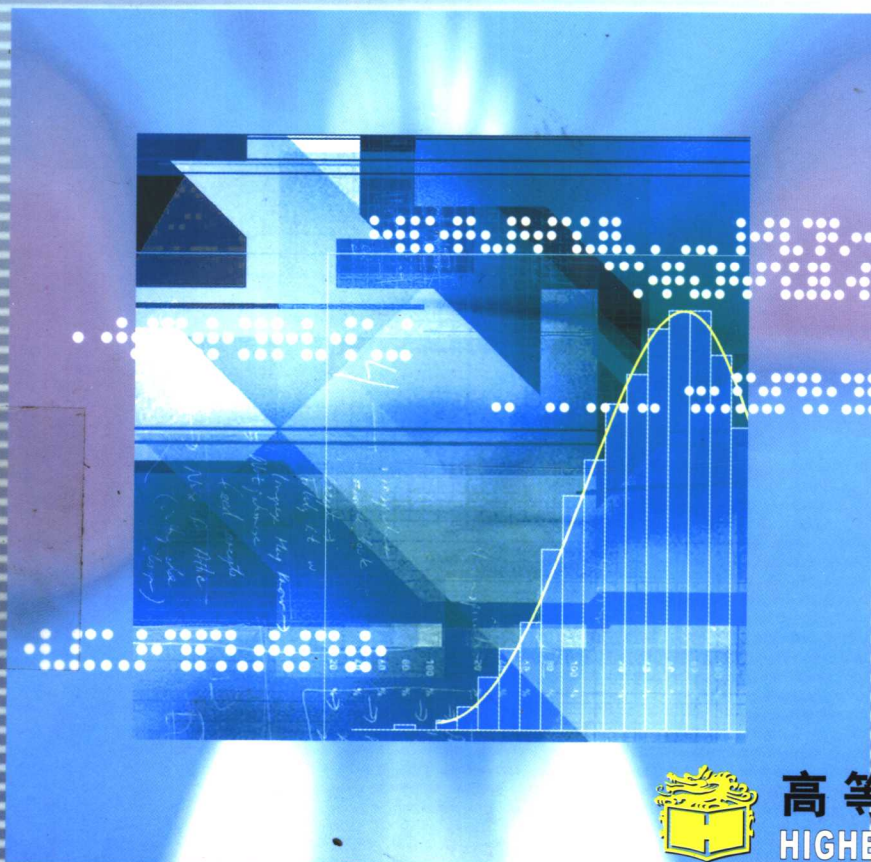


● 高等学校研究生系列教材

高级数据库技术

Advanced Database Technology

汤 庸 叶小平 汤 娜 吉永杰



高等教育出版社
HIGHER EDUCATION PRESS

内容提要

本书系统介绍现代数据库领域中的有代表性的新技术。全书共 12 章,第 1 章简要回顾数据库基础知识并概述数据库新技术的发展,第 2 章讨论对象关系数据库和面向对象数据库,第 3 章到第 5 章讨论时态数据库、实时数据库、主动数据库技术,第 6 章到第 7 章介绍 Web 数据库技术、移动数据库技术,第 8 章到第 11 章介绍多媒体数据库、知识库技术、空间数据库、数据仓库与数据挖掘技术等,第 12 章介绍协同数据库并列举一个综合实例。

本书可作为高校计算机专业本科高年级学生或研究生相关课程的教材或教学参考书,也可供广大数据库研究与开发人员参考使用。

图书在版编目(CIP)数据

高级数据库技术/汤庸等编著. —北京:高等教育出版社,2005.1

ISBN 7-04-015863-9

I. 高... II. 汤... III. 数据库系统—高等学校—教学参考资料 IV. TP311.13

中国版本图书馆 CIP 数据核字(2004)第 126311 号

策划编辑 倪文慧 责任编辑 倪文慧 市场策划 陈 振
封面设计 王凌波 责任印制 孔 源

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 100011
总 机 010-58581000

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京市南方印刷厂

开 本 787×1092 1/16
印 张 25.25
字 数 540 000

版 次 2005 年 1 月第 1 版
印 次 2005 年 1 月第 1 次印刷
定 价 32.00 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:15863-00

前 言

数据库技术从 20 世纪 60 年代中期产生至今已有 40 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三代: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 年代末到 80 年代得到广泛应用;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了更广泛的拓展,数据库技术发展进入了一个新的时期,现代数据库技术融合多种技术,数据库新技术在不断产生和发展。

现代数据库应用范围非常广泛,涉及许多相关技术。在实际信息系统和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。现代数据库技术所涵盖的范围很大,难以在一本书中详细介绍每种数据库技术,本书只是选取了有代表性的一些数据库技术,目的是通过所介绍的内容使读者对数据库新技术有一个大体的印象,为进一步从事数据库研究和开发提供一些思路。

全书共 12 章,大致分为五部分。

第一部分是本书的预备知识和基础知识,包括第 1 章和第 2 章。第 1 章简要回顾数据库系统基础概念和理论,概述数据库新技术的发展等。第 2 章讨论对象数据库技术,主要包括面向对象数据库基本概念和关系数据库的扩展:对象关系数据库。

