

普通高等教育“十三五”规划教材 (软件工程专业)

数据库原理

SHU JU KU YUAN LI

Database

主 编 杨俊杰 张 玮
副主编 侯 睿 熊建芳 关 心 张志洁

- 内容清晰、脉络分明、可读性和操作性强
- 采用项目驱动、案例引导的编写模式
- 注重应用型人才的专业技能和工程师实用技术的培养



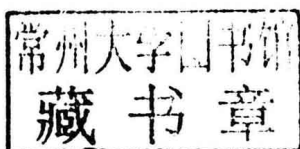
中国水利水电出版社
www.waterpub.com.cn

普通高等教育“十三五”规划教材（软件工程专业）

数据库原理

主 编 杨俊杰 张 玮

副主编 侯 睿 熊建芳 关 心 张志洁



中国水利水电出版社
www.waterpub.com.cn

· 北京 ·

内 容 提 要

本书全面介绍了数据库系统的基本原理及其实现技术。全书共 9 章, 内容主要包括数据库的基本概念、数据模型、关系数据库、SQL 语言、存储过程、触发器、关系数据库的规范化理论、数据库的安全性、完整性、事务管理、并发控制、SQL 查询优化与系统调优、数据库设计等。

本书除了在每章后均配有习题外, 还在第 3 章和第 4 章配有课堂练习, 为理实一体化教学提供参考素材。

书中所涉及例子均在 SQL Server 2010 环境下测试通过。

本书既可作为普通高等院校计算机及相关专业的数据库课程教材, 也可作为读者自学计算机技术的参考书。

图书在版编目 (C I P) 数据

数据库原理 / 杨俊杰, 张玮主编. — 北京 : 中国水利水电出版社, 2018. 1

普通高等教育“十三五”规划教材. 软件工程专业
ISBN 978-7-5170-6213-4

I. ①数… II. ①杨… ②张… III. ①数据库系统—高等学校—教材 IV. ①TP311. 43

中国版本图书馆CIP数据核字(2017)第326292号

策划编辑: 陈红华 责任编辑: 封 裕 加工编辑: 赵佳琦 封面设计: 李 佳

书 名	普通高等教育“十三五”规划教材(软件工程专业) 数据库原理 SHUJUKU YUANLI
作 者	主 编 杨俊杰 张 玮 副主编 侯 睿 熊建芳 关 心 张志洁
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 68367658 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	三河市铭浩彩色印装有限公司
规 格	184mm×260mm 16 开本 15.75 印张 384 千字
版 次	2018 年 1 月第 1 版 2018 年 1 月第 1 次印刷
印 数	0001—3000 册
定 价	35.00 元

凡购买我社图书, 如有缺页、倒页、脱页的, 本社营销中心负责调换
版权所有·侵权必究

编 委 会

主任委员 杨俊杰

副主任委员 吴 涛 赵法信 张立敏 曾绍庚 杨义文

罗 良 吴 东 张 玮

委 员 (排名不分先后)

洪伟铭 王晓晔 卢利琼 关 心 张志洁

张 强 莫长江 沈 旭 魏 武 蔡广基

孔艺权 李 君 石 艳 刘 瑶 梁 伟

侯 睿 陈 霞 刘劲武 邹海涛 刘宇欣

闵 笛 周迎春 张子石 梁 莉 彭增焰

阳 松 谭福超 陈茂东 宋联金

项目总策划 石永峰

序

为了深入贯彻落实教育部《关于加强高等学校本科教学工作，提高教学质量的若干意见》精神，紧密配合教育部《关于国家精品开放课程建设的实施意见》和广东省教育厅《广东省高等教育“创新强校工程”实施方案（试行）》，加快发展应用型普通院校的计算机专业本科教育，形成适应学科发展需求、校企深度融合的新型教育体系，在有关部门的大力支持下，我们组织并成立了“普通高等教育‘十三五’规划教材编审委员会”（以下简称“编委会”），讨论并实施应用型普通高等院校计算机类专业精品示范教材的编写与出版工作。编委会成员为来自教学科研一线的教师和软件企业的工程技术人员。

按照教育部的要求，编委会认为，精品示范教材应该能够反映应用型普通高等院校教学改革与课程建设的需要，教材的建设以提高学生的核心竞争力为导向，培养高素质的计算机高级应用人才。编委会结合社会经济发展的需求，设计并打造计算机科学与技术专业的系列教材。本系列教材涵盖软件技术、移动互联、软件与信息管理等专业方向，有利于建设开放共享的实践环境，有利于培养“双师型”教师团队，有利于学校创建共享型教学资源库。教材由个人申报，经编委会认真评审，由中国水利水电出版社审定出版。

本套规划教材的编写遵循以下几个基本原则：

（1）突出应用技术，全面针对实际应用。根据实际应用的需要组织教材内容，在保证学科体系完整的基础上，不过度强调理论的深度和难度，而是注重应用型人才专业技能和工程师实用技术的培养。

