

21世纪高等学校规划教材 | 计算机科学与技术



# 数据库原理及应用

周 炜 著



清华大学出版社

21世纪高等学校规划教材 | 计算机科学与技术

# 数据库原理及应用

周 炜 著

清华大学出版社  
北京

## 内 容 简 介

本书是作者多年教学经验和研究成果的结晶,系统地研究和介绍了数据库系统的基本原理、方法和发展现状,用总篇幅的三分之二重点研究了关系数据库的基本理论、规范化理论和 SQL 语言。全书分为 8 章:数据库系统概论,DBS 需求分析和概念设计,关系数据库基本理论,关系规范化理论和 DBS 逻辑设计,关系数据库结构化查询语言 SQL,DBS 物理设计和实现、运行与维护,DBMS 的事务管理和安全性控制,数据库系统新技术简介。在本书最后提供了 11 个教学实验,供参考。

本书内容丰富,体系完整,论证严密,行文流畅,深入浅出,特色鲜明。

本书可以作为计算机科学与技术和其他相关专业本科生、大专生的教材,也可作为其他有关专业师生和工程技术人员的参考书和自学用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

## 图书在版编目(CIP)数据

数据库原理及应用/周炜著.—北京:清华大学出版社,2011.9

(21 世纪高等学校规划教材·计算机科学与技术)

ISBN 978-7-302-26153-7

I. ①数… II. ①周… 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(2011)第 122624 号

责任编辑:郑寅堃

责任校对:李建庄

责任印制:王秀菊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954,jsjic@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015,zhiliang@tup.tsinghua.edu.cn

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185×260 印 张:14 字 数:349 千字

版 次:2011 年 9 月第 1 版 印 次:2011 年 9 月第 1 次印刷

印 数:1~3000

定 价:23.00 元

产品编号:043107-01

# 编审委员会成员

(按地区排序)

清华大学

周立柱 教授  
覃 征 教授  
王建民 教授  
冯建华 教授  
刘 强 副教授

北京大学

杨冬青 教授  
陈 钟 教授  
陈立军 副教授

北京航空航天大学

马殿富 教授  
吴超英 副教授  
姚淑珍 教授

中国人民大学

王 珊 教授  
孟小峰 教授  
陈 红 教授

北京师范大学

周明全 教授

北京交通大学

阮秋琦 教授  
赵 宏 副教授

北京信息工程学院

孟庆昌 教授

北京科技大学

杨炳儒 教授

石油大学

陈 明 教授

天津大学

艾德才 教授

复旦大学

吴立德 教授

吴百锋 教授

杨卫东 副教授

同济大学

苗夺谦 教授

徐 安 教授

华东理工大学

邵志清 教授

华东师范大学

杨宗源 教授

应吉康 教授

东华大学

乐嘉锦 教授

孙 莉 副教授

浙江大学	吴朝晖	教授
	李善平	教授
扬州大学	李 云	教授
南京大学	骆 斌	教授
	黄 强	副教授
南京航空航天大学	黄志球	教授
	秦小麟	教授
南京理工大学	张功萱	教授
南京邮电学院	朱秀昌	教授
苏州大学	王宜怀	教授
	陈建明	副教授
江苏大学	鲍可进	教授
中国矿业大学	张 艳	教授
武汉大学	何炎祥	教授
华中科技大学	刘乐善	教授
中南财经政法大学	刘腾红	教授
华中师范大学	叶俊民	教授
	郑世珏	教授
	陈 利	教授
江汉大学	颜 彬	教授
国防科技大学	赵克佳	教授
	邹北骥	教授
中南大学	刘卫国	教授
湖南大学	林亚平	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永锋	教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
仰恩大学	张思民	教授
云南大学	刘惟一	教授
电子科技大学	刘乃琦	教授
	罗 蕾	教授
成都理工大学	蔡 淮	教授
	于 春	副教授
西南交通大学	曾华燊	教授

# 出版说明

---

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

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,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上。精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展,顺应并符合21世纪教学发展的规律,代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版

社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。推出的特色精品教材包括:

