬勃向前的发展局面,成为任何一种计算机信息系统都必须具备的基本技术支撑。在数据库教学和研究过程中,通常遇到两种基本情形,一是以关系数据库为代表的经典数据库仍旧占据整个数据库应用的主流地位,仍旧是所有数据库学习者和应用者的必备基础,再就是由于各种新型应用需求的驱动,各类新型数据管理技术不断涌现,成为人们在深入学习和研究数据库技术过程中必须面对的挑战。由关系数据库到各类新型数据库之间技术跨度较大,理论衔接并不明显,因此在经典技术与新近发展之间建立适当的连接桥梁,有效地实现两者之间的“平滑”过渡,就成为学习、研究者需要关注的一个基本课题,也是计算机专业本科高年级和研究生阶段开设和建设高级数据库技术与应用课程的初衷。

由汤庸、叶小平、汤娜、吴凌坤编著的本书第 1 版——普通高等教育“十一五”国家级规划教材《高级数据库技术与应用》得到不少兄弟院校师生的欢迎和使用,并提出了各类宝贵的意见和建议。另外,通过在中山大学、华南师范大学等学校的多次使用,也逐渐积累和形成了关于教学目标、课程选材以及内容组织等高级数据库课程建设基本问题的一些想法和体会,所有这些都会努力在本书第 2 版中予以体现。

本书不是在第 1 版基础上的简单修订,而是突破第 1 版的结构重新构思和写作的。全书共 6 章和 1 个附录。

第 1 章绪论,扼要叙述数据库技术的发展,结合作者学习和研究数据库的体会,尝试阐述高级数据库技术的基本内涵和外延,同时也讨论了数据库技术的发展趋势。

第 2 章关系数据库基础,简要回顾与叙述关系数据库的经典内容,试图通过对关系数据库原理与技术的分析梳理,为后续基本内容的展开进行必要的铺垫,提供技术基础。

第 3 章高级数据模型,主要讨论对象关系和面向对象数据模型,以及时态数据与空间数据模型。前两者是通常所讲的第三代数据库的基础,后两者是数据管理技术在两个最基

本维度——时间和空间上的扩展和深入。

第4章智能数据处理，主要包括知识库技术、主动数据库技术、数据仓库与数据挖掘技术等。数据的智能分析与知识推理是海量数据管理技术发展的必然，也是高级数据库技术中的基本内容。

第5章网络环境数据管理，主要包括分布式数据库、移动数据库、Web数据库和XML数据库管理。由于计算机网络技术的飞速发展，在网络环境下有效管理数据已经成为21世纪数据库技术的突出特征。

第6章高级数据库应用实践，主要从高级数据库管理系统、多种数据库技术综合应用、大数据时代信息系统应用三个角度介绍数据库实践，包括时态数据库管理系统TempDB的设计与开发、数据库技术综合应用实例——基于时态知识数据库的财政工资系统、大数据技术应用实例——学术社交网络“学者网”。通过这些数据库应用示例的介绍，希望有助于读者从原理探讨与技术开发结合的层面上进行相关内容的研究与学习。

最后，对第1版中的附录进行了修订，包括：数据库学术资源（列出2013年1月发布的《中国计算机学会推荐国际学术会议和期刊目录》（第三版）中“数据库、数据挖掘与内容检索”领域国际学术刊物和会议清单），TempDB 2.0中ATSQL2的BNF定义，“高级数据库技术”课程网站。

和第1版相比，本书结构更趋合理，内容更加成熟，素材更为新颖。本书从数据库新技术的整体出发进行构思，以“高级数据库技术内涵与外延—新型数据模型—智能数据处理与知识发现—网络环境数据管理—高级数据库技术应用”为主线，涵盖了高级数据库技术的主要方面。全书内容取自国内外数据库领域研究的新近重要成果，主要参考了SI-MOD、VLDB、ICDE等国际会议以及ACM TODS、VLDB Journal、IEEE TKDE等国内外权威期刊，基本上反映了高级数据库技术全貌。本书写作过程注重对相关技术的分析、梳理和总结，重点突出、思路明确，有助于读者理解高级数据库技术的来龙去脉和发展趋势，可读性较强。相对于第1版，本书大量加入近年来国内外新近研究成果，同时也参考我们自己的相关研究工作，力争在组织结构上更为合理，在课程内容上始终反映学科研究前沿和应用进展。