第二部分是数据库系统方面的技术延伸,包括第 3 章至第 5 章。第 3 章介绍时态数据库技术,第 4 章介绍实时数据库技术,第 5 章介绍主动数据库技术。这些数据库技术在数据库原理上无根本的变化,但是在系统的体系结构方面采用和集成了新的技术。

第三部分介绍基于分布环境下的数据库技术,主要包括第 6 章 Web 数据库和第 7 章移动数据库。按照人们一般的观点,数据库技术与计算机网络技术的彼此融合和相互促进,将是数据库发展的基本方向之一。

第四部分是数据库研究与应用的较深入的课题,包括第 8 章多媒体数据库技术、第 9 章知识库系统、第 10 章空间数据库和第 11 章数据仓库。

最后一部分为第 12 章,介绍数据库的协同工作技术和一个综合应用实例。在这个综合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。

本书作者长期从事数据库教学和科研工作,承担了多项数据库方面国家自然科学基金、省市重点项目以及国际国内合作研究与开发项目,并从 1999 年起先后给研究生和本科生开设了高级数据库课程。本书是在教学讲义的基础上,结合科研成果完成的。限于作者水平,

书中难免存在缺陷和错误,衷心希望广大读者提出建议和批评意见。

本书由汤庸统筹,叶小平、汤娜、吉永杰编写了主要内容,左亚尧老师编写了第 11 章、潘炎博士编写了第 6 章,参加编写的还有毛承洁老师,何庆博士,冀高峰博士,研究生何炜、谭柱成、康向峰、谢建勤、张励、邢军等。本书的编写参考了大量的国内外有关文献,其中主要的已在书后参考文献中列出,在此对其作者表示诚挚的谢意。

本书可作为高等学校计算机专业本科高年级学生或研究生课程教材或教学参考书,也可供广大数据库研究与开发人员参考使用。

作 者

2004 年 11 月于中山大学

目 录

第1章 绪论	(1)	2.2.6 面向对象的结构化方法	(39)
1.1 数据库系统基本概念	(1)	2.3 面向对象数据模型	(41)
1.1.1 数据库系统	(1)	2.3.1 基本概念	(41)
1.1.2 数据库体系结构	(3)	2.3.2 基本模型	(41)
1.1.3 数据库二级映射	(6)	2.4 面向对象数据库系统	(44)
1.2 数据模型	(7)	2.4.1 面向对象数据库系统	(44)
1.2.1 概念数据模型	(7)	2.4.2 面向对象数据库管理系统	(45)
1.2.2 结构数据模型	(9)	2.5 面向对象数据操作	(46)
1.2.3 层次模型	(10)	2.5.1 数据定义	(46)
1.2.4 网状模型	(11)	2.5.2 扩充的数据类型及抽象 数据类型	(49)
1.2.5 关系模型	(12)	2.5.3 数据查询	(50)
1.2.6 面向对象模型	(13)	2.6 对象关系数据模型	(54)
1.3 数据库理论基础	(14)	2.6.1 关系数据模型扩充	(55)
1.3.1 关系数据库规范化	(14)	2.6.2 组合类型	(56)
1.3.2 数据库完整性	(16)	2.6.3 聚集类型	(56)
1.3.3 数据库的事务处理	(18)	2.6.4 引用类型	(58)
1.3.4 数据索引与查询处理	(20)	2.6.5 对象与类型	(58)
1.4 数据库新技术概述	(23)	2.7 ORDB 中的定义语言	(59)
1.4.1 数据模型的新特征	(23)	2.7.1 数据类型定义	(59)
1.4.2 数据库系统的新要求	(24)	2.7.2 继承性的定义	(61)
1.4.3 数据库新技术的范畴	(25)	2.7.3 引用类型的定义	(62)
1.4.4 现代数据库实现途径	(26)	2.8 ORDB 中的查询语言	(63)
本章小结	(28)	2.8.1 以关系为值的属性	(63)
习题	(28)	2.8.2 路径表达式	(64)
第2章 对象数据库技术	(29)	2.8.3 函数的定义和使用	(65)
2.1 对象数据库系统概述	(29)	2.9 对象关系数据库系统	(65)
2.2 面向对象方法	(30)	本章小结	(67)
2.2.1 面向对象与面向计算机	(30)	习题	(67)
2.2.2 面向对象方法基本概念	(31)	第3章 时态数据库技术	(69)
2.2.3 类的概念	(34)	3.1 时态数据库的起源与发展	(69)
2.2.4 继承关系	(34)	3.1.1 开创与形成时期	(69)
2.2.5 组合关系	(38)		