(1) 21世纪高等学校规划教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 21世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 21世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。

(4) 21世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。

(5) 21世纪高等学校规划教材·信息管理与信息系统。

(6) 21世纪高等学校规划教材·财经管理与应用。

(7) 21世纪高等学校规划教材·电子商务。

(8) 21世纪高等学校规划教材·物联网。

清华大学出版社经过三十年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail:weijj@tup.tsinghua.edu.cn

# 前言

数据库系统尤其是关系数据库系统的诞生是数据管理技术乃至软件技术发展历史上的一场革命,它从根本上改变了数据管理技术的面貌,影响了软件技术发展的历史进程和现状。如今,数据库技术已经成为软件技术的一个分支,成为信息技术的核心和基础,数据库系统原理也已经成为我国计算机科学与技术、信息管理与信息系统等专业本科生、大专生的一门必修课。

本书内容是根据作者多年的教学经验和研究成果整理的。书中知识点不超出本科生的认知水平和学习范围;很多内容属于作者的独立研究成果,是第一次面世,并有作者自己研制的一批实用算法;改进了现有的一些算法;对几乎所有的定理和算法都给出了正确性证明(本书用一个黑色小方块(■)作为理论证明的结束标志);引入了一些新的概念,如核心属性、边缘属性、引用属性、正则函数依赖等,同时对现有的一些概念如数据库、数据库系统、数据库管理系统等给出了有自己特色的定义;每章都有一批风格独特的例题和习题;有关编程的所有例题都经过计算机实际运行证明正确。

本书以数据库系统设计过程为主线安排章节顺序,系统地研究和介绍了数据库系统的基本原理、方法、应用和发展现状,用总篇幅的三分之二重点研究了关系数据库的基本理论、规范化理论和 SQL 语言。在 SQL 语言部分增加了局部变量和流程控制、用户自定义函数等内容。全书分为 8 章:第 1 章为数据库系统概论;第 2 章讨论和介绍 DBS 需求分析与概念结构设计;第 3 章研究关系数据库基本理论;第 4 章研究关系数据库规范化理论和 DBS 逻辑结构设计;第 5 章研究关系数据库结构化查询语言 SQL;第 6 章介绍 DBS 的物理设计和实现、运行与维护;第 7 章介绍 DBMS 的事务管理和安全性控制;第 8 章介绍数据库系统新技术。每章后面配有一定数量难度不一的习题可供选做,书后还附有实验教学参考计划。

本书可以作为计算机科学与技术和其他相关专业本科生、大专生的教材,也可作为其他有关专业师生和工程技术人员的参考书或自学用书。打 \* 号的内容供选学。

若以本书作为教材,可在教学中暂时跳过一些复杂的理论证明。学时安排建议:本科生 40~50 学时,大专生 50~60 学时;上机实验 10~20 学时。

本书绝大部分内容已在空军工程大学导弹学院计算机科学与技术专业本科生和大专生中讲授多年。但由于作者水平有限,书中一定还有作者未发现的错误、缺点和纰漏,恳请广大读者批评指正,作者不胜感激!

作 者

2011 年 7 月





# 目 录

<b>第 1 章 数据库系统概论</b> .....	1
1.1 信息、数据和数据库.....	1
1.1.1 信息、数据和数据管理 .....	1
1.1.2 数据管理技术及其发展.....	1
1.2 数据抽象 .....	4
1.2.1 数据模型.....	4
1.2.2 概念模型.....	5
1.2.3 逻辑模型及其要素.....	5
1.2.4 外部模型.....	7
1.2.5 内部模型.....	7
1.3 数据库管理系统(DBMS) .....	7
1.3.1 DBMS 的基本概念 .....	7
1.3.2 DBMS 的组成和各部分功能 .....	8
1.3.3 DBMS 的工作过程 .....	9
1.3.4 DBMS 的主要功能 .....	9
1.4 数据库系统(DBS).....	10
1.4.1 DBS 的概念 .....	10
1.4.2 DBS 的组成 .....	10
1.4.3 DBS 的三级模式结构 .....	12
1.4.4 DBS 的数据独立性 .....	14
1.4.5 DBS 的全局结构 .....	14
1.4.6 DBS 的体系结构分类 .....	15
1.5 数据库系统设计.....	16
1.5.1 DBS 设计的基本任务 .....	17
1.5.2 DBS 设计的特点 .....	17
1.5.3 DBS 的设计过程 .....	18
1.6 小结.....	20
1.7 习题.....	20
<b>第 2 章 DBS 需求分析和概念设计</b> .....	22
2.1 需求分析.....	22
2.1.1 需求分析的任务和方法 .....	22