（2）教材采用项目驱动、案例引导的编写模式。以实际问题引导出相关原理和概念，在讲述实例的过程中将知识点融入，通过分析归纳，介绍解决工程实际问题的思想和方法，然后进行概括总结。教材内容清晰、脉络分明、可读性和可操作性强，同时，引入案例教学和启发式教学方法，便于激发学习兴趣。

（3）专家教师共建团队，优化编写队伍。由来自高校的一线教师、行业专家、企业工程师协同组成编写队伍，跨区域、跨学校交叉研究、协调推进，把握行业发展方向，将行业创新融入专业教学的课程设置和教材内容。

本套教材凝聚了众多长期在教学、科研一线工作的老师和数十位软件工程师的经验和智慧。衷心感谢该套教材的各位作者为教材出版所做的贡献。我们期待广大读者对本套教材提出宝贵意见，以便进一步修订，使该套教材不断完善。

丛书编委会

2017年12月

前 言

“数据库原理”是本科院校计算机相关专业的一门基础课。本书将数据库基本原理、方法和应用技术相结合,兼顾理论和应用,每个知识点都通过实例进行讲解,在 SQL 编程的相关章节提供一定的课堂练习,为理实一体化教学提供参考素材。本书能满足地方应用型本科院校人才培养的要求。

本书共分 9 章:

第 1 章 数据库系统概述。简要介绍了数据库的基本概念、数据模型、数据库体系结构等。

第 2 章 关系模型基本理论。简要介绍了关系模型的基本概念、传统的关系运算和专门的关系运算,并通过几个简单的例子说明关系运算的基本应用。

第 3 章 结构化查询语言 SQL。简要介绍了 SQL 语言、数据库对象、SQL Server 数据库的存储结构,详细讲解了 SQL 语言的数据定义、数据更新、数据查询语句的语法和应用,并给出了本书示例要用的一个示例数据库。在本章的部分小节,还给出了课堂练习。

第 4 章 T-SQL 编程。详细讲解了 T-SQL 程序设计、函数、存储过程、触发器、游标和 SQL 异常处理的语法和应用。本章的小节后附有课堂练习。

第 5 章 关系数据库的规范化理论。介绍了函数依赖、关系模式的规范化、关系模式分解的概念和基本应用。

第 6 章 数据库的安全性与完整性。主要介绍了数据库完整性约束的分类、完整性约束的定义、完整性约束的验证以及 SQL Server 中的完整性约束机制。

第 7 章 事务与并发控制。介绍了事务和并发控制的基本概念,讲解了事务控制的基本语法,并通过实例分析事务的处理过程。

第 8 章 SQL 查询优化与系统调优。介绍了关系数据库查询处理的步骤,讲解了 SQL 查询处理优化方法和计算机硬件调优策略。

第 9 章 数据库设计。介绍了数据库设计各阶段所采用的方式方法及处理手段。

本书由杨俊杰、张玮任主编,侯睿、熊建芳、关心、张志洁任副主编。在编写过程中,编者参考并引用了相关教材的部分内容,还有部分网络资料,限于篇幅和来源,无法面面俱到地罗列,在此一并对这些资料的作者致以衷心的感谢。

本书的出版得到了广东省计算机科学与技术专业综合改革试点项目(粤教高函(2013)113号)、广东省计算机实验教学示范中心项目(粤教高函(2015)133号)、广东高校优秀青年教师培养计划项目(编号:YQ2014117)等经费的资助。

由于作者水平所限,书中难免存在不妥之处,敬请广大读者批评指正,并欢迎读者将意见通过邮箱 yangjunjie1998@lingnan.edu.cn 反馈给编者。

编 者

2018 年 1 月

目 录

序

前言

第1章 数据库系统概述	1	3.3 示例数据库	32
1.1 数据库系统概述	1	3.4 SQL Server 数据库的存储结构	34
1.1.1 数据库的基本概念	1	3.4.1 逻辑存储结构	34
1.1.2 数据库技术的发展概述	3	3.4.2 物理存储结构	35
1.1.3 数据库系统的特点	5	3.5 数据定义	37
1.1.4 数据库系统的应用	6	3.5.1 数据库的创建和管理	37
1.2 数据模型	7	3.5.2 表的创建和管理	46
1.2.1 数据模型的组成三要素	7	3.5.3 表的完整性管理	50
1.2.2 数据抽象的过程	8	3.5.4 索引的创建和管理	53
1.2.3 概念模型	9	课堂练习	55
1.2.4 逻辑模型	11	3.6 数据更新	55
1.3 数据库体系结构	13	3.6.1 插入数据	56
1.3.1 数据库系统三级模式结构	13	3.6.2 更新数据	57
1.3.2 数据库系统的二级映射与 数据独立性	14	3.6.3 删除数据	58
习题1	14	课堂练习	59
第2章 关系模型基本理论	16	3.7 数据查询	59
2.1 关系模型	16	3.7.1 基本查询	59
2.1.1 关系数据库的基本概念	16	3.7.2 连接查询	67
2.1.2 关系的完整性	18	3.7.3 嵌套查询	70
2.2 关系代数	19	3.7.4 集合查询	75
2.2.1 传统的关系运算	20	3.7.5 SELECT 各子句的编写顺序和 执行顺序	78
2.2.2 专门的关系运算	21	课堂练习	79
2.2.3 关系代数操作实例	23	3.8 视图	79
习题2	24	习题3	82
第3章 结构化查询语言 SQL	26	第4章 T-SQL 编程	85
3.1 SQL 语言介绍	26	4.1 T-SQL 编程基础	85
3.1.1 SQL 的产生与发展	26	4.1.1 运算符与表达式	85
3.1.2 SQL 的特点	27	4.1.2 语句块和注释	91
3.1.3 SQL 的语句结构	27	4.1.3 流程控制语句	91
3.1.4 T-SQL	28	课堂练习	96
3.2 数据库对象	30	4.2 函数的使用	97