3.1.2 理论与模型探索时期	(70)	4.2.4 实时数据库与其他高级数据库 的联系	(124)
3.1.3 应用与发展时期	(71)	4.3 实时数据库的特征	(125)
3.1.4 研究与应用现状	(72)	4.3.1 RTDB 的数据特征	(126)
3.2 基本时态元素	(73)	4.3.2 RTDB 的事务特征	(127)
3.2.1 时间模型	(73)	4.4 实时数据库管理系统	(129)
3.2.2 时间点	(74)	4.4.1 实时数据库系统的主要问题	(129)
3.2.3 时间区间与时间跨度	(74)	4.4.2 实时数据库系统结构	(130)
3.2.4 时间粒度与时间量子	(75)	4.4.3 实时数据库执行模型	(132)
3.2.5 时间元素 Now 与 UC	(78)	4.5 实时事务模型	(132)
3.3 时态数据库基本概念	(80)	4.5.1 实时事务的建模特性	(133)
3.3.1 三种基本时间	(80)	4.5.2 实时事务的分类	(133)
3.3.2 两个时间维	(82)	4.5.3 实时事务的特性	(135)
3.3.3 快照数据库	(85)	4.5.4 实时事务模型	(136)
3.3.4 回滚数据库	(86)	4.5.5 嵌套实时事务模型	(136)
3.3.5 历史数据库	(87)	4.6 实时事务调度与并发控制	(137)
3.3.6 双时态数据库	(88)	4.6.1 实时事务调度	(138)
3.4 时态关系操作	(89)	4.6.2 实时事务的并发控制	(138)
3.4.1 基本概念和术语	(89)	4.7 实时数据库应用	(140)
3.4.2 HRDM 数学模型	(90)	4.7.1 国内外实时数据库的发展 现状	(140)
3.4.3 历史关系代数	(91)	4.7.2 实时数据库系统在企业信息化 中的应用	(141)
3.5 时态数据库查询语言	(95)	4.7.3 基于实时数据库的高级数据 应用技术	(141)
3.5.1 双时态概念数据模型 BCDM	(96)	本章小结	(145)
3.5.2 TempSQL 模型及语言	(98)	习题	(145)
3.5.3 TQuel 模型及语言	(102)	第 5 章 主动数据库技术	(146)
3.5.4 TSQL2 语言	(107)	5.1 主动数据库的产生	(146)
3.6 时态数据库软件 TimeDB	(112)	5.1.1 数据库的被动服务与主动 服务	(146)
3.6.1 TimeDB 基础	(113)	5.1.2 实际应用的主动性需求	(147)
3.6.2 TimeDB 的时态扩展	(114)	5.1.3 什么是主动数据库	(147)
3.6.3 与商业 DBMS 兼容性问题	(116)	5.2 触发器技术	(148)
3.6.4 一个例子	(118)	5.2.1 标准 SQL 中的主动数据库 功能	(148)
本章小结	(119)	5.2.2 在商业 DBMS 中的触发器	(149)
习题	(120)	5.2.3 触发器的使用	(150)
第 4 章 实时数据库技术	(121)		
4.1 实时应用的数据处理	(121)		
4.2 实时数据库简介	(121)		
4.2.1 实时数据库系统的定义	(122)		
4.2.2 实时系统的主要特性	(122)		
4.2.3 实时数据库与传统数据库	(123)		

5.2.4 触发器的优缺点	(152)	6.3 Web 数据库的数据交换技术	(183)
5.3 主动数据库体系结构	(153)	6.3.1 XML 技术的基本思想	(183)
5.4 主动(ECA)规则	(155)	6.3.2 主要的数据库交换技术	(189)
5.4.1 ECA 规则的构成	(155)	6.4 Web 数据库中的安全问题	(190)
5.4.2 ECA 规则描述	(156)	6.4.1 防火墙技术	(190)
5.4.3 事件	(157)	6.4.2 用户身份认证	(191)
5.4.4 条件	(161)	6.4.3 授权控制	(192)
5.4.5 动作	(162)	6.4.4 监视跟踪	(193)
5.5 主动数据库的实现	(163)	6.4.5 审计	(193)
5.5.1 实现途径	(163)	6.4.6 备份与故障恢复	(194)
5.5.2 事件监视器	(164)	本章小结	(194)
5.5.3 基于组件的主动数据库模型 探讨	(165)	习题	(194)
5.6 面向主动对象的数据库	(167)	第 7 章 移动数据库	(196)
5.6.1 面向主动对象数据库的组成 ..	(168)	7.1 移动数据库简介	(196)
5.6.2 主动对象类及面向主动对象的 数据库模型	(168)	7.1.1 移动数据库与移动计算	(196)
5.6.3 面向主动对象数据库的 ECA 规则	(169)	7.1.2 移动数据库的特点	(197)
5.6.4 面向主动对象数据库的实现 途径	(169)	7.1.3 移动数据库的典型系统模型 ..	(198)
5.6.5 面向主动对象数据库的优点 ..	(170)	7.2 移动数据库的关键技术	(199)
本章小结	(170)	7.2.1 移动数据库的复制与 缓存技术	(199)
习题	(171)	7.2.2 数据广播	(203)
第 6 章 Web 数据库	(172)	7.2.3 移动查询技术	(204)
6.1 Web 数据库的体系结构	(172)	7.2.4 移动事务处理技术	(207)
6.1.1 Web 数据库的主要优点	(172)	7.3 移动 Agent 技术	(210)
6.1.2 主机集中式模式	(173)	7.3.1 Agent 与移动 Agent	(211)
6.1.3 客户端/服务器模式	(173)	7.3.2 开发移动 Agent 系统需要注意的 问题	(213)
6.1.4 三层结构的浏览器/服务器 模式	(174)	7.3.3 Agent 对移动数据库系统的 支持	(214)
6.1.5 $N(N \geq 4)$ 层结构的浏览器/服 务器模式	(175)	7.3.4 基于 Agent 的移动数据库 模型	(214)
6.2 Web 数据库的访问技术	(176)	7.4 一个移动数据库的应用例子	(222)
6.2.1 ODBC	(176)	7.4.1 RMS 存储系统简介	(222)
6.2.2 OLE DB	(177)	7.4.2 股票查询示例程序的系统结构 及相关实现	(222)
6.2.3 JDBC	(180)	7.4.3 移动数据库的应用前景展望	(225)
6.2.4 对象/关系映射	(181)	本章小结	(226)
		习题	(226)