2.1.2	数据流图 .....	23
2.1.3	数据字典 .....	26
2.2	概念结构设计 .....	28
2.2.1	概念结构设计任务和方法 .....	28
2.2.2	概念结构设计的步骤 .....	28
2.3	E-R 图设计 .....	29
2.3.1	E-R 模型中的数据描述 .....	29
2.3.2	E-R 图的图元 .....	33
2.3.3	E-R 图的设计原则 .....	34
2.3.4	局部 E-R 图之间的三种冲突 .....	34
2.3.5	E-R 图的设计步骤 .....	35
2.4	小结 .....	37
2.5	习题 .....	38
<b>第 3 章</b>	<b>关系数据库基本理论 .....</b>	<b>40</b>
3.1	集合论的基本概念 .....	40
3.1.1	集合的基本概念 .....	40
3.1.2	集合的代数运算及性质 .....	41
3.1.3	集合的运算性质 .....	42
3.2	关系数据库的基本概念 .....	42
3.2.1	集合上的关系与关系数据模型 .....	42
3.2.2	关系模式、关系实例与关系数据库 .....	43
3.2.3	关系数据库模型中的数据完整性约束 .....	46
3.2.4	关系运算的分类 .....	47
3.3	关系代数 .....	47
3.3.1	关系代数的基本运算 .....	47
3.3.2	关系代数的组合运算 .....	50
3.3.3	关系代数的扩展运算 .....	53
3.3.3.1	关系的外连接和半连接 .....	53
3.3.3.2	关系的改名、赋值、外部并、广义投影和聚集运算 .....	54
3.3.4	关系代数的安全性 .....	54
3.3.5	关系代数表达式的优化 .....	55
3.3.5.1	关系代数表达式的优化问题 .....	55
3.3.5.2	关系代数表达式的等价变换规则 .....	55
3.3.5.3	关系代数表达式的优化策略 .....	57
3.4	关系演算 .....	58
3.4.1	元组关系演算 .....	58
3.4.2	域关系演算* .....	58
3.4.3	安全关系演算与关系代数的等价性 .....	59

3.5	关系逻辑*	59
3.5.1	关系逻辑的要素	59
3.5.2	关系逻辑规则的安全性	60
3.5.3	从关系代数到关系逻辑的转换	60
3.5.4	递归过程	61
3.6	小结	62
3.7	习题	62
<b>第4章</b>	<b>关系规范化理论和 DBS 逻辑设计</b>	<b>64</b>
4.1	函数依赖	64
4.1.1	函数依赖的定义	64
4.1.2	Armstrong 公理系统与函数依赖推理规则	65
4.1.3	函数依赖集的正则闭包	67
4.1.4	属性集关于函数依赖集的闭包	71
4.1.5	部分函数依赖和传递函数依赖	73
4.1.6	键码	74
4.1.7	极小函数依赖集与正则覆盖	80
4.2	多值依赖	85
4.2.1	多值依赖的定义与基本性质	85
4.2.2	多值依赖推理规则	87
4.2.3	多值依赖与函数依赖的主要区别和共同点	89
4.3	关系模式的规范化	90
4.3.1	数据冗余和操作异常	90
4.3.2	消除不良数据依赖的主要途径——关系模式分解	92
4.3.3	对关系模式分解的要求	93
4.3.3.1	保持函数依赖的分解	93
4.3.3.2	无损连接分解	94
4.3.3.3	对关系模式分解的要求	102
4.3.4	关系模式的范式	102
4.4	第一范式	103
4.4.1	第一范式的定义	103
4.4.2	第一范式的缺点	104
4.5	基于消除不良函数依赖的范式	104
4.5.1	第二范式	104
4.5.2	第三范式	106
4.5.2.1	第三范式的定义和性质	106
4.5.2.2	第三范式的函数依赖特性	106
4.5.2.3	第三范式的缺点	107
4.5.3	BC 范式	108