4.2.1 内置函数	97	5.4.1 规范化的含义	143
4.2.2 用户定义函数	101	5.4.2 第一范式	143
课堂练习	104	5.4.3 第二范式	145
4.3 存储过程	104	5.4.4 第三范式	146
4.3.1 存储过程的特点和类型	104	5.4.5 BCNF 范式	146
4.3.2 存储过程的创建和执行	105	5.5 多值依赖与 4NF	147
4.3.3 存储过程的参数和执行状态	108	5.5.1 问题的引入	147
4.3.4 存储过程的管理	111	5.5.2 多值依赖的基本概念	149
课堂练习	112	5.5.3 第四范式	150
4.4 触发器	113	5.6 关系模式分解	151
4.4.1 触发器简介	113	5.6.1 无损分解	151
4.4.2 触发器分类	113	5.6.2 保持函数依赖	154
4.4.3 创建触发器	115	5.7 连接依赖与 5NF	156
4.4.4 管理触发器	118	5.7.1 连接依赖	156
课堂练习	119	5.7.2 第五范式——5NF	158
4.5 游标	120	5.8 关系模式规范化的步骤	158
4.5.1 游标简介	120	习题五	159
4.5.2 游标的操作	121	第 6 章 数据库的安全性与完整性	161
4.5.3 游标应用举例	127	6.1 问题的提出	161
课堂练习	129	6.2 数据库的安全性	162
4.6 异常处理	129	6.2.1 数据库安全性问题的概述	162
4.6.1 异常捕获与异常抛出	129	6.2.2 数据库的安全性机制	163
4.6.2 异常处理	131	6.2.3 SQL Server 的安全性策略	167
习题 4	133	6.3 数据库的完整性	174
第 5 章 关系数据库的规范化理论	134	6.3.1 数据库完整性的概述	174
5.1 关系模式的设计问题	134	6.3.2 数据库完整性的分类	175
5.1.1 关系模式可能存在的异常	134	6.3.3 数据库完整性的定义与验证	177
5.1.2 关系模式中存在异常的原因	135	6.3.4 SQL Server 的完整性策略	178
5.1.3 关系模式规范化	136	习题 6	178
5.2 函数依赖	137	第 7 章 事务与并发控制	179
5.2.1 关系模式的简化表示	137	7.1 事务概述	179
5.2.2 函数依赖的基本概念	137	7.1.1 事务的特性	179
5.2.3 码的函数依赖表示	139	7.1.2 事务的类型	180
5.2.4 函数依赖和码的唯一性	139	7.2 事务的控制	181
5.3 函数依赖的公理系统	140	7.2.1 启动事务	182
5.3.1 属性的闭包与 F 逻辑蕴含的 充要条件	140	7.2.2 终止事务	183
5.3.2 最小函数依赖集 F_{min}	141	7.2.3 事务控制语句的使用	184
5.4 关系模式的规范化	142	7.2.4 事务和批的差别	185
		7.3 事务处理实例分析	186

7.4 并发控制	189	9.1.2 数据库设计的方法	214
7.4.1 并发控制概述	189	9.1.3 数据库设计的过程	215
7.4.2 封锁协议	190	9.2 需求分析	215
7.4.3 活锁和死锁	193	9.2.1 需求分析的任务	215
7.4.4 并发调度的可串行性	193	9.2.2 需求分析的方法	216
7.4.5 两段锁协议	194	9.2.3 用户需求调查的方法	216
7.4.6 基于时标的并发控制	196	9.2.4 数据流图	216
习题 7	197	9.2.5 数据字典	217
第 8 章 SQL 查询优化与系统调优	198	9.3 概念结构设计	218
8.1 概述	198	9.3.1 数据模型	218
8.2 关系数据库查询处理	198	9.3.2 概念模型	219
8.2.1 查询处理步骤	198	9.3.3 概念结构设计的方法与步骤	221
8.2.2 实现查询操作的算法示例	200	9.4 逻辑结构设计	232
8.3 SQL 查询处理优化方法	202	9.4.1 E-R 模型向关系模型的转换	232
8.3.1 基于索引的优化	202	9.4.2 关系模型的优化	235
8.3.2 SQL 语句优化	203	9.4.3 设计用户子模式	236
8.4 其他优化方法	207	9.5 物理设计	236
8.5 计算机硬件调优	211	9.5.1 确定数据库的物理结构	237
8.5.1 数据库对象的放置策略	211	9.5.2 评价物理结构	238
8.5.2 使用磁盘硬件优化数据库	211	9.6 数据库的实施与维护	238
习题 8	212	9.6.1 数据库的建立与调整	239
第 9 章 数据库设计	214	9.6.2 数据库系统的试运行	239
9.1 数据库设计概述	214	9.6.3 数据库系统的运行和维护	240
9.1.1 数据库设计的特点	214	习题 9	241