第 8 章 多媒体数据库技术	(228)	9.2.3 知识的产生式表示	(261)
8.1 多媒体数据库概述	(228)	9.2.4 知识的语义网络表示	(262)
8.1.1 多媒体数据	(228)	9.2.5 知识的框架表示	(263)
8.1.2 多媒体数据分类	(229)	9.3 面向对象的时态知识表达方法	(266)
8.1.3 多媒体数据特征	(230)	9.3.1 知识对象的结构	(266)
8.1.4 多媒体数据库	(233)	9.3.2 描述事实的时态知识	(267)
8.1.5 多媒体数据库的主要技术	(234)	9.3.3 描述时序关系的时态知识	(268)
8.2 多媒体数据模型	(235)	9.3.4 面向对象时态知识表达举例	(269)
8.2.1 多媒体数据模型的层次	(235)	9.4 知识库系统原理	(272)
8.2.2 扩展的关系数据模型	(237)	9.4.1 知识库系统中的一阶谓词逻辑 表示	(272)
8.2.3 面向对象模型	(237)	9.4.2 子句	(273)
8.2.4 语义数据模型	(238)	9.4.3 Horn 子句	(274)
8.3 多媒体元数据及其生成	(238)	9.4.4 Datalog 表示方法	(274)
8.3.1 多媒体元数据	(238)	9.4.5 基于证明论的知识库系统	(275)
8.3.2 多媒体元数据生成	(239)	9.4.6 基于模型论的知识库系统	(277)
8.4 多媒体数据库查询	(242)	9.5 知识库系统体系结构	(279)
8.4.1 多媒体数据库查询实例	(242)	本章小结	(280)
8.4.2 多媒体数据库的查询类型	(243)	习题	(280)
8.4.3 查询过程的选择	(243)	第 10 章 空间数据库	(281)
8.4.4 多媒体数据库查询语言	(244)	10.1 空间数据库概述	(281)
8.5 多媒体数据库管理系统	(246)	10.1.1 空间数据库的意义	(281)
8.5.1 多媒体数据库视图	(247)	10.1.2 空间数据的基本特征	(282)
8.5.2 物理存储视图	(248)	10.1.3 空间数据库作为常规数据库的 扩充	(283)
8.5.3 概念数据视图	(248)	10.2 空间数据模型	(283)
8.5.4 分布视图	(251)	10.2.1 空间数据模型	(283)
8.5.5 过滤视图	(251)	10.2.2 空间对象所处的环境	(284)
8.5.6 用户视图	(252)	10.2.3 空间对象之间关系	(284)
本章小结	(253)	10.2.4 空间数据操作的谓词描述	(285)
习题	(253)	10.2.5 空间关系的集合描述与 判断	(287)
第 9 章 知识库系统	(254)	10.2.6 空间关系的代数描述与 运算	(288)
9.1 知识库系统概论	(254)	10.2.7 空间数据查询语言	(289)
9.1.1 知识的概念	(254)	10.3 空间索引	(289)
9.1.2 知识库、知识库管理系统和知识 库系统	(256)	10.3.1 空间索引概述	(290)
9.1.3 知识库语言	(259)	10.3.2 空间对象的近似表示	(291)
9.2 传统的知识表示模式	(260)		
9.2.1 知识表示基本概念	(260)		
9.2.2 知识的谓词逻辑表示	(260)		

10.3.3 基于大小非零空间对象查询 的 R 树	(292)	11.2.6 确定粒度的级别	(326)
10.4 空间数据库的系统结构	(295)	11.2.7 确定分割标准	(327)
10.4.1 以 RDBMS 为基础的空间数据库 系统结构	(295)	11.2.8 数据仓库的数据组织	(327)
10.4.2 以 OODBMS 或 ORDBMS 为基础 的集成空间结构	(296)	11.2.9 数据预处理	(328)
10.5 地理信息系统 (GIS)	(296)	11.2.10 数据维护模块	(331)
10.5.1 GIS 概述	(296)	11.2.11 数据仓库的元数据	(332)
10.5.2 地理信息系统的组成	(297)	11.2.12 数据仓库管理	(334)
10.5.3 Oracle Spatial 在 GIS 中 的应用	(299)	11.3 数据仓库的应用技术	(335)
10.5.4 GIS 在各领域中的应用	(300)	11.3.1 OLAP 技术	(335)
10.5.5 GIS 应用实例	(302)	11.3.2 数据挖掘	(339)
10.6 时空数据库	(304)	11.3.3 领域数据仓库的数据挖掘	(343)
10.6.1 时空变化的分类	(304)	本章小结	(348)
10.6.2 时空数据建模	(305)	习题	(348)
10.6.3 时空索引和时空数据查询	(306)	第 12 章 协同数据库与综合实例	(350)
本章小结	(307)	12.1 数据库中的协同工作	(350)
习题	(307)	12.1.1 计算机支持的协同工作	(350)
第 11 章 数据仓库	(309)	12.1.2 基于 XML 的数据库协同 工作	(354)
11.1 关于数据仓库	(309)	12.1.3 时态 XML 扩展	(356)
11.1.1 从事务数据到决策知识	(309)	12.2 工作流技术	(359)
11.1.2 数据库与数据仓库	(312)	12.2.1 工作流技术的起源与发展	(359)
11.1.3 数据仓库的特征	(314)	12.2.2 工作流的定义	(360)
11.1.4 数据仓库类型	(315)	12.2.3 工作流的相关概念	(361)
11.1.5 数据仓库系统的组织与 体系结构	(316)	12.2.4 工作流管理系统 (WFMS)	(362)
11.1.6 数据仓库的典型应用	(317)	12.2.5 工作流管理系统的分类	(363)
11.2 构建数据仓库	(318)	12.2.6 WfMC 工作流参考模型	(364)
11.2.1 数据仓库的开发流程	(318)	12.2.7 现代的工作流产品介绍	(366)
11.2.2 确定主题	(319)	12.3 一个综合实例	(368)
11.2.3 数据仓库的多维建模	(320)	12.3.1 SIDSS 系统框架	(368)
11.2.4 数据仓库的数据建模步骤	(323)	12.3.2 工资信息的时态性	(369)
11.2.5 确定数据仓库的集成和维护 体系结构	(324)	12.3.3 工资的时态数据模型	(372)
		12.3.4 工资政策时态知识模型	(376)
		12.3.6 系统的协同机制	(385)
		本章小结	(386)
		习题	(387)
		参考文献	(388)