本书可用作高等学校计算机专业本科高年级学生选修课和研究生基础课的教材，也可作为从事数据库研究和应用人员学习与工作的参考书。本书第2版由汤庸统筹组织，全书主要内容由汤庸和叶小平负责编写，陈洁敏和汤娜参与了第6章的部分写作。华南师范大

学和中山大学数据库科研团队的教师以及“高级数据库技术”研讨班的博士生和其他学习者都提出了许多好的建议和见解，本书编写过程中参阅了国内外同行专家的相应著作和参考文献。在此，一并表示衷心感谢！由于水平所限，加上数据库新技术发展迅速，书中疏漏之处在所难免，诚请专家、老师和广大读者不吝指正。

本书得到广东省研究生示范课程（编号：2013SFKC08）的支持。

编者
2015年7月

第1版前言

数据库技术从20世纪60年代中期产生至今已有四十多年的历史，数据库技术一直是最活跃、发展速度最快的IT技术之一。一般来说，数据库发展可以分为3个阶段：第一代数据库以网状模型和层次模型为特征；第二代数据库以关系模型为基础，关系数据库在20世纪70年代末和80年代得到了广泛应用；20世纪90年代，随着网络和多媒体技术的迅猛发展，数据库的应用也得到进一步拓展，现代数据库技术融合多种技术，数据库技术的发展进入一个新的时期。

现代数据库应用范围非常广泛，在实际数据库应用开发中难以使用单一的数据库技术实现，往往需要融合应用许多方面的相关技术，网络化、智能化、多维化和协同化已成为现代数据库应用的主要特征。现代数据库技术所涵盖的范围很广，每种具体数据库技术都可以编写成独立教材，因此本书中不可能详细介绍每种数据库技术。本书根据网络化、智能化、多维化和协同化等特征，选取介绍其中一些基本的和有代表性的数据库技术，目的是使读者可以通过这门课程的学习对数据库新技术有一个整体的认识，为进一步从事相关研究和开发提供一些思路。

全书共分6章。第1章简要回顾数据库发展，讨论新一代数据库技术的范畴和实现途径，给出现代数据库技术及应用展望。

第2章简要介绍关系数据库理论，主要包括关系数据模型、关系模式设计和事务处理等关系理论基础，然后介绍面向对象数据库、对象关系数据库的基本概念和基础知识等。本章知识点是高级数据库技术的基础。

第3章介绍基于时间的数据库技术。时间是自然界重要属性，也是新一代数据库的重要特征。本章主要介绍多维数据库技术，包括时间数据演算、时态数据库技术、实时数据库技术、时空数据库技术等。

第4章介绍基于知识的数据库技术。主要以知识库为基础，介绍主动数据库技术、数据挖掘技术、决策支持系统等需要处理知识的数据库应用技术。

第5章介绍基于网络的数据库技术。主要包括分布式数据库技术、Web数据库技术、

移动数据库技术、XML 数据库、Web 搜索技术等。

第6章介绍数据库协同管理、数据交换与信息发布、工作流数据库等数据库协同技术。然后介绍时态知识数据库综合应用和两个综合应用实例，在实例中综合应用了面向对象技术、时态数据库技术、主动数据库、知识库、数据库协同工作等多种数据库技术。

最后，附录A给出了对于学习和研究数据库新技术有帮助的一些资源，包括重要学术会议、学术期刊和相关教材的介绍，附录B给出了时态数据库 TempDB 中 ATSQL 语言的 BNF 描述。

本书定位为“导论”，旨在为读者对数据库新技术的学习和研究起到一个“引导”作用。主要特色是基础技术、发展方向与综合实践相结合。

① 本书按现代数据库应用的主要特征——多维化、智能化、网络化、协同化——这4个方面来组织内容，分别介绍每个方面最基础的数据库技术，然后介绍几种相关的数据库新技术，使读者既掌握相关基础，又了解相关技术领域及其发展方向。