第 1 章 数据库系统概述

- 了解：数据与数据处理的概念、数据库技术的产生背景与发展概况。
- 理解：数据库系统的组成与特点、数据独立性的概念、数据模型的概念。
- 掌握：关系模型的基本知识、关系数据库的设计方法。

1.1 数据库系统概述

1.1.1 数据库的基本概念

1. 数据

数据（Data）是用来记录信息的可识别的符号，是信息的具体表现形式。数据是对现实世界的事物采用计算机能够识别、存储和处理的方式进行的描述，我们可以从中得到需要的信息。目前，数据既包含字母、文字及其他特殊字符，又包含图形、图像和声音等多媒体符号。

通常人们用很多事实来描述所感兴趣或要保存的事物。例如某所大学的校长要了解教师张雅清的一些基本情况，她的教师编号是 04018，每月基本工资为 5400 元，祖籍为广东湛江，出生日期是 1988 年 3 月 20 日，电话是 13623456789，家庭住址是湛江市赤坎区寸金路 29 号，等等。我们知道这些事实就可以每月处理张雅清老师的工资单，在生日时发贺卡给她，打印她的工资条，万一遇到紧急情况可以通知她的家人，等等。

2. 数据库

数据库（Database，DB）是长久存储在计算机中有组织、可共享、具有逻辑关系和确定意义的数据的集合。通俗一点来说，它是一个电子文件库，库里包含被计算机数据化了的文件。比如，你可以把单位同事的姓名、地址、电话号码等信息存储在计算机中的 Excel 表格中，这就是一个简单的数据库例子。

3. 数据库管理系统

数据库管理系统（Database Management System，DBMS）是一种重要的程序设计系统，它由一个相互关联的数据集合和一组访问这些数据的程序组成。这些数据集合就是数据库，并且允许用户根据需求增加、更改、删除以及检索数据。

通俗一点来说，数据库管理系统是一个通用的管理数据库的软件系统，是由一组计算机程序构成的。数据库管理系统负责数据库的定义、建立、操纵、管理和维护，能够对数据库进行有效的管理，包括存储管理、安全性管理、完整性管理等。

4. 数据库系统

数据库系统（Database System，DBS）是一个计算机应用系统，它是把计算机硬件、软件以及数据和有关人员组合起来为用户提供信息服务的系统。因此，数据库系统是由计算机系统、

数据库及其描述机制、数据库管理系统和有关人员组成的具有高度组织性的整体。数据库系统如图 1.1 所示。

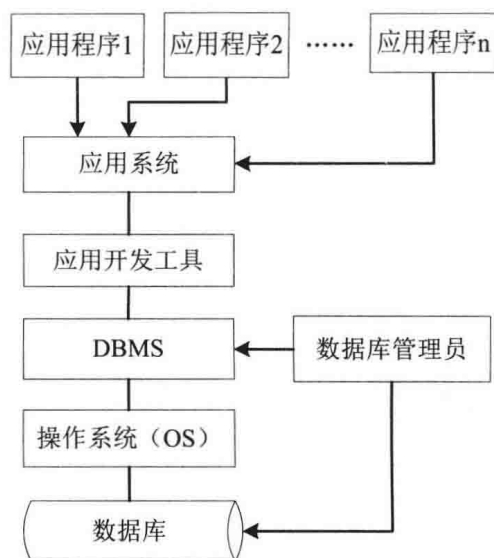


图 1.1 数据库系统

（1）计算机硬件。

计算机硬件是数据库系统的物质基础，是存储数据库及运行数据库管理系统的硬件资源，主要包括计算机主机、存储设备、输入/输出设备及计算机网络环境。

（2）计算机软件。

计算机软件包括操作系统、数据库管理系统、数据库应用系统等。

数据库管理系统是数据库系统的核心软件之一，它提供数据定义、数据操纵、数据库管理、数据库建立和维护及通信等功能。数据库管理系统提供对数据库中的数据资源进行统一管理和控制的功能，将用户、应用程序与数据库数据相互隔离，是数据库系统的核心，其功能的强弱是衡量数据库系统性能优劣的主要指标。数据库管理系统必须运行在相应的系统平台上，有操作系统和相关系统软件的支持。

数据库管理系统功能的强弱随系统而异，大系统功能较强、较全，小系统功能较弱、较少。目前较流行的数据库管理系统有 MySQL、SQL Server、Oracle 和 Sybase 等。