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40多年历史,一般来说,数据库发展可以分为三代: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年代末到80年代得到广泛应用;20世纪90年代,随着网络 and 多媒体技术的迅猛发展,数据库的应用也得到更为广泛的拓展,数据库技术发展进入了一个新的时期,现代数据库技术融合多种技术,面向对象是其主要特征。

作为本书的预备知识,本章简要回顾数据库基础知识和基本概念,主要包括数据模型、关系数据库理论、数据库系统的一些基本概念,最后讨论新一代数据库的需求和特征,新一代数据库技术的范畴和实现途径。

1.1 数据库系统基本概念

本节简要回顾数据库的几个基本概念:数据库、数据库管理系统(DBMS)、数据库体系结构、数据库概念模式/内模式映射、外模式/概念模式映射等。

1.1.1 数据库系统

数据库系统(Database System, DBS)是指一个计算机存储记录信息的系统,即:DBS首先是一个计算机系统;该系统的目标是存储数据信息并支持用户检索和更新所需要的数据信息。

在人们的使用当中,DBS通常有下述两种含义。

DBS的个体含义:指一个具体的数据库管理系统软件和它建立起来的数据库。

DBS的科学含义:指研究、开发、建立、维护和应用数据库系统所涉及的理论、方法和技术所构成的科学领域。在这一含义下,数据库系统是软件研究领域的一个重要的分支,常称为数据库领域。本书中的DBS正是基于这种科学含义的。

数据库系统由数据库、数据库管理系统、支持数据库运行的软、硬件环境和用户等四部分组成。

1. 数据库

数据库可以看做是一个电子文件柜,它是基于计算机系统的持久性数据的“仓库”或者

“容器”。是一个长期存储在计算机内、有组织的和可共享的数据集合。其特点是集中了各种应用的数据,并对其进行统一的构造与存储,同一数据库可以为不同的应用服务。

2. 数据库管理系统 (DBMS)

数据库是一个多级系统结构,需要一组软件提供相应的工具进行数据的管理和控制,以达到保证数据的安全性和一致性的基本要求。这样一组软件就是数据库管理系统(Database Management System, DBMS)。它具有数据组织定义、数据操作与查询优化、数据控制及数据维护、数据管理以及提供各种接口等功能。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件(如图 1-1 所示),是整个 DBS 的核心与主体。

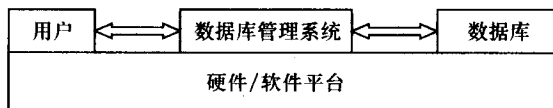


图 1-1 数据库系统结构图

由于对数据库的查询操作和优化处理由 DBMS 完成,数据库可以独立于具体应用程序;同时由于 DBMS“隔离”了用户和数据库中的实际数据,保证数据库中应用程序与数据的独立性,数据库又可以为多个不同的用户共享。数据独立性为数据库系统广泛有效的应用奠定了基础,数据共享性节省了大量人力物力,DBMS 正是在这两个基本点上发挥着重要作用。

3. 支持数据库系统的硬件/软件平台

(1) 硬件平台

数据库系统中的硬件平台可以有狭义和广义之分。

狭义硬件平台是指数据存储和数据处理所必不可少的硬件设施,它们主要有:

① 二级存储设备:其中包括相关的 I/O 设备(磁盘驱动器等)、设备控制器和 I/O 通道等。二级存储设备(大部分为磁盘)用来存放数据;

② 硬件处理器和相应的主存:主要用于支持数据库系统软件的执行。

广义硬件平台是指具有广泛应用和较强功能的硬件设施,一般包括以下两类:

① 计算机:这是系统硬件的基础平台,常用的有微型机、小型机、中型机和大型机。

② 网络:过去数据库一般建在单机上,现在较多建在网络上,从发展趋势来看,数据库系统今后将以建在网络上为主,其结构形式又以客户机/服务器(C/S)方式和浏览器/服务器(B/S)方式为主。

