

高等学校计算机专业规划教材

# 数据库原理与应用

## ——基于SQL Server 2014



蒙祖强 许嘉 编著

清华大学出版社

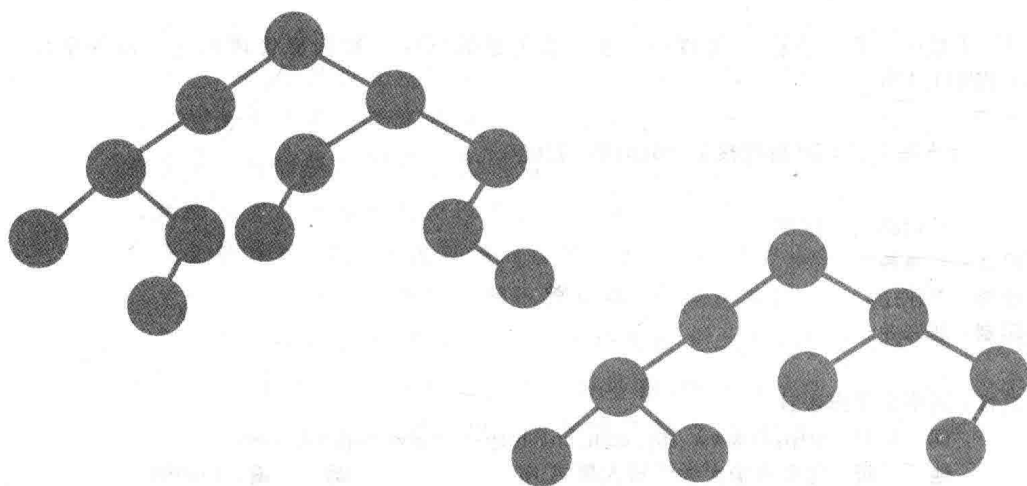


高等学校计算机专业规划教材

# 数据库原理与应用

## ——基于SQL Server 2014

蒙祖强 许嘉 编著



清华大学出版社  
北京

## 内 容 简 介

本书首先介绍关系数据库的理论基础知识和数据库的设计技术和方法,然后结合 SQL Server 2014 详细介绍了基于关系数据库基础知识的数据库开发和应用技术。全书共 13 章,内容包括数据库概述、关系数据库理论基础、数据库设计技术、SQL Server 2014 简介与安装、数据库查询语言 SQL、Transact-SQL 程序设计、数据库的创建和管理、索引与视图、存储过程和触发器、事务管理与并发控制、数据的完整性管理、数据的安全性控制、数据库备份和恢复等。

本书的特点是技术性和实践性强,内容通俗易懂,同时又兼顾应有的理论基础知识,理论知识和实践知识按照适当比例有机结合。本书实例翔实、逻辑性强、结构清晰、条理清楚、重点突出,每章后面都配有适量的习题(包括上机题)以及习题答案,供教师和学生参考使用。

本书主要面向应用型和应用研究型本科院校,可作为计算机及相关专业的数据库课程教材,也可作为数据库爱好者和初学者的学习用书,还可以作为从事数据库系统开发人员的参考用书。

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

数据库原理与应用——基于 SQL Server 2014/蒙祖强,许嘉编著. —北京:清华大学出版社,2018  
(高等学校计算机专业规划教材)

ISBN 978-7-302-49392-1

I. ①数… II. ①蒙… ②许… III. ①关系数据库—数据库管理系统—高等学校—教材  
IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2018)第 012096 号

责任编辑:龙启铭 常建丽

封面设计:何凤霞

责任校对:李建庄

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:22.25

字 数:517千字

版 次:2018年3月第1版

印 次:2018年3月第1次印刷

印 数:1~1500

定 价:49.00元

产品编号:038371-01



数据库技术是计算机技术的重要组成部分,也是发展最快、应用最广的计算机技术之一,自20世纪60年代中后期出现以来,经历了50多年的发展,目前日臻完善,造就了诸多的辉煌,带动了一系列的软件产业,出现了Oracle、DB2、SQL Server等十分成熟且深受用户喜爱的数据库产品。如今不管是在工作,还是在生活和学习中,数据库已经成为各类信息系统和应用系统的技术基础,与人们的生活息息相关。随着信息技术的进一步发展,数据库技术将发挥更重要的基础作用。