数据库应用系统是系统开发人员利用数据库系统资源开发出来的、面向某一类实际应用的应用软件系统。从实现技术角度而言，它是以数据库技术为基础的计算机应用系统。

（3）数据库。

数据库是数据库系统中按照一定的方式组织的、存储在外部存储设备上的、能被多个用户共享的、与应用程序相互独立的相关数据集合。它不仅包括描述事物的数据本身，而且包括相关事物之间的联系。

数据库中的数据往往不像文件系统那样只面向某一项特定应用，而是面向多种应用，可以被多个用户、多个应用程序共享。其数据结构独立于使用数据的程序，对数据的增加、删除、更改和检索都由数据库管理系统进行统一管理和控制，用户对数据库进行的各种操作也都是由数据库管理系统实现的。

(4) 数据库系统的有关人员。

数据库系统的有关人员主要有 3 类：最终用户、数据库应用系统开发人员和数据库管理员 (Database Administrator, DBA)。最终用户指通过应用系统的用户界面使用数据库的人员，他们一般对数据库知识了解不多。数据库应用系统开发人员包括系统分析员、系统设计员和程序员。系统分析员负责应用系统的分析，他们和最终用户、数据库管理员相配合，参与系统分析；系统设计员负责应用系统设计和数据库设计；程序员则根据设计要求进行编码。数据库管理员是数据管理机构的一组人员，他们负责对整个数据库系统进行总体控制和维护，以保证数据库系统的正常运行。

1.1.2 数据库技术的发展概述

1. 数据管理的诞生

在数据库系统出现以前，各个应用都拥有自己的专用数据，通常存放在专用文件中，这些数据与其他文件中的数据大量重复，造成了资源与人力的浪费。随着机器内存储数据的日益增多，数据重复的问题越来越突出。于是人们就想到将数据集中存储、统一管理，这样就演变成数据库管理系统进而形成数据库技术。数据库系统的萌芽出现于 20 世纪 60 年代，当时计算机开始广泛地应用于数据管理。人们对数据的共享提出了越来越高的要求，而传统的文件系统已经不能满足人们的需要，能够统一管理和共享数据的数据库管理系统应运而生。那时的数据管理非常简单，通过大量的分类、比较和表格绘制的机器运行数百万穿孔卡片来管理数据，其运行结果在纸上打印出来或者制成新的穿孔卡片，而数据管理就是对所有这些穿孔卡片进行物理的存储和处理。数据模型是数据库系统的核心和基础，各种数据库管理系统软件都是基于某种数据模型开发出来的。因此，通常也按照数据模型的特点将传统数据库系统分成网状数据库、层次数据库和关系数据库三类。

2. 关系数据库的由来

网状数据库和层次数据库已经很好地解决了数据的集中和共享问题，但是在数据独立性和抽象级别上仍有很大的欠缺。用户在对这两种数据库进行存取时，仍然需要明确数据的存储结构，指出存取路径。而后来出现的关系数据库较好地解决了这些问题。1970 年，IBM 研究员 E.F.Codd 博士提出了关系模型的概念，奠定了关系模型的理论基础。后来 Codd 又论述范式理论和衡量关系系统的 12 条标准，用数学理论奠定了关系数据库的基础。关系模型有严格的数据基础，抽象级别比较高而且简单清晰，便于理解和使用。1976 年，霍尼韦尔公司 (Honeywell) 开发了第一个商用关系数据库系统——Multics Relational Data Store。关系数据库系统以关系代数作为坚实的理论基础，经过几十年的发展和实际应用，技术越来越成熟和完善。其代表产品有 Oracle、IBM 公司的 DB2 和 Informix、微软公司的 MS SQL Server、ADABAS D 等。

3. 面向对象数据库系统

面向对象数据库系统 (Object-oriented Database System, OODBS) 是将面向对象的模型、方法和机制与先进的数据库技术有机结合而形成的新型数据库系统。它从关系模型中脱离出来，强调在数据库框架中发展类型、数据抽象、继承和持久性。它的基本设计思想：把面向对象语言向数据库方向扩展，使应用程序能够存取并处理对象；扩展数据库系统，使其具有面向对象的特征，提供一种综合的语义数据建模概念集，以便对现实世界中复杂应用的实体和联系建模。因此，面向对象数据库系统首先是一个数据库系统，具备数据库系统的基本功能；其次

是一个面向对象的系统,针对面向对象的程序设计语言的永久性对象存储管理而设计,充分支持完整的面向对象概念和机制。面向对象数据库系统对一些特定应用领域(如 CAD 等),能较好地满足其应用需求。

然而,数年的发展表明面向对象的关系型数据库系统产品的市场发展情况并不理想,理论上的完美并没有带来市场的热烈反应。不成功的主要原因在于,这种数据库产品的主要设计思想是企图用新型数据库系统来取代现有的数据库系统。这对许多已经运用数据库系统并积累了大量工作数据的客户,尤其是大客户来说,无法承受新旧数据间的转换而带来的巨大工作量及巨额开支。另外,面向对象的关系型数据库系统使查询语言变得极为复杂,从而使得无论是数据库的开发商家还是应用客户都视其复杂的应用技术为畏途。