4.5.3.1	BC 范式的定义	108
4.5.3.2	BC 范式的函数依赖特性	108
4.5.3.3	BC 范式的缺点	109
4.5.4	关于传递函数依赖定义的讨论	109
4.6	第三范式和 BC 范式的有关算法	110
4.6.1	第三范式的有关算法	110
4.6.1.1	3NF 的判定算法	110
4.6.1.2	关系模式分解成 3NF 的算法	111
4.6.2	关系模式分解成 BCNF 的算法	112
4.7	基于消除不良多值依赖的范式——第四范式(4NF)	113
4.7.1	第四范式的概念	113
4.7.2	关系模式分解成 4NF 的算法	113
4.8	基于消除不良连接依赖的范式——第五范式(5NF)	114
4.9	各范式间的关系	114
4.10	数据库的逻辑结构设计	115
4.10.1	逻辑结构设计中的数据描述	115
4.10.2	逻辑结构设计任务	115
4.10.3	E-R 图向关系数据库模型的转换	115
4.10.4	关系数据库模型的优化	116
4.10.5	关系数据库模型的外模式设计	117
4.11	小结	118
4.12	习题	118
<b>第 5 章</b>	<b>关系数据库结构化查询语言 SQL</b>	<b>121</b>
5.1	SQL 概述	121
5.1.1	SQL 的产生与发展	121
5.1.2	SQL 的组成和特点	121
5.1.3	SQL 的数据类型、运算符、表达式、标识符、通配符和函数	123
5.1.3.1	SQL 的数据类型	123
5.1.3.2	SQL 的运算符和表达式	124
5.1.3.3	SQL 的标识符和通配符	125
5.1.3.4	SQL 的常用函数	126
5.1.4	SQL 实现的数据完整性约束	128
5.2	SQL 的局部变量和流程控制	130
5.2.1	SQL 的局部变量、BEGIN...END 语句块和 PRINT 语句	130
5.2.2	SQL 的 IF...ELSE 语句	130
5.2.3	SQL 的 CASE 语句	131
5.2.4	SQL 的 WHILE 循环语句	131
5.2.5	SQL 的 GOTO 语句和 RETURN 语句	131

5.3	SQL 的数据定义——CREATE、ALTER、DROP 语句	132
5.3.1	CREATE 语句	132
5.3.1.1	创建空数据库和基本表	132
5.3.1.2	创建索引	134
5.3.2	ALTER 语句	135
5.3.3	DROP 语句	136
5.4	SQL 的数据查询——SELECT 语句	137
5.4.1	SELECT 语句的一般形式和执行过程	137
5.4.2	单表查询	140
5.4.3	连接查询	142
5.4.4	联合查询——实现并运算的查询	144
5.4.5	嵌套查询——相关子查询和不相关子查询	145
5.4.6	将查询结果直接组织成新基本表	147
5.5	SQL 的数据修改——INSERT、DELETE、UPDATE 语句	147
5.5.1	INSERT 语句	148
5.5.2	DELETE 语句	149
5.5.3	UPDATE 语句	149
5.6	SQL 的视图	149
5.6.1	视图的创建和查询	149
5.6.2	修改视图定义	151
5.6.3	修改视图数据	151
5.7	嵌入式 SQL	151
5.7.1	ESQL 的使用规定	152
5.7.2	ESQL/C 的使用方法	154
5.8	用户自定义函数	157
5.8.1	标量函数	157
5.8.2	内嵌表值函数	158
5.8.3	多语句表值函数*	159
5.9	用户自定义存储过程	160
5.9.1	随机调用的存储过程	160
5.9.2	触发器*	163
5.10	小结	164
5.11	习题	165
<b>第 6 章</b>	<b>DBS 物理设计和实现、运行与维护</b>	<b>166</b>
6.1	数据库的存储结构	166
6.1.1	数据库文件的组织	166
6.1.2	数据库文件的结构	167
6.1.3	数据库访问技术	168