“数据库原理”课程是计算机科学与技术和相关专业的主干课程。这门课程的主要特点是实践性强,同时又要求具备较好的理论基础。根据多年的教学体会,我们发现有的学生学完这门课后,仍然难以胜任数据库设计、创建、开发、维护和管理的基本工作,觉得所学的数据库知识比较混乱。出现这种现象的原因可能是多方面的,但我们认为其主要原因之一就是学生类型与教材的搭配问题。高校培养的人才类型大致可以分为研究型人才、应用型人才和应用研究型人才。应该说,后二者占的比例比较高,他们希望能掌握技术性、实践性比较强的数据库知识,以便为他们毕业后的工作提供技术和方法支持。如果使用理论性很强、内容比较广泛的教材,他们内心深处可能产生一种抵触情绪而导致学习积极性欠佳,进而导致学习效果差。因此,针对技术型、应用研究型人才的培养,编写合适的、与时俱进的学习教材,以开发学生的实践性思维能力和概括能力,是教学工作者面临的一项重要任务。

基于以上考虑,我们组织人手,在总结多年教学经验的基础上编写了这部数据库教材。为了避免内容的分散,本教材阐述的数据库知识主要依托SQL Server 2014,其中涉及的SQL代码都是在SQL Server 2014中调试完成的(实际上,绝大部分SQL代码也适用于其他版本的SQL Server)。全书共13章,第1章介绍数据库系统涉及的基本概念;第2章介绍关系数据库理论基础;第3章介绍数据库设计技术;第4章介绍SQL Server 2014简介与安装;第5章比较系统地介绍了数据库查询语言SQL,以及数据表的创建方法;第6章介绍Transact-SQL程序设计;第7章介绍数据库的创建和管理;第8章介绍索引与视图的创建、管理和使用方法;第9章介绍存储过程和触发器的开发和使用方法;第10章介绍事务管理与并发控制的技术和方法;第11章和第12章分别介绍如何实现和保证数据的完整性和安全性;第13章

比较全面地介绍数据库的备份方法及其恢复技术。

本书的特点是由浅入深、通俗易懂、技术性和实践性强,同时又兼顾应有的理论基础知识,理论知识和实践知识按照适当比例有机结合。本书实例翔实、逻辑性强、结构清晰、条理清楚、重点突出,此外,每章后面都配有适量的习题(包括上机题)并在书末附有习题答案,供教师和学生参考使用。

本书主要面向应用型和应用研究型本科院校,可作为计算机及相关专业的数据库课程教材,也可作为数据库爱好者和初学者的学习用书,还可以作为从事数据库系统开发人员的参考用书。本书中所有的实例代码以及教学用的教学大纲和 PPT 课件都可以从清华大学出版社网站(<http://www.tup.com.cn/>)上免费下载。读者如有问题或需要技术支持,可与编辑联系,也可以直接与作者联系 [mengzuqiang@163.com](mailto:mengzuqiang@163.com)。

全书由蒙祖强教授执笔,许嘉副教授修订了第 1 章和第 2 章并对全书进行审阅。此外,参与本书编写、资料整理和调试程序的还有秦亮曦、刘智斌、黄柏雄、顾平、姚怡、李虹利、郭英明等。

感谢所有关心、支持本书编写和出版的人员,包括广西大学李陶深教授、陈宁江教授以及其他一些老师、研究生和技术人员,同时感谢清华大学出版社的领导和编辑,他们为本书的编写和出版提供了很大帮助。本书还参考了相关文献和网络资源,在此对这些资料的著者们表示衷心感谢。

蒙祖强

2018 年 1 月



## 第 1 章 数据库概述 /1

1.1	数据管理技术	1
1.1.1	数据管理的概念	1
1.1.2	数据管理技术的发展过程	2
1.2	大数据分析与管理技术	5
1.2.1	大数据	5
1.2.2	大数据存储技术	6
1.2.3	大数据处理模式	9
1.2.4	大数据处理的基本流程	11
1.3	数据库系统概述	12
1.3.1	数据库系统	12
1.3.2	数据库系统的模式结构	13
1.3.3	数据库管理系统	15
1.4	数据模型	16
1.4.1	数据模型的基本要素	16
1.4.2	4 种主要的逻辑模型	17
1.5	概念模型的描述	20
1.5.1	概念模型	20
1.5.2	实体及其联系	21
1.5.3	E-R 图	22
	习题 1	24