4. 数据库技术的现状

1980 年以前,数据库技术的发展主要体现在数据库的模型设计上。进入 20 世纪 90 年代后,计算机领域中其他新兴技术的发展对数据库技术产生了重大影响。数据库技术与网络通信技术、人工智能技术、多媒体技术等相互渗透、相互结合,使数据库技术的新内容层出不穷。数据库的许多概念、应用领域,甚至某些原理都有了重大的发展和变化,形成了数据库领域众多的研究分支和课题,产生了一系列新型数据库。分析目前数据库的应用情况可以发现:经过多年的积累,企业和部门积累的数据越来越多,许多企业面临着“数据爆炸”而知识缺乏的困境。如何解决海量数据的存储管理、挖掘大量数据中包含的信息和知识,已成为目前亟待解决的问题。所以,除了数据库技术核心问题的研究外,市场的需求导致了以下几种数据库的发展及一些研究热点:分布式数据库、并行数据库、主动数据库、多媒体数据库、模糊数据库、知识数据库、XML 数据库、数据仓库和联机分析处理(OLAP)、数据挖掘、面向对象数据库及数据可视化技术。

5. 数据库技术发展的趋势

(1) 非结构化数据库。

非结构化数据库是部分研究者针对关系数据库模型过于简单、不便表达复杂的嵌套需要以及支持数据类型有限等局限性,从数据模型入手而提出的全面基于因特网应用的新型数据库理论。这种数据库的最大区别就在于它突破了关系数据库结构定义不易改变和数据定长的限制,支持重复字段、子字段以及变长字段,实现了对变长数据和重复字段进行处理和数据项的变长存储管理,在处理连续信息(包括全文信息)和非结构信息(重复数据和变长数据)中有着传统关系数据库所无法比拟的优势。但研究者认为此种数据库技术并不会完全取代现在流行的关系数据库,而是它们的有益补充。

(2) 数据库技术与多学科技术的有机结合。

有学者指出,数据库与学科技术的结合将会建立一系列新数据库,如分布式数据库、并行数据库、知识库、多媒体数据库等,这将是数据库技术重要的发展方向。其中,许多研究者都把多媒体数据库作为研究的重点,并认为将多媒体技术和可视化技术引入多媒体数据库是未来数据库技术发展的热点和难点。

(3) 数据仓库和电子商务。

随着信息技术的高速发展,数据库应用的规模、范围和深度不断扩大,一般的事务处理已不能满足应用的需要,企业需要在大量数据的基础上进行决策,而数据仓库(Data Warehouse, DW)技术的兴起满足了这一需求。数据仓库作为决策支持系统(Decision Support

System, DSS) 的有效解决方案, 涉及三个方面的技术内容: 数据仓库技术、联机分析处理 (Online Analytical Processing, OLAP) 技术和数据挖掘 (Data Mining, DM) 技术。

数据仓库、OLAP 和数据挖掘是作为三种独立的数据处理技术出现的。数据仓库用于数据的存储和组织, OLAP 集中于数据的分析, 数据挖掘则致力于知识的自动发现。它们都可以分别应用到信息系统的设计和实现中, 以提高相应部分的处理能力。但是, 由于这三种技术内在的联系性和互补性, 将它们结合起来就是一种新的 DSS 架构。这一架构以数据库中的大量数据为基础, 系统则由数据驱动。数据仓库技术应用遍及通信、零售业、金融及制造业等领域。

(4) 面向专门应用领域的数据库技术。

许多研究者从实践的角度对数据库技术进行研究, 提出了适合应用领域的数据库技术, 如工程数据库、统计数据库、科学数据库、空间数据库、地理数据库等。这类数据库在原理上也没有多大的变化, 但是它们却与一定的应用相结合, 从而加强了系统对有关应用的支撑能力, 尤其表现在数据模型、语言、查询方面。部分研究者认为, 随着研究工作的继续深入和数据库技术在工作中的应用, 数据库技术将会朝着更多专门应用领域发展。

(5) 内存数据库系统。

内存数据库 (Main Memory Database, MMDB) 系统是实时系统和数据库系统的有机结合。它抛弃了磁盘数据管理的传统方式, 基于全部数据都在内存中这一前提重新设计了体系结构, 并且在数据缓存、快速算法、并行操作方面进行了相应的改进, 所以数据处理速度比传统数据库的数据处理速度要快很多, 一般都在 10 倍以上。内存数据库的最大特点是其“主拷贝”或“工作版本”常驻内存, 即活动事务只与实时内存数据库的内存拷贝打交道。内存数据库系统目前广泛应用于航空、军事、电信、电力及工业控制等领域。