(2) 软件平台

操作系统:数据库系统的基础软件平台,目前常用有 Windows 与 UNIX (Linux)等。

数据库系统开发工具:例如实用程序、应用开发程序、设计辅助、报表书写器和事务管理

器与事务监控器(TP Monitor)。这些包括过程程序设计语言(C、C++)、可视化开发工具(Visual Basic、Power Builder、Delphi)以及近期与 Internet 有关的 HTML 和 XML 等。

接口软件:在网络环境下的数据库系统中,数据库与应用、数据库与网络间存在着多种接口,它们需要用接口软件进行连接,否则数据库系统整体就无法运作,这些接口软件包括 ODBC、JDBC、LECB、CORBA、COM、DCOM 等。

4. 用户

一般认为,存在着三类数据库用户(当然它们相互之间可以有所重叠)。

第一类用户:应用程序员。应用程序员负责编写数据库应用程序,这些程序通过向 DBMS 发出数据库操作语句请求来访问数据库,这些程序通常可以是具有批处理特征或者联机特征的应用程序,目的是允许最终用户通过联机工作站或者终端访问数据库。

第二类用户:最终用户。最终用户通过联机工作站或者终端与数据库系统进行交互。实行交互的应用软件为数据库系统本身固有,无需用户自己编写。

第三类用户:数据库管理员。由于数据库的共享性,数据库的规划、设计、维护和监视需由专人管理,这就是数据库管理员(Database Administer, DBA)。

从数据库技术运行的角度来说,数据库管理员是三类用户中的灵魂人物。DBA 需要根据企业的数据情况与要求,制定数据库建设与维护的策略,并对这些策略的执行提供技术支持。数据库管理员负责技术层的全局控制。

DBA 的主要工作包括以下内容:

数据库设计(Database Design):由于数据库的集成性与共享性,必须有专人对多个应用的数据需求作全面的规划、设计和集成,这是 DBA 的基本任务。

数据库维护(Database Maintenance):DBA 必须对数据库中数据的安全性、完整性、并发控制及系统恢复进行实施与维护。

改善系统性能和提高系统效率:DBA 必须随时监视数据库运行状态,不断调整内部结构,保持系统的最佳状态与最高效率。

1.1.2 数据库体系结构

在实际应用当中,DBS 的种类很多,它们可以支持不同的数据模型,使用不同的数据库语言,建立在不同的操作系统之上,数据的存储结构也各不相同,但它们在体系结构上通常都具有相同的特征,即采用三级模式结构(早期微型计算机上的小型数据库系统除外)并提供两级映像功能。掌握数据库的三级结构及其联系与转换应当是深入学习和掌握数据库理论、技术与方法的必由之路。

模式(Schema)是数据库中全体数据的逻辑结构和特征的描述,其主体就是数据模型,此外,一般还包括允许的各种操作、数据完整性和安全性等方面的控制。模式的一个具体值称为模式的一个实例(Instance)。同一模式可以有很多实例。模式是相对稳定的,而实例是经

常变动的;模式反映数据的结构及其联系,而实例反映数据库某一时刻的状态。

DBS 体系结构分为三层:外模式、概念模式和内模式。

1. 外模式

(1) 外模式概念

外模式(External Schema),也称用户模式(User's Schema)或子模式(Subschema)。外模式就是用户所看到的数据视图,它最接近用户,是用户和数据库系统的接口,是用户给出的对数据的描述。外模式是与某一具体应用有关的数据的逻辑结构和特征描述。概念模式给出系统全局的数据描述,外模式则给出每个用户的局部描述,即外模式为用户所见到的概念模式的一个部分。由于外模式通过概念模式推导而出,一般可以看做是概念模式的一个子集。

(2) 外模式的不惟一性

不同的用户由于需求的不同,看待数据的方式也会不同,对数据的保密要求及使用的程序设计语言也会不同,从而不同用户的外模式的描述方式一般是互不相同的。即使是对概念模型相同的数据,在外模式也可以对结构、类型、长度和保密级别等进行不同的限制,从而产生不同的外模式。这样,一个概念模式可以有若干个外模式,每个用户只关心与其有关的外模式。由此可以屏蔽大量与特定需求无关的信息,有利于数据保护,对数据所有者和用户都极为方便。

(3) 外模式 DDL

在一般的 DBMS 中都提供相关的外模式描述语言(外模式 DDL)。外模式 DDL 在嵌入形式下和用户选用的程序设计语言具有相容的语法格式。例如在 PowerBuilder 中使用外模式 DDL 必须符合 PowerBuilder 的语法要求。

2. 概念模式

(1) 概念模式

概念模式(Conceptual Schema),也称公共逻辑模式或逻辑模式,它是介于内模式和外模式之间的层次,与结构数据模型对应。概念模式是数据库系统中全局数据逻辑结构和特征的描述,是全体用户的公共数据视图,这种描述是一种抽象描述,不涉及具体硬件平台与软件环境。概念模式主要描述数据的概念记录类型和它们之间的关系,还包括一些数据间的语义约束。概念模式是惟一的。