## 第 2 章 关系数据库理论基础 /26

2.1	关系模型	26
2.1.1	关系模型的数据结构——关系	26
2.1.2	关系操作	29
2.1.3	关系的完整性约束	29
2.2	关系代数	32
2.2.1	基本集合运算	32
2.2.2	关系运算	33



2.3	关系数据库	36
2.3.1	关系数据库的概念	36
2.3.2	关系数据库的特点	37
2.4	函数依赖	37
2.4.1	函数依赖的概念	37
2.4.2	候选码和主码	39
2.4.3	函数依赖的性质	40
2.5	关系模式的范式	40
2.5.1	第一范式(1NF)	41
2.5.2	第二范式(2NF)	42
2.5.3	第三范式(3NF)	44
2.5.4	BC 范式(BCNF)	46
2.6	关系模式的分解和规范化	48
2.6.1	关系模式的规范化	48
2.6.2	关系模式的分解	49
	习题 2	51
<b>第 3 章 数据库设计技术</b>		<b>/55</b>
3.1	数据库设计概述	55
3.2	需求分析	56
3.2.1	系统调研过程	56
3.2.2	需求分析的方法	57
3.2.3	形成数据字典	61
3.3	数据库结构设计	62
3.3.1	概念结构设计	62
3.3.2	逻辑结构设计	66
3.3.3	物理结构设计	68
3.4	数据库的实施、运行和维护	70
3.4.1	数据库的实施	70
3.4.2	数据库系统的运行和维护	72
	习题 3	73
<b>第 4 章 SQL Server 2014 简介与安装</b>		<b>/75</b>
4.1	SQL Server 的发展历史	75
4.2	SQL Server 2014 的组件和管理工具	77
4.2.1	SQL Server 2014 的组件	77
4.2.2	SQL Server 2014 的管理工具	78
4.3	SQL Server 2014 的几个版本	78



4.4	SQL Server 2014 的安装 .....	79
4.4.1	安装 SQL Server 2014 的要求 .....	79
4.4.2	SQL Server 2014 的安装过程 .....	81
4.4.3	SQL Server 2014 的使用方法 .....	88
	习题 4 .....	90
<b>第 5 章</b>	<b>数据库查询语言 SQL</b>	<b>/91</b>
5.1	SQL 概述 .....	91
5.1.1	SQL 的发展 .....	91
5.1.2	SQL 的特点 .....	92
5.1.3	SQL 的基本功能 .....	93
5.2	SQL 的数据类型 .....	93
5.2.1	字符串型 .....	94
5.2.2	数值型 .....	94
5.2.3	日期时间型 .....	95
5.2.4	货币型 .....	96
5.2.5	二进制型 .....	96
5.2.6	其他数据类型 .....	96
5.2.7	自定义数据类型 .....	97
5.3	SQL 的数据定义功能 .....	97
5.3.1	数据表的创建和删除 .....	97
5.3.2	数据表的修改 .....	101
5.4	SQL 的数据查询功能 .....	102
5.4.1	基本查询 .....	104
5.4.2	带 DISTINCT 的查询 .....	106
5.4.3	带 WHERE 子句的条件查询 .....	107
5.4.4	带 BETWEEN 的范围查询 .....	108
5.4.5	带 IN 的范围查询 .....	109
5.4.6	带 GROUP 的分组查询 .....	109
5.4.7	带 LIKE 的匹配查询和空值的查询 .....	111
5.4.8	使用 ORDER 排序查询结果 .....	112
5.4.9	连接查询 .....	113
5.4.10	嵌套查询 .....	119
5.4.11	查询的集合运算 .....	122
5.5	SQL 的数据操纵功能 .....	124
5.5.1	数据插入 .....	125
5.5.2	数据更新 .....	127
5.5.3	数据删除 .....	129