② 书中介绍了国际数据库界重要的学术会议和期刊，点评了现有的高级数据库相关著作，目的是培养读者的研究素质，引导读者了解数据库技术的新动向和新方向。这些学习资源是本书内容的延伸，这样可以保持本书内容的新颖性。

③ 注重理论研究和实践应用的结合，书中融入了我们近期数据库研究和开发成果，介绍的应用实例，涉及面向对象、时态数据库、主动数据库、知识库、协同技术等多种数据库新技术综合应用，有较好的示范性，启发读者在实际研究工作中注重多种研究成果的综合应用。

本书与我们编写的《数据库系统原理教程》、《数据库系统实验指导教程》（清华大学出版社）一同入选普通高等教育“十一五”国家级规划教材。我们在3部教材编写过程中，统筹安排，形成数据库课程链。前两部教材作为计算机专业必选课程教材，本书可作为高年级选修课和研究生教材。建议计划40~60学时，教学方式采用讲授与讨论相结合的方式，数据库技术部分采用讲授（32~40学时），根据书中讨论题组织数据库文献阅读和专题讨论（8~20学时），有条件的可以进行综合实例分析与实践。

本书由汤庸统筹，叶小平、汤娜、吴凌坤、李建国等参加编写主要内容，参加编写的还有马慧、刘海、左亚尧、陈国华、潘炎、何庆、冀高峰、朱君、郭欢等博士生。数据库课程组的老师印鉴教授、刘玉葆、万海、范昭赋博士以及“高级数据库技术”课程讨论班的研究生给予了许多支持，中国计算机学会数据库专业委员会副主任、四川大学唐常杰教

授审阅了书稿并提出了宝贵意见，本书的编写还参阅了大量的著作和相关文献，在此一并表示衷心的感谢！书中错漏之处，敬请指正。

本书编写工作得到国家自然科学基金项目（60673135、60373081），教育部新世纪优秀人才支持计划（NCET-04-0805）和广东省自然科学基金重点项目（重点04105503、7003721、05003348）、教育部-微软精品课程（2006）等项目支持。

编 者

2007年7月于中山大学

目 录

第 1 章 绪论	1
1.1 数据库技术的发展阶段	1
1.1.1 数据库技术的初级阶段	2
1.1.2 关系模型与关系数据库	3
1.1.3 第三代数据库技术	4
1.1.4 互联网时代的数据库技术	5
1.2 高级数据库技术概述	6
1.2.1 数据库技术的内涵和外延	6
1.2.2 高级数据库技术范畴	7
1.2.3 高级数据库技术应用特征	9
1.3 数据库技术的学术资源与 热点研究	10
1.3.1 数据库技术的主要 学术资源	11
1.3.2 数据库技术的研究热点	15
1.4 本书内容组织	17
思考讨论题	19
主要参考文献	19

第 2 章 关系数据库基础	20
2.1 关系数据模型	20
2.1.1 数据结构	20
2.1.2 数据操作	22
2.1.3 数据完整性约束	23

2.2 关系数据库标准语言	23
2.2.1 SQL 的发展与基本功能	23
2.2.2 关系定义	25
2.2.3 数据查询	27
2.2.4 数据更新	29
2.3 关系模式设计	31
2.3.1 函数依赖	31
2.3.2 Armstrong 公理系统	32
2.3.3 关系模式范式	37
2.3.4 多值依赖与连接依赖	38
2.4 关系数据库保护	43
2.4.1 完整性保护	43
2.4.2 安全性保护	46
2.5 关系数据库事务处理	47
2.5.1 并发控制	48
2.5.2 故障恢复	54
本章小结	56
思考讨论题	58
主要参考文献	58

第 3 章 高级数据模型	60
3.1 对象关系数据模型	60
3.1.1 对象关系数据模型	61
3.1.2 对象关系数据创建	65
3.1.3 对象关系数据操作	72