6.2	DBS 的物理结构设计 .....	169
6.3	DBS 的实现 .....	169
6.4	DBS 的运行与维护 .....	170
6.5	小结 .....	172
6.6	习题 .....	172
<b>第 7 章 DBMS 的事务管理和安全性控制 .....</b>		<b>173</b>
7.1	事务 .....	173
7.1.1	事务的概念 .....	173
7.1.2	事务的 ACID 性质 .....	174
7.2	DBMS 对事务的并发控制 .....	175
7.2.1	事务并发执行可能带来的问题 .....	175
7.2.2	封锁技术 .....	176
7.2.3	活锁与死锁 .....	176
7.2.4	两段锁协议 .....	176
7.3	DBS 的安全性 .....	177
7.3.1	安全性级别 .....	177
7.3.2	权限控制 .....	177
7.3.3	计算机病毒、木马和流氓软件的防护 .....	179
7.4	数据库的恢复技术 .....	179
7.4.1	DBS 故障分类和恢复策略 .....	179
7.4.2	检查点 .....	180
7.4.3	数据库镜像 .....	180
7.5	小结 .....	181
7.6	习题 .....	181
<b>第 8 章 数据库系统新技术简介 .....</b>		<b>182</b>
8.1	数据库访问接口技术 .....	182
8.1.1	ODBC .....	182
8.1.2	OLE DB .....	186
8.1.3	ADO .....	186
8.1.4	JDBC .....	190
8.1.5	ADO.NET .....	190
8.2	对象数据库系统 .....	191
8.3	并行数据库系统 .....	192
8.4	分布式数据库系统 .....	193
8.4.1	分布式数据库系统的定义、特点和分类 .....	193
8.4.2	分布式数据存储 .....	193
8.4.3	DDBS 的分布透明性 .....	194

8.4.4 DDBMS 的功能和组成 .....	195
8.5 现代信息集成技术 .....	195
8.5.1 数据仓库技术 .....	196
8.5.2 联机分析处理技术 .....	196
8.5.3 数据挖掘技术 .....	197
8.6 XML 技术 .....	197
8.7 小结 .....	198
8.8 习题 .....	198
<b>实验教学参考计划</b> .....	199
实验 1 SQL 的数据定义语句 .....	199
实验 2 SELECT 语句单表查询 .....	200
实验 3 SELECT 语句连接查询和联合查询 .....	201
实验 4 SELECT 语句嵌套查询 .....	201
实验 5 INSERT 语句、DELETE 语句和 UPDATE 语句 .....	202
实验 6 SQL 的视图 .....	203
实验 7 嵌入式 SQL .....	203
实验 8 SQL 的自定义函数 .....	205
实验 9 SQL 的用户自定义存储过程 .....	206
实验 10 ODBC 应用程序 .....	206
实验 11 ASP 网页 .....	207
<b>参考文献</b> .....	208

# 数据库系统概论

数据库技术诞生于 20 世纪 60 年代末,迄今已有 50 年多的历史。现在,数据库技术已经成为软件学科的一个重要分支,数据库原理也成为国内计算机等信息类专业本科生、大专生的一门必修课。

本章介绍数据库系统的基本概念,旨在使读者对数据库系统有一个基本的了解。

## 1.1 信息、数据和数据库

### 1.1.1 信息、数据和数据管理

**定义 1.1** 信息是事物表达其运动状态与方式的一种属性,这种属性是事物运动、变化、联系、差异的产物,是事物为了与外部环境进行协调以减少运动、变化、联系、差异所产生的不确定性而进行的表达。

在这里,“事物”泛指一切可能的研究对象,包括客观世界的物质,也包括主观世界的精神现象;“运动”泛指一切意义上的变化,包括机械运动、化学运动、思维运动和社会运动;“运动方式”是指事物的运动在时间上所呈现的过程和规律;“运动状态”则是事物的运动在空间上所展示的形状与态势。