习题 5 .....	130
<b>第 6 章 Transact-SQL 程序设计 /135</b>	
6.1 Transact-SQL 概述 .....	135
6.1.1 Transact-SQL .....	135
6.1.2 Transact-SQL 元素 .....	136
6.2 Transact-SQL 的变量和常量 .....	136
6.2.1 变量的定义和使用 .....	137
6.2.2 Transact-SQL 常量 .....	140
6.3 Transact-SQL 运算符 .....	143
6.4 Transact-SQL 流程控制 .....	145
6.4.1 注释和语句块 .....	145
6.4.2 IF 语句 .....	146
6.4.3 CASE 语句 .....	149
6.4.4 WHILE 语句 .....	152
6.4.5 GOTO 语句 .....	153
6.4.6 TRY...CATCH 语句 .....	154
6.4.7 RETURN 语句 .....	154
6.4.8 WAITFOR 语句 .....	155
6.5 Transact-SQL 函数 .....	155
6.5.1 系统内置函数 .....	155
6.5.2 用户自定义函数 .....	162
习题 6 .....	167
<b>第 7 章 数据库的创建和管理 /169</b>	
7.1 数据库和数据库文件 .....	169
7.1.1 数据库的组成 .....	169
7.1.2 文件组 .....	170
7.2 数据库的创建 .....	171
7.2.1 创建数据库的 SQL 语法 .....	171
7.2.2 创建使用默认参数的数据库 .....	173
7.2.3 创建指定数据文件的数据库 .....	173
7.2.4 创建指定大小的数据库 .....	174
7.2.5 创建带多个数据文件的数据库 .....	175
7.2.6 创建指定文件组的数据库 .....	176
7.3 查看数据库 .....	177
7.3.1 服务器上的数据库 .....	177
7.3.2 数据库的基本信息 .....	178



7.3.3	数据库中的数据表	180
7.4	修改数据库	180
7.4.1	更改数据库的名称	180
7.4.2	修改数据库的大小	181
7.5	数据库的分离和附加	182
7.5.1	用户数据库的分离	182
7.5.2	用户数据库的附加	183
7.6	删除数据库	185
	习题 7	185
<b>第 8 章</b>	<b>索引与视图</b>	<b>/187</b>
8.1	索引概述	187
8.1.1	索引的概念	187
8.1.2	何种情况下创建索引	188
8.1.3	索引的原理——B_树	188
8.2	索引的类型	189
8.2.1	聚集索引和非聚集索引	189
8.2.2	唯一索引与非唯一索引	191
8.2.3	组合索引	191
8.3	创建索引	192
8.3.1	创建聚集索引	192
8.3.2	创建非聚集索引	194
8.3.3	创建唯一索引	194
8.3.4	创建组合索引	195
8.4	查看和删除索引	195
8.4.1	查看索引	195
8.4.2	删除索引	196
8.5	视图概述	197
8.5.1	视图的概念	197
8.5.2	视图的优缺点	198
8.6	视图的创建、更新与删除	199
8.6.1	创建视图	199
8.6.2	更新视图	202
8.6.3	删除视图	202
8.7	查看视图	203
8.7.1	视图的定义代码	203
8.7.2	视图的结构信息	203
8.7.3	数据库中的视图	203



习题 8 .....	204
<b>第 9 章 存储过程和触发器 /207</b>	
9.1 存储过程 .....	207
9.1.1 存储过程的概念 .....	207
9.1.2 存储过程的类型 .....	208
9.1.3 存储过程的创建和调用 .....	209
9.1.4 存储过程的修改和删除 .....	213
9.2 触发器 .....	214
9.2.1 关于触发器 .....	214
9.2.2 创建触发器 .....	216
9.2.3 修改触发器 .....	220
9.2.4 禁用和删除触发器 .....	221
习题 9 .....	223
<b>第 10 章 事务管理与并发控制 /226</b>	
10.1 事务的基本概念 .....	226
10.1.1 事务 .....	226
10.1.2 事务的 ACID 特性 .....	227
10.2 事务的管理 .....	227
10.2.1 启动事务 .....	228
10.2.2 终止事务 .....	229
10.2.3 嵌套事务 .....	234
10.3 并发控制 .....	236
10.3.1 并发控制的概念 .....	236
10.3.2 几种并发问题 .....	237
10.3.3 基于事务隔离级别的并发控制 .....	239
10.3.4 基于锁的并发控制 .....	245
习题 10 .....	251
<b>第 11 章 数据的完整性管理 /253</b>	
11.1 关于数据完整性 .....	253
11.1.1 数据完整性的概念 .....	253
11.1.2 数据完整性的分类 .....	253
11.2 实体完整性的实现 .....	254
11.2.1 实体完整性的定义 .....	254
11.2.2 实体完整性的检查 .....	257
11.3 参照完整性的实现 .....	257



11.3.1	参照完整性的定义	257
11.3.2	参照完整性的检查	258
11.4	用户定义完整性的实现	260
11.4.1	域完整性的实现	260
11.4.2	表级约束完整性的实现	264
习题 11		265