3.2 面向对象数据库	75	4.3.3 知识库系统	146
3.2.1 面向对象数据模型	75	4.3.4 数据库与知识库	147
3.2.2 ODMG 模型	78	4.4 基于知识的数据库技术	149
3.2.3 ODMG 数据定义	79	4.4.1 主动数据库技术	149
3.2.4 ODMG 数据查询	83	4.4.2 决策支持系统	153
3.3 时态数据模型与时态数据库	85	4.4.3 数据仓库技术	157
3.3.1 时间结构与时间演算	85	4.4.4 数据挖掘技术	173
3.3.2 时态数据库	88	本章小结	181
3.3.3 时态关系数据模型	92	思考讨论题	182
3.3.4 时态查询语言 TSQL2	95	主要参考文献	182
3.4 空间数据模型及其操作	101	第5章 网络环境数据管理	184
3.4.1 空间数据类型与数据操作	102	5.1 分布数据管理	184
3.4.2 空间点索引	111	5.1.1 分布式数据库系统	184
3.4.3 空间区域索引	113	5.1.2 分布式数据库体系结构	189
本章小结	116	5.1.3 分布式数据存储	192
思考讨论题	118	5.1.4 分布式数据查询	196
主要参考文献	118	5.1.5 分布式事务管理	199
第4章 智能数据处理	120	5.2 移动数据管理	201
4.1 知识的概念	120	5.2.1 移动数据管理概述	202
4.1.1 数据、信息与知识	120	5.2.2 移动数据库关键技术	204
4.1.2 知识类型	122	5.2.3 移动代理技术	209
4.2 知识表示与推理	124	5.2.4 移动数据管理与分布 数据管理	211
4.2.1 知识表示的基本概念	124	5.3 Web 数据管理	212
4.2.2 知识的基本表示方式	126	5.3.1 Web 数据库体系结构	212
4.2.3 知识的本体表示	134	5.3.2 Web 数据库访问连接	215
4.2.4 知识库语言	143	5.4 XML 数据管理	222
4.3 知识库与知识库系统	144	5.4.1 XML 基础	222
4.3.1 知识库概念	144	5.4.2 XML 数据库	233
4.3.2 知识库管理系统	145		

5.4.3 XML 数据库模式设计	234	6.2.4 TempKB 设计与实现	265
本章小结	238	6.3 大数据技术应用及实例	269
思考讨论题	239	6.3.1 大数据与数据库技术	270
主要参考文献	240	6.3.2 云计算与信息服务	271
第 6 章 高级数据库应用实践	241	6.3.3 社交网络与大数据	273
6.1 时态数据库管理系统设计	241	6.3.4 学术社交网络大数据应用 ...	277
6.1.1 TempDB 概述	241	本章小结	286
6.1.2 TempDB 体系架构	242	思考讨论题	286
6.1.3 时态 SQL 语法解析	243	主要参考文献	286
6.1.4 TempDB 应用示例	251	附录	288
6.2 数据库技术综合应用实例	256	附录 A 数据库学术资源	288
6.2.1 财政工资时态模型	257	附录 B TempDB 2.0 中 ATSQL2 的	
6.2.2 时态工资知识表达	260	BNF 定义	293
6.2.3 工资变更主动机制	263	附录 C “高级数据库技术”课程网站	296

第1章 绪 论

数据库技术从 20 世纪 60 年代中期产生至今已有近半个多世纪的历史，并且一直是最活跃、发展速度最快的 IT 技术之一。自 20 世纪 90 年代以来，随着互联网和多媒体技术的迅猛发展，应用中的新需求直接推动了数据库技术的研究与发展。

作为本书绪论，本章首先简要回顾数据库技术的发展，描述数据库的学科内涵及技术外延，其次讨论高级数据库的理论范畴和应用特征，同时介绍数据库学术资源与研究热点，最后介绍本书的组织结构。