(2) 概念模式 DDL

DBMS 提供概念模式的 DDL 语言(概念 DDL)来严格定义概念模式。定义模式时不仅要定义数据的逻辑结构,而且还要定义数据之间的联系,定义与数据有关的安全性和完整性要求。更一般的观点认为,概念模式的根本目的是描述整个组织的情况,除了数据本身之外,还要描述数据的流动等情况。但是目前的系统实际上还不能支持这种程度的概念模式,应用中的系统支持的“概念模式”实际上只是把单个的外模式合并起来,再加上一些安全性

和完整性的约束。但是将来的系统很可能会支持更加复杂、功能更加强大的概念模式。

3. 内模式

(1) 内模式概念

内模式 (Internal Schema) 也称为存储模式 (Storage Schema) 或物理模式 (Physical Schema), 是数据库物理结构和存储方式的描述, 即数据库的“内部视图”。内部视图是整个数据库的底层表示, 它由内部记录型中各个类型的值组成。内模式定义了数据库中的各种存储记录、存储记录的物理表示、存储结构与物理存取方法, 如数据存储的文件结构、索引、集簇等存取方式和存取路径等。一个数据库只有一个内模式。

(2) 内模式不深入到设备级

内模式虽然称为物理模式, 但它的物理性质主要表现在操作系统级和文件级上, 本身并不深入到设备级上 (如磁盘及磁盘操作), 内部视图仍然不是物理层, 它不涉及物理记录的形式, 例如物理块或页, 也不考虑具体设备的柱面与磁道大小。因此只能说, 内模式是最接近物理存储的数据的存储方式。当然, 近年来内模式也有向设备级发展的趋势 (如原始磁盘、磁盘分块技术等)。

(3) 内模式 DDL

DBMS 一般提供相关的内模式描述语言 (内模式 DDL) 来严格定义内模式。当前, 许多 DBMS 产品基本上都可以自动完成内模式的大部分定义工作, 几乎不需要用户介入。当用户使用内模式 DDL 定义模式的同时, 通常 DBMS 也就自动完成了相应的内模式的定义工作。

4. 三种模式的关系

在数据库系统体系结构当中, 内模式处于最低层, 它最接近于反映数据在计算机物理结构中的实际存储形式; 概念模式处于中层, 它反映了设计者的数据全局逻辑要求; 而外模式处于最外层, 它反映了用户对数据的实际要求。

三种模式给出了数据库系统的体系结构, 它说明作为数据库主体的数据必须按照这些模式所描述的框架结构进行组织。以概念模式为框架组织的数据库称为概念数据库 (Conceptual Database), 以内模式为框架组成的数据库称为物理数据库 (Physical Database), 以外模式为框架组成的数据库称为用户数据库 (User's Database)。当然, 这三种数据库中只有物理数据库是真实存在于计算机外存中, 其他两种数据库并不真正存在于计算机之内, 而是通过所谓“二级映射”功能由物理数据库映射而成。

在关系系统当中, 如下关系一般是成立的:

(1) 概念模式一定是关系的, 在该层可见的实体是关系的表和关系的操作符。

(2) 外部视图也是关系的或接近是关系的。

(3) 内模式不是关系的, 因为该层的实体不是关系表的照搬。事实上, 不管是什么系统, 其内模式都是一样的 (如存储记录、指针、索引、哈希表等), 关系模式与内模式无关。

1.1.3 数据库二级映射

数据库系统体系结构三级模式实质上是对数据的3个级别抽象,它的基本意义在于将DBS中数据的具体物理实现留给物理模式,使得用户与全局设计者不必关心数据库的具体实现与物理背景。为了能够保证在数据库系统内部实现这3个抽象层次的联系和转换,还必须在这3个模式之间提供两个(两级)映射,这就是概念模式/内模式映射和外模式/概念模式映射。

1. 概念模式/内模式映射

该映射定义了概念视图和数据库的对应关系,它说明了概念记录 and 字段在内部层次怎样表示。如果数据库的存储结构发生改变,即变动了存储结构的定义,概念模式/内模式映射也必须进行相应的改变,以保证概念模式能够保持不变(这是数据库管理员的工作)。也就是说,为了保持数据的物理独立性,内模式变化所带来的影响必须与概念模式隔离开来。

2. 外模式/概念模式映射

该映射定义了特定的外部视图和概念视图之间的对应关系。一般而言,这两层之间存在的差异与概念模式和内模式之间的差异类似。例如,字段可能有不同的数据类型,字段和记录名可能改变,几个概念字段能合成一个单一的字段等。又如,可能同时存在多个外部视图,多个用户共享一个特定的外部视图,不同的外部视图可能存在交叉等。