## 第 12 章 数据的安全性控制 /268

12.1	SQL Server 2014 安全体系结构	268
12.1.1	主体	268
12.1.2	安全对象	270
12.1.3	权限	271
12.2	角色	272
12.2.1	服务器角色	272
12.2.2	数据库角色	272
12.2.3	应用程序角色	273
12.3	服务器级的安全控制	274
12.3.1	身份验证模式	274
12.3.2	创建登录	275
12.3.3	查看登录	278
12.3.4	登录的权限管理	279
12.3.5	删除登录	283
12.4	数据库级的安全控制	283
12.4.1	数据库用户的管理	283
12.4.2	安全对象的权限管理	289
12.5	架构级的安全控制	296
12.5.1	架构及其管理	296
12.5.2	安全对象的权限管理	301
习题 12		303

## 第 13 章 数据库备份和恢复 /305

13.1	备份和恢复	305
13.1.1	备份和恢复的概念	305
13.1.2	恢复模式及其切换	305
13.1.3	备份类型	307
13.2	完整数据库备份与恢复	308
13.2.1	完整数据库备份	308
13.2.2	完整数据库恢复	309



13.3 差异数据库备份与恢复.....	310
13.3.1 差异数据库备份.....	310
13.3.2 差异数据库恢复.....	311
13.4 事务日志备份与恢复.....	312
13.4.1 事务日志备份.....	312
13.4.2 事务日志恢复.....	313
13.5 一种备份案例.....	314
习题 13 .....	318

**参考答案 /319**

**参考文献 /342**

本章介绍数据库系统的基本概念,包括数据管理技术的发展历程、大数据管理、数据库系统的概念及其模式结构、数据库管理系统的概念、数据的 4 种逻辑模型、概念模型的描述、E-R 图的概念等。通过本章的学习,读者应该了解或掌握下列内容:

- 了解数据管理技术发展的几个过程。
- 了解大数据的概念、存储技术、处理模式和处理流程等。
- 了解数据库系统的组成部分和数据库系统的模式结构。
- 了解数据库管理系统在数据库系统中的作用。
- 了解数据的 4 种逻辑模型。
- 掌握概念模型及 E-R 图的绘制。

### 1.1 数据管理技术

本节主要介绍数据管理的相关概念及其发展过程,以帮助读者更好地认识数据库技术出现的背景。

#### 1.1.1 数据管理的概念

走进数据库技术领域,首先遇到的概念是信息、数据、数据处理和数据管理等基本概念。这些概念和术语将贯穿读者对数据库技术学习和应用的全过程。正确理解这些概念,对读者学习和掌握数据库技术有重要的意义。

##### 1. 信息与数据

信息(information)是现实世界中客观事物的反映,这种反映主要体现为事物属性的表现形式,是对事物存在方式或运动状态的刻画。具体讲,信息是经过加工后的客观事物反映的数据表现形式,它可能对人类的行为产生影响,具有潜在或明显的实际应用价值。

信息的主要特征体现在:

- 可传递性。信息是可以传递的,但其传递的前提是必须有载体,且传递过程消耗能量。
- 可感知性。信息是可以被人类“感觉”到的,但感觉的方式可能由于信息源的不同而呈现多样性。例如,耳朵可以听到信息,眼睛可以看到信息。
- 可管理性。信息是可以被管理的,我们可以通过一定的方法对信息进行加工、存储、传播、再生和增值等。

信息作为一种资源,与物质和能量一样,在社会活动中产生了深刻而重要的影响。现在,信息已经变成一种产业,“信息社会”“信息经济”等词语相继产生,折射出人类对信息的一种依托。

数据(data)是描述事务的符号记录,是信息的符号化表示,是信息的载体。也就是说,数据是信息表示的一种符号形式。这种符号形式可以是语言、图表、数字、声音等。在数据库中,符号形式多表现为记录格式。但不管用什么样的符号形式,其目的只有一个,那就是客观地反映信息的内容。信息的内容不会随着数据表现形式的不同而改变。

在概念上,信息和数据既有区分,又有联系。数据是信息的载体,可以有多种表现形式,其目的都是为了揭示信息的内容;信息是数据的内涵,仅由客观事物的属性来确定,与数据形式无关。但在实际应用中,如果不需要特别强调信息和数据的差异,这两个概念往往是互换的。例如,“信息处理”和“数据处理”通常是有相同内涵的两个概念。