1.1 数据库技术的发展阶段

数据库技术是随数据管理任务的需要而产生并随着对数据管理要求的提高而不断发展。在数据库出现以前，数据管理经历了人工管理阶段和文件管理阶段。

通常将数据库技术的发展分为 3 个阶段：第一阶段从 20 世纪 60 年末到 20 世纪 70 年代，此时数据库采用格式化数据模型即层次和网状模型，由此建立的数据库称为第一代数据库；第二阶段从 20 世纪 80 年代到 20 世纪 90 年代，此时数据库以关系模型为基础，称为第二代数据库；第三阶段是从 20 世纪 90 年代开始，此时数据库以面向对象为主要特征，适应多媒体、网络等众多应用，称为第三代数据库。进入 21 世纪，特别在云计算与大数据时代来临的今天，传统的数据库阶段划分已不能表达数据库技术的内涵和外延，数据库技术进入了一个新的发展时期。因此，可以提出将数据管理技术发展划分为 5 个阶段，如图 1-1 所示。

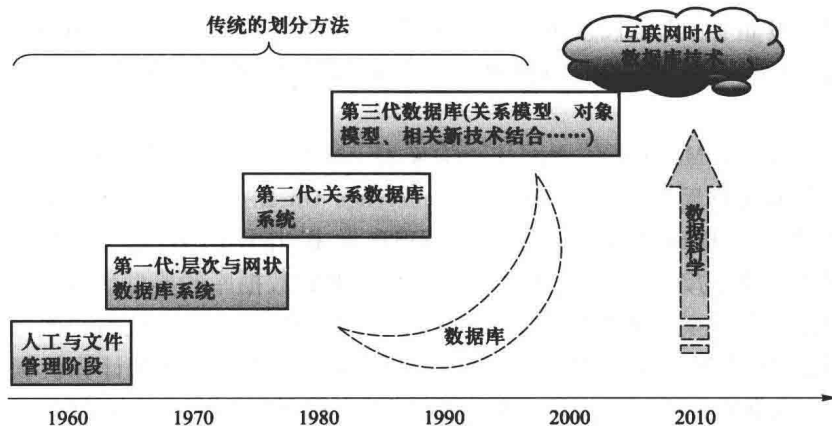


图 1-1 数据管理技术发展阶段划分

1.1.1 数据库技术的初级阶段

从历史发展来看,开发计算机的初衷是进行复杂科学计算,但随着计算机技术发展,其应用远远超出了科学计算范围,数据处理逐步成为计算机的主要用途。在计算机软件 and 硬件发展的基础上,数据管理也从人工管理、文件系统发展到20世纪60年代的数据库系统。

1. 人工管理与文件管理数据

20世纪50年代,计算机主要用于数值计算,当时的外存只有纸带、卡片和磁带,没有可直接存取的设备,也没有形成软件的整体概念,没有操作系统和数据管理软件,处理的数据量小,且数据间缺乏逻辑组织,数据依赖于特定的应用程序,缺乏独立性,这期间的数据由用户直接管理。

20世纪50年代后期到60年代中期,出现了磁鼓、磁盘等数据存储设备,新的数据处理系统是把计算机中的数据组织成相互独立的数据文件,系统可对文件中的记录进行存取,并可实现对文件的修改、插入和删除,这就是文件系统。文件系统实现了记录内的结构化,数据可长期保存。但是,文件中数据仍然是面向特定应用程序的,数据共享性和独立性差,且冗余度大,数据管理和维护的代价也很大。随着对数据管理性能要求的提高,如更高的共享性、更好的独立性和更有效率的数据查询与更新等,就要求数据管理的方法和技术要有新的突破,数据库技术应运而生。

2. 数据库技术产生

20世纪60年代的层次和网状数据模型库可看做是数据库技术的开端,它们为统一管理与共享数据提供了有力的支撑,以层次和网状模型为基础的数据库被称为第一代数据库系统。

层次模型(Hierarchical Model)使用树形结构来表示数据以及数据间的联系。层次数据库系统的代表是20世纪60年代末由IBM公司开发的信息管理系统(Information Management System, IMS),该系统用于帮助美国国家宇航局管理宏大的“阿波罗登月计划”中的烦琐资料。

网状模型(Network Model)使用网状结构来表示数据以及数据间的联系。20世纪60年代末70年代初,美国数据库系统语言协会(Conference On Data SYstem Language, CODASYL)下属的数据库任务组(Data Base Task Group, DBTG)提出了DBTG报告。DBTG报告确定并建立了网状数据库系统的许多概念、方法和技术,是网状数据库的典型代表。