数据库系统的功能从早期的数据存储、查询到联机事务处理, 再到数据挖掘, 从单纯的数据库发展到与之相关的模型库、知识库的集成, 其所取得的成就是令人瞩目的。当然, 所有的这些都有许多局限性, 还有许多关键问题等待解决, 而且, 随着应用领域日益广泛, 硬件技术的不断提高, 数据库技术还要面临新的挑战。当前数据库技术的发展呈现出与多种学科知识相结合的趋势, 凡是有数据 (广义的) 产生的领域就可能需要数据库技术的支持, 它们相结合后就会出现一种新的数据库成员而壮大数据库家族。此外, 虽然我国目前对数据库技术的研究和应用都还处于较低的水平, 但是随着计算机的普及和社会信息化, 数据库技术将处于越来越重要的地位, 在未来的信息社会中, 数据库技术必将得到更大、更快的发展。

1.1.3 数据库系统的特点

数据库系统的出现是计算机数据管理技术的重大进步, 它克服了文件系统的缺陷, 提供了对数据更高级、更有效的管理。

1. 数据结构化

在文件系统中, 文件的记录内部是有结构的。例如, 学生数据文件的每个记录是由学号、姓名、性别、出生年月、籍贯、简历等数据项组成的。但这种结构只适用于特定的应用, 对其他应用并不适用。

在数据库系统中, 每一个数据库都是为某一应用领域服务的。例如, 学校信息管理涉及多个方面的应用, 包括对学生的学籍管理、课程管理、成绩管理等, 还包括教工的人事管理、教学管理、科研管理、住房管理和工资管理等, 这些应用彼此之间都有着密切的联系。因此,

在数据库系统中不仅要考虑某个应用的数据结构,还要考虑整个组织(即多个应用)的数据结构。这种数据组织方式使数据结构化了,这就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。而在文件系统中,尽管其记录内部已有了某些结构,但记录之间没有联系。数据库系统实现了整体数据的结构化,这既是数据库的主要特点之一,也是数据库系统与文件系统的本质区别。

2. 数据共享性高、冗余度低

数据共享是指多个用户或应用程序可以访问同一个数据库中的数据,而 DBMS 提供并发和协调机制,可以保证在多个应用程序同时访问、存取和操作数据库数据时不产生任何冲突,从而保证数据不遭到破坏。

数据冗余既浪费存储空间,又容易产生数据的不一致。在文件系统中,由于每个应用程序都有自己的数据文件,所以数据存在着大量的重复。

数据库从全局观念来组织和存储数据,数据已经根据特定的数据模型结构化了,在数据库中用户的逻辑数据文件和具体的物理数据文件不必一一对应,从而有效地节省了存储资源,减少了数据冗余,保证了数据的一致性。

3. 具有较高的数据独立性

数据独立性是指应用程序与数据库的数据结构之间相互独立。在数据库系统中,因为采用了数据库的三级模式结构,保证了数据库中数据的独立性;在数据存储结构改变时,不影响数据的全局逻辑结构,保证了数据的物理独立性;在全局逻辑结构改变时,不影响用户的局部逻辑结构和应用程序,保证了数据的逻辑独立性。

4. 有统一的数据控制功能

在数据库系统中,数据由 DBMS 进行统一控制和管理。DBMS 提供了一套有效的数据控制手段,包括数据安全性控制、数据完整性控制、数据库的并发控制和数据库的恢复等,增强了多用户环境下数据的安全性和一致性保护。

1.1.4 数据库系统的应用

数据库技术应用在很多方面,从应用类型方面来分,有信息系统、事务处理系统、管理信息系统、决策支持系统和数据挖掘系统等。

1. 信息系统 (Information System)

信息系统是一个由人员、活动、数据、网络和技术等要素组成的集成系统,其目的是对组织的业务数据进行采集、存储、处理和交换,以支持和改善组织的日常业务运作,满足管理人员解决问题和制定决策对信息的各种需求。比如医院管理子系统通常包含门诊、住院两部分,而管理的主线则为药品和收款金额。因为其数据量巨大、实时性强,所以在数据库系统选型时必须选择高效、稳定的大型数据库系统。

2. 事务处理系统 (Transaction Processing System)

事务处理系统是利用计算机对工商业、社会服务性行业等中的具体业务进行处理的信息系统。基于计算机的事务处理系统又称为电子数据处理系统,它以计算机、网络为基础,对业务数据进行采集、存储、检索、加工和传输。

3. 管理信息系统 (Management Information System)

管理信息系统是对一个组织机构进行全面管理的以计算机为基础的集成化的人—机系

统,具有分析、计划、预测、控制和决策功能。它把数据处理功能与管理模型的优化计算、仿真等功能结合起来,能准确、及时地向各级管理人员提供决策用的信息。

4. 决策支持系统 (Decision Support System)

决策支持系统是计算机科学、行为科学和系统科学(包括控制论、系统论、信息论、运筹学、管理科学等)相结合的产物,是以支持半结构化和非结构化决策过程为特征的一类计算机辅助决策系统,用于支持高级管理人员进行战略规划和宏观决策。

5. 数据挖掘

数据挖掘又称为数据库中的知识发现(Knowledge Discovery from Database, KDD),它是一个从大量数据中挖掘出未知、有价值的模式或规律等知识的复杂过程。数据挖掘的全过程如下:

- 数据清洗(Data Cleaning),其作用就是清除数据噪音和与挖掘主题明显无关的数据。
- 数据集成(Data Integration),其作用就是将来自多数据源中的相关数据组合到一起。
- 数据转换(Data Transformation),其作用就是将数据转换为易于数据挖掘的数据存储形式。
- 数据挖掘(Data Mining),它是知识挖掘的一个基本步骤,其作用就是利用智能方法挖掘数据模式或规律知识。
- 模式评估(Pattern Evaluation),其作用就是根据一定评估标准(Interesting Measures)从挖掘结果筛选出有意义的模式知识。
- 知识表示(Knowledge Presentation),其作用就是利用可视化和知识表达技术,向用户展示所挖掘出的相关知识。

现如今的数据库积累越来越多,在那些数据库中总有很多数据是我们不需要的,为了对其进行更深层次的分析,以便更好地利用这些数据,我们就利用挖掘技术重新根据需求开发系统,这个系统就是数据挖掘系统。

1.2 数据模型

数据模型是现实世界数据特征的抽象,用于描述一组数据的概念和定义。数据模型是数据库中数据的存储方式,是数据库系统的基础。在数据库中,数据的物理结构即数据的存储结构,是数据元素在计算机存储器中的表示及其配置;数据的逻辑结构则是数据元素之间的逻辑关系,它是数据在用户或程序员面前的表现形式,数据的存储结构不一定与逻辑结构一致。因此,了解和掌握数据模型的基本概念是学习数据库的基础。

1.2.1 数据模型的组成三要素

由于数据模型是现实世界的事物及其联系的一种模拟和抽象表示,是一种形式化描述数据、数据间联系以及有关语义约束规则的方法,这些规则规定数据如何组织以及允许进行何种操作。因此,数据模型所描述的要素有三个部分,分别是数据结构、数据操作和数据约束。

1. 数据结构

数据结构用于描述系统的静态特征,包括数据的类型、内容、性质及数据之间的联系等。它是数据模型的基础,也是刻画一个数据模型性质最重要的方面。例如前面介绍的教师

(04018, 张雅清, 5400, 广东湛江, 1988-3-20, 13623456789, 湛江市赤坎区寸金路 29 号) 属于记录型数据结构, 即教师 (教师编号, 姓名, 基本工资, 籍贯, 出生日期, 电话号码, 居住地址)。

因此, 在数据库系统中, 通常按照数据结构的类型来命名数据模型, 如层次结构、网状结构和关系结构的数据模型分别命名为层次模型、网状模型和关系模型。

2. 数据操作

数据操作用于描述系统的动态特征, 包括数据的插入、修改、删除和查询等。数据模型必须定义这些操作的确切含义、操作符号、操作规则及实现操作的语言。

3. 数据约束

数据的约束条件实际上是一组完整性规则的集合。完整性规则是给定数据模型中的数据及其联系所具有的制约和存储规则, 用以限定符合数据模型的数据库及其状态的变化, 以保证数据的正确性、有效性和相容性。

数据模型应该反映和规定数据必须遵守的、基本的、通用的完整性约束。此外, 数据模型还应该提供定义完整性约束条件的机制, 以反映具体涉及的数据必须遵守的、特定的语义约束条件, 如学生信息中的“性别”只能为“男”或“女”, 学生选课信息中的“课程号”的值必须为学校已开设课程的课程号等。

1.2.2 数据抽象的过程

从现实世界中的客观事物到数据库中存储的数据是一个逐步抽象的过程, 这个过程经历了现实世界、观念世界和机器世界三个阶段, 对应于数据抽象的不同阶段采用不同的数据模型。首先将现实世界的事物及其联系抽象成观念世界的概念模型, 然后再转换成机器世界的数据库模型。概念模型并不依赖于具体的计算机系统, 它不是 DBMS 所支持的数据模型, 而是现实世界中客观事物的抽象表示。概念模型经过转换成为计算机上某一 DBMS 支持的数据模型。因此, 数据模型是对现实世界进行抽象和转换的结果, 这一过程如图 1.2 所示。

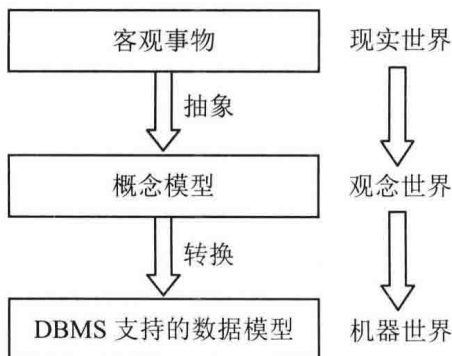


图 1.2 数据抽象的过程

1. 对现实世界的抽象

现实世界就是客观存在的世界, 其中存在着各种客观事物及其相互之间的联系, 而且每个事物都有自己的特征或性质。计算机处理的对象是现实世界中的客观事物, 在对其实施处理