三级模式与二级映射功能如图1-2所示。

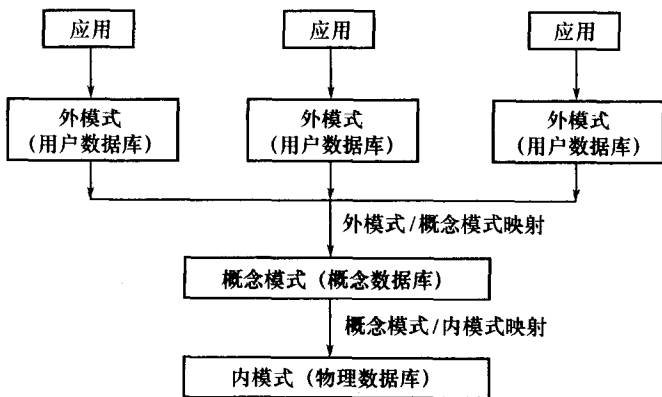


图 1-2 三级模式与二级映射功能

(1) 三级模式结构实现了数据使用的非过程化

数据库的三级模式结构是对数据的3个抽象级别。它把数据的具体组织留给DBMS,用户只需“抽象”处理数据,不必关心数据在机器中的存储和表示,从而实现了数据使用的非过

程化,大大减轻了各种用户使用数据库的负担。

(2) 二级功能映射实现三级模式间的转换

可以通过二级功能映射建立三级模式之间的联系与转换,使概念模式与外模式虽然并不“物理”存在,但也能通过映射而获得其功能意义上的“实体”。

(3) 二级功能映射保证系统的数据独立性

数据库系统有着许多优点,其中最重要的就是数据的独立性。数据独立性的实质是将数据模型与其实现分离开来。

如果用户和用户的应用程序对于数据库物理结构的改变保持不变,系统就提供了物理独立性,所以概念模式/内模式映射是物理独立性的关键;如果用户和用户的应用程序对于数据库逻辑结构的改变(在概念或逻辑概念层的改变)能保持不变,系统就提供了逻辑独立性,所以,外模式/概念模式映射就是逻辑独立性的关键。

1.2 数据模型

数据模型(Data Model)是一种抽象模型。数据是现实世界中事物及其特征的抽象,数据模型则是数据特征的抽象。事物个体特征通过各个事物个体之间的相互关系表现出来,数据模型从本质上来说,是数据间相互联系或者说约束条件的描述。数据模型正是从一般抽象的层面上模拟和描述了数据库系统的静态特征、动态行为和约束条件。

数据模型的概念是数据库系统的核心概念和基础,数据是按计算机系统的观点对数据建模。因此,数据库设计的核心问题之一就是要有有一个“好”的数据模型。无论是市场上的哪一种数据库管理系统产品,都是基于某一类的数据模型的实现。根据数据模型的不同,DBMS 可以分为:网状型、层次型、关系型和面向对象型,目前主流产品大多为关系型和面向对象的关系型。

根据实际问题的需要和应用目的的不同,可以将数据模型分为三种类型。

1.2.1 概念数据模型

概念数据模型(Conceptual Data Model)也称为信息模型。概念数据模型的实质是面向用户的模型,它是用户所容易理解的现实世界特征的数据抽象,其基本特征是按用户观点对数据和信息进行建模,与具体 DBMS 无关。概念数据模型作为数据库设计员与用户之间进行交流的语言,服务于数据库设计的应用目的。

概念模型的作用与意义在于描述现实世界的概念化结构,使得数据库设计人员在设计的初始阶段能够摆脱计算机系统及 DBMS 的具体技术问题束缚,集中精力分析数据之间的联系,充分而有效地与企业人员进行交流和沟通,使得设计能真实反映企业的客观实

际情况。

概念数据模型经过转换就可以变为 DBMS 支持的逻辑数据模型,进而在 DBMS 中得以实现。最常用和最著名的概念模型是实体-联系(E-R)模型,简称 E-R 模型。其中的基本概念有:

(1) 实体(Entity):客观存在并可相互区别的事物称为实体。实体可以是具体的人、事、物,也可以是抽象的概念或联系。

(2) 属性(Attribute):实体所具有的某一特性称为属性。一个实体可以由若干个属性来刻画。

(3) 码(Key):惟一标识实体的属性集。

(4) 域(Domain):属性的取值范围称为该属性的域。

(5) 实体型(Entity Type):用实体名及其属性名集合来抽象和刻画同类实体,称为实体型。

(6) 实体集(Entity Set):同型实体的集合称为实体集。

(7) 联系(Relationship):现实世界中事物内部以及事物之间的联系在信息世界中反映为实体内部的联系和实体之间的联系。

一对一联系(1:1):如果对于实体集 A 中的每一个实体,实体集 B 中至多有一个实体与之联系,反之亦然,则称实体集 A 与实体集 B 具有一对一联系。记为 1:1。

一对多联系(1:n):如果对于实体集 A 中的每一个实体,实体集 B 中有 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中至多只有一个实体与之联系,则称实体集 A 与实体 B 有一对多联系。记为 1:n。

多对多联系($m:n$):如果对于实体集 A 中的每一个实体,实体集 B 中有 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中也有 m 个实体($m \geq 0$)与之联系,则称实体集 A 与实体 B 具有多对多联系。记为 $m:n$ 。

实体型之间的一对一、一对多、多对多联系不仅存在于两个实体型之间,也存在于两个以上的实体型之间。

同一个实体集内的各实体之间也可以存在一对一、一对多、多对多的关系。

E-R 模型的图示形式就称为 E-R 图。E-R 图提供了用图形表示实体型、属性和联系的方法:

实体型:用矩形表示,矩形框内写明实体名。

属性:用椭圆形表示,并用无向边将其与相应的实体连接起来。

联系:用菱形表示,菱形框内写明联系名,并用无向边分别与有关实体连接起来,同时在无向边旁标上联系的类型(1:1、1:n 或 $m:n$)。

需要注意的是,联系本身也是一种实体型,也可以有属性。如果一个联系具有属性,则这些属性也要用无向边与该联系连接起来。