**定义 1.2** 数据是人们对事物所含信息的符号表示。数据具有概括性、结构性和独立性。

为了记载信息,人们使用了各种各样的符号和它们的组合,这些符号及其组合就是数据。数据可以是数值数据,也可以是非数值数据,如声音、图像等。在计算机中,数据是能输入计算机并能为其处理的符号序列。

数据是信息的符号表示,信息是数据的内涵,是数据的语义解释。数据是符号化的信息,信息是语义化的数据。数据是信息的具体表示形式,信息是数据的有意义的表现。

### 1.1.2 数据管理技术及其发展

**定义 1.3** 数据管理是对数据的分类、组织、编码、存储、检索和维护。

数据管理技术的发展先后经过了人工管理、文件系统管理、数据库系统管理 3 个阶段。

#### 1. 人工管理阶段

人工管理阶段是指 20 世纪 50 年代中期以前。这一阶段的主要背景是:



- (1) 计算机主要用于科学计算,数据量小、结构简单。如高阶方程、曲线拟合等。
- (2) 外存为顺序存取设备(如磁带、卡片、纸带),没有磁盘等直接存取设备。
- (3) 没有操作系统,没有数据管理软件。用户用机器指令编码,通过纸带机输入程序和数据,程序运行完毕后,用户取走纸带和运算结果,再让下一用户上机操作。

人工管理阶段的主要特点是:

- (1) 没有专门的软件对数据进行管理,用户完全负责数据管理工作。数据的组织、存储结构、存取方法、输入、输出等,均由用户自己设计和完成。
- (2) 数据完全面向特定的应用程序。每个用户使用自己的数据,数据不能共享;而且数据用完就撤走,不能长期保存。
- (3) 程序中存取数据的子程序随着数据存储结构的改变而改变,程序与数据没有独立性。

## 2. 文件系统管理阶段

文件系统管理阶段是指 20 世纪 50 年代后期至 60 年代中期这一时期。这一阶段的主要背景是:

- (1) 计算机不但用于科学计算,还用于信息管理。
- (2) 外存有了磁盘、磁鼓等直接存取设备,无须顺序存取,由地址可直接访问所需记录。
- (3) 有了操作系统(OS),OS 中有专门管理数据的子系统——文件系统。

文件系统管理阶段的主要特点是:

- (1) 数据可以以文件形式长期保存在外部存储器上。
- (2) 数据的存取基本上以记录为单位。
- (3) 系统提供一定的数据管理功能。这主要表现为文件组织和数据存取方法多样化,有索引文件、链接文件、直接存取文件、倒排文件等;支持对文件的增、删、改、查等基本操作,应用程序不必考虑其物理细节。
- (4) 一个数据文件对应一个或几个应用程序,数据不一定属于某个特定的应用程序,可以重复使用,但数据与应用程序之间的依赖关系并未改变,数据仍是面向应用程序的。
- (5) 文件的逻辑结构与存储结构由系统进行转换,因而数据的逻辑结构与存储结构有了区别,数据存储结构的改变不一定反映在应用程序上,应用程序与数据有一定的独立性(设备独立性)。

文件系统管理的主要缺点是:

- (1) 数据与程序的独立性差。文件系统只是解脱了程序员对物理设备的存取负担,但它并不理解数据的语义,只负责存储。数据的语义信息只能由程序来解释,也就是说,数据收集以后怎么组织,以及数据取出来之后按什么含义使用,只有全权管理它的程序知道。因此,一个应用程序若想共享另一个应用程序生成的数据,必须与该应用程序沟通,了解数据的语义与组织方式。可见,文件系统的出现并没有从根本上改变数据与程序紧密结合的状况,数据逻辑结构的改变必然导致应用程序的修改。

- (2) 数据联系弱,共享性差,冗余度大。由于数据面向应用程序,即使不同应用程序所需要的数据只有小部分不同,也必须建立各自的数据文件,而不能共享相同的数据。由于数据文件之间缺乏联系,相互孤立,导致数据分散管理,使得同样的数据在多个数据文件中出