这个时期代表人物是美国科学家巴赫曼(Charles W. Bachman),他因“数据库技术方面的杰出贡献”而获得1973年的计算机界最高奖ACM图灵奖(此奖被誉为计算机领域中的诺贝尔奖),并被称为“数据库先驱”和“网状数据库之父”。巴赫曼在1973年的图灵奖颁奖会上做了题为“The Programmer as Navigator”的演讲(刊载于1974年11月的Communications of ACM)。

如果说层次数据库是数据库系统的先驱,那么网状数据库则奠定了数据库的概念、方法和技术基础。在层次模型与网状模型基础上,数据库系统得到蓬勃发展,从而形成了历史上著名的“数据库时代”。但这两种数据库均脱胎于文件系统,具有比较简单的数据结构,受到物理结构的影响较大,用户使用数据库时需要对数据的物理结构有比较详细的了解,这给数据库的应用和推广带来了诸多困难与麻烦。

3. 层次和网状数据模型的主要贡献

层次数据库的数据模型对应定向有序树,网状模型对应有向图。这两种数据库奠定了现代数据库发展的基础,具有如下共同特点。

(1) 支持三级模式的体系结构

三级模式通常指外模式、模式、内模式。模式之间具有转换(或称为映射)功能。层次数据库和网状数据库均支持三级模式结构,通过外模式与模式、模式与内模式之间的映像,保证了数据库系统具有数据与程序的物理独立性和一定的逻辑独立性。

(2) 用存取路径表示数据之间的联系

数据库不仅存储数据,而且存储数据之间的联系。数据之间的联系在层次和网状数据库系统中是用存取路径来表示和实现的。这是数据库系统和文件系统的主要区别之一。

(3) 独立的数据定义语言

层次数据库系统和网状数据库系统都有独立的数据定义语言,用以描述数据库的外模式、模式、内模式以及相互映像。但是模式一经定义,就很难修改。修改模式必须首先把数据全部卸出,然后重新定义、重新生成诸模式,最后编写实用程序把卸出的数据按新模式的定义装入新数据库中。因此在许多实际运行的层次、网状数据库系统中,模式是不轻易重构的。这就要求数据库设计人员在建立数据库应用系统时,不仅应充分考虑用户的当前需求,还要充分了解需求可能的变化和发展。因此对数据库设计的要求比较高。层次和网状数据库通过独立的数据定义语言较好地处理了这一问题。

(4) 导航的数据操纵语言

所谓导航就是指用户不仅要了解“要干什么”,而且要指出“怎么干”。用户必须使用某种高级语言编写程序,一步一步地“引导”程序按照数据库中某一条预先定义的存取路径来访问数据库,最终达到要访问的数据目标。层次和网状数据库的数据查询和数据操纵语言通常都需嵌入某一种高级语言当中,如 COBOL、FORTRAN、PL/1 中。

1.1.2 关系模型与关系数据库

关系数据库系统形成于 20 世纪 70 年代中期并在 80 年代得到了充分的发展,它具有简单的结构方式并且使用方便。因此 20 世纪 80 年代开始它逐步取代层次和网状数据库,成为占主导地位的数据库,发展为第二代数据库系统。第二代数据库的主要特征是支持关系数

据模型。

1. 关系数据库的产生

1970年, IBM公司的San Jose研究试验室的研究员Edgar Frank Codd在其论文《大型共享数据库数据的关系模型》中开创性地提出了关系数据模型, 为关系数据库技术奠定了理论基础。正是由于其在关系数据库研究方面的杰出贡献, Codd在1981年被授予ACM图灵奖, 并被称为“关系数据库之父”。

2. 关系模型基本组成

关系数据库以关系模型为基础, 关系模型基本组成包括数据结构、关系操作、数据完整性3部分。关系数据结构包括域及域上定义的关系等。关系操作基于关系运算, 而关系运算包括基于集合论的关系代数和基于数理逻辑的关系演算。关系数据完整性包括实体完整性、参照完整性以及与应用有关的用户定义完整性。

3. 关系模型主要特点

关系模型概念简洁单一, 实体间的联系采用关系来表示; 以关系运算为数据操作的数学支撑, 易于形式化描述; 数据独立性强, 数据的物理存储和存取路径对用户透明; 关系数据库语言是非过程化的, 可以将用户从编程数据库记录的导航式检索中解脱出来, 大大降低了用户准入门槛。

4. 典型的关系数据库

20世纪70年代是关系数据库理论研究和原型开发的时代, 大量的理论成果和实践经验终于使关系数据库从实验室走向了社会, IBM公司的San Jose研究试验室开发的System R和加州大学伯克利分校研制的Ingres是典型代表。20世纪80年代几乎所有新开发的系统都是关系型数据库, 其中涌现出许多的商品化关系数据库管理系统, 如DB2、Ingres、Oracle、Informix、Sybase和SQL Sever等。

1.1.3 第三代数据库技术

20世纪80年代, 数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长。随着科学技术的不断进步, 各个行业领域对数据库技术提出了更多的需求, 关系数据库已不能完全满足需求, 以关系数据库为代表的传统数据库已经很难胜任新领域的要求。这就为新数据库技术的发展提供了强大的现实驱动。为了支持众多的实际应用需求, 必须将数据库技术与其他现代信息、数据处理技术, 如面向对象技术、时序和实时处理技术、人工智能技术、多媒体技术“完善”集成, 以形成第三代数据库技术。

1990年, 美国的高级DBMS功能委员会多位成员和关系数据库领域著名学者发表了《第三代数据库系统宣言》, 提出了第三代数据库管理系统应具有的特征, 包括3条基本原则和13

个命题。其基本思想如下。

1. 第三代数据库系统应支持数据管理、对象管理和知识管理

第三代数据库系统应集数据管理、对象管理和知识管理为一体，支持丰富的对象结构和规则，即面向对象数据模型是第三代数据库系统的重要特征。但由于面向对象技术方法的复杂性和其他各种原因，第三代数据库并没有像人们预期那样有效地取代关系数据库，关系数据库至今仍然占据着数据库领域的统治地位。

2. 第三代数据库系统必须保持或继承第二代数据库系统的技术

第三代数据库系统必须保持第二代数据库系统的非过程化数据存取方式和数据独立性，即第三代数据库是关系模式的扩展。在 1998 年，James Gray 成为第三位获得 ACM 图灵奖的数据库专家，他是 SQL 的主要奠基人，并且通过“事务处理技术”解决数据库的完整性（Integrity）、安全性（Security）、并行性（Concurrency）等问题。

3. 第三代数据库系统必须对其他系统开放

数据库系统的开放性表现在：支持数据库语言标准，支持标准网络协议，有良好的可移植性、可连接性、可扩展性和互操作性等。数据库技术的强大生命力在于能够与计算机各类新技术积极融合，同时也能有效适应计算机各类新型应用。例如，对于分布式计算学科有分布式数据库技术，对于人工智能学科有知识库技术，对于多媒体有多媒体数据库技术，对于 Internet 有 XML 数据库等；同时，对于辅助决策应用有 OLAP、数据仓库和数据挖掘，对于移动对象应用有移动对象数据库，对于地理信息和遥感应用有空间数据库和时空数据库技术等。这些数据库新技术研究的发展导致了众多不同于第二代数据库的新型系统诞生，构成了当今数据库系统的大家族。

尽管这些数据库中数据模型大多具有面向对象的基本特征，但却难以像第一代和第二代数据库那样基于某个统一的数据模型。“第三代数据库”概念并没有得到普遍认可，人们通常使用“新一代数据库”或“现代数据库”概念来描述这种情形。

1.1.4 互联网时代的数据库技术

20 世纪末进入互联网时代，数据库技术的内涵和外延发生了很大变化，数据库管理从数据查询发展到信息检索、数据挖掘和知识发现等应用，但是关系模型仍然是数据库技术及应用的核心技术和基础理论。

进入 21 世纪，特别在云计算与大数据时代来临的今天，从数据库到数据科学，又迎来了前所未有的挑战和发展机遇。本书将在后续章节介绍相关技术。