## 2. 数据处理与数据管理

数据处理也称信息处理,泛指用计算机对各种类型数据进行的处理操作,这些操作包括对数据进行采集、转换、分类、存储、排序、加工、维护、统计和传输等一系列活动。操作过程可能是重复而复杂的,最终可能会产生新的数据和信息。数据处理的目的是从原始数据中提取有价值的、可作决策依据的信息。

在复杂的数据处理过程中,有些操作是基本的,如数据存储、分类、统计和检索等,这些基本的数据操作过程通常称为数据管理。实际上,数据管理是数据处理的任务之一,是数据管理的核心内容。数据库系统的基本功能就是数据管理。

### 1.1.2 数据管理技术的发展过程

数据管理技术始于 20 世纪 50 年代中期以前。60 多年来,随着计算机技术的发展,数据管理技术已经发生了深刻的变化,其发展主要经历了人工管理、文件系统和数据库系统 3 个阶段。下面以这 3 个阶段为主线,介绍数据管理技术的发展过程及其特点。

#### 1. 人工管理阶段

这一阶段主要指从计算机诞生到 20 世纪 50 年代中期的这一个时期。世界上第一台计算机 ENIAC 于 1946 年 2 月 14 日在美国宾夕法尼亚大学诞生,在随后的近十年中,计算机的主要应用是科学计算,处理的是数字数据,数据量不大。计算机没有操作系统(实际上,当时根本没有操作系统的概念),也没有数据的管理软件,是以批处理方式对数据进行计算。计算机硬件本身也没有磁盘,所使用的“存储设备”是磁带、卡片等。

这个时期数据管理技术的特点体现在:

- 数据不保存。一方面是当时计算机处理的数据量很小,不需要保存;另一方面是计算机本身就没有有效的存储设备。
- 数据缺乏独立性和有效的组织方式。这体现在数据依赖于应用程序,缺乏共享性。其原因在于,数据的逻辑结构与程序是紧密联系在一起的,程序 A 处理的数据,对程序 B 而言可能就无法识别,更谈不上处理。解决的办法是修改数据的逻辑结构,或者修改应用程序。而且,数据的逻辑结构严重依赖于数据的物理结构,要更改数据的逻辑结构,就必须重新组织数据的存储结构、存取方法、输入/输出

方式等。显然,这种数据管理方法仅适用于小量数据,对大量数据则是低效的。

- 数据为程序所拥有,冗余度高。由于数据缺乏独立性,一组数据只能为一个程序所拥有,不能同时为多个程序所共享,这就造成了一份数据的多个副本,各程序之间存在大量重复的数据,从而产生大量的冗余数据。

## 2. 文件系统阶段

这一阶段是指从20世纪50年代后期到60年代中期的这段时间。这个时期,计算机除了用于科学计算以外,还大量用于数据的管理。计算机已经有了操作系统,并且在操作系统上已经开发了一种专门用于数据管理的软件——文件管理系统。在文件管理系统中,数据的批处理方法发展到文件的批处理方式,且还可以实现一定程度的联机实时处理。计算机硬件本身已经出现了磁盘、磁鼓等外部存储设备。

在文件系统阶段,数据管理技术的主要特点如下:

- 计算机的应用已从单纯的科学计算逐步转移到数据处理上来,特别是在该阶段的后期,数据处理已经成为计算机应用的主要目的。但这时候处理操作还比较简单,主要限于对文件的插入、删除、修改和查询等基本操作。
- 数据按照一定的逻辑结构组成文件,并通过文件实现数据的外部存储。即数据是以文件的方式存储在外部存储设备中,如磁盘、磁鼓等。
- 数据具有一定的独立性。由于数据是以文件的方式存储,文件的逻辑结构与存储结构可以自由地进行转换,所以多个程序可以通过文件系统对同一数据进行访问,实现了一定程度的数据共享。这样,程序员可以把更多的时间花在程序设计中,而不必把太多精力放在数据的存储设计上。
- 文件形式具有多样化。出于对文件管理的需要,除了数据文件以外,还产生了索引文件、链接文件、顺序文件、直接存取文件和倒排文件等。
- 基本上以记录为单位实现数据的存取。

相对人工管理阶段而言,上述特点是文件管理系统的优点。但这些优点是相对的,需要改进的地方也是明显的,这体现在:

- 数据和程序并不相互独立,数据冗余度仍然比较大。一个程序基本上对应着一个或一组文件,即数据还是面向应用的,不同的程序还需建立自己的数据文件,能真正实现数据共享的情况并不多。原因在于,文件中数据的逻辑结构与其对应的程序密切相关,而且没有统一标准的逻辑结构。这样,同一份数据可能以不同的方式在多个文件中重复出现,数据的冗余度依然比较大。
- 难以保证数据一致性。由于文件之间没有关联机制,所以当对一个数据进行更改时,难以保证对该数据的其他副本进行同样的更改。
- 文件的数据表达能力还十分有限。这主要体现在,文件中的数据结构比较单一,也比较简单,还难以表示复杂的数据结构。

## 3. 数据库系统阶段

数据库系统阶段始于20世纪60年代中后期,一直到现在。这时计算机除了用于科学计算外,更多的时候是用于数据管理,而且数据的量已经很大,管理功能也越来越强大。计算机硬件本身也发生了深刻的变化,出现了大容量磁盘和高主频的CPU等。在软件



上,数据的管理软件已经由原来的文件系统上升到数据库管理系统(Database Management System, DBMS)。不管是管理的功能,还是管理的效率,都有了长足的进步。这一时期,数据管理的主要特点是:数据集中存放在一个地方,这个地方就是所谓的数据库。应用程序要实现对数据库中的数据进行访问,则必须通过数据库管理系统来完成。这种基于数据库的数据管理技术就是所谓的数据库技术。关于数据库技术,很难用一言两语来概括,但我们可以从它的一些特点来了解这种技术。其特点体现在:

(1) 数据组织的结构化。数据库在本质上可以看作是文件的集合,数据库中的文件所存放的数据是按照一定的结构组织起来的,而且文件之间是有关联的。在文件系统中,尽管采用了像记录这一类简单的数据结构,但记录间没有关联,彼此是独立的。所以,从整体上看,文件系统中的数据是“涣散”的,而数据库中的数据是结构化的,具有统一的逻辑结构。数据的结构化是数据库的主要特征之一,是数据库和文件系统的根本区别。

(2) 减少数据冗余度,增强数据共享性。由于数据库中的文件彼此可以建立关联,如果同一数据在多个文件中重复出现,则可以将这些重复的数据单独组成文件,然后通过文件的关联将数据间的逻辑关系连接起来,从而可以最大限度地减少数据库中出现同一数据的多个副本的情况,这就有效地减少了数据的冗余。另外,从整个系统上看,数据不再面向某一个特定的应用程序,而是面向由所有应用程序组成的系统。所以,一个数据可以为多个应用程序所共享,一个应用程序也可以同时访问多个数据。这既可以提高数据的共享性,又可以减少数据的冗余度。

(3) 保证数据的一致性。通过建立文件间的关联,使得在对某一个数据进行更新时,与之相关的数据也得到相应更改,从而保证数据的一致性。这在文件系统中是难以实现的。

(4) 具有较高的数据独立性。在数据库系统中,数据独立性包含两个方面:一个是数据的物理独立性;另一个是数据的逻辑独立性。所谓数据的物理独立性,是指在数据的物理存储结构发生改变时,数据的逻辑结构可以不变的特性;数据的逻辑独立性则是指在总体逻辑结构改变时,应用程序可以保持不变的一种特性。

数据库之所以具有较好的数据独立性,主要是因为数据库提供了两个映像功能:数据的存储结构和逻辑结构之间的映像或转换功能以及数据的总体逻辑结构和局部逻辑结构之间的映像功能。前者保证了数据的物理独立性,后者则保证了数据的逻辑独立性。由于具有物理独立性,使得在数据的物理结构发生改变时,数据的逻辑结构并没有改变,从而基于该逻辑结构的应用程序也不需要改变;同时,由于具有逻辑独立性,使得在数据的总体逻辑结构发生改变时,一些应用程序涉及的局部逻辑结构没有改变,所以也不需要修改这些应用程序。可见,数据的独立性可以有效保证应用程序和数据的分离,从而极大地简化了应用程序的设计过程,减小了应用程序的维护代价。

(5) 以数据项为单位进行数据存取。相对文件系统而言,数据库可以实现更小粒度的数据处理,满足更多的应用需求。

(6) 具有统一的数据控制功能。这些功能包括数据的安全性控制、完整性控制、并发控制和一致性控制等功能。

总之,数据管理技术发展的这3个阶段是一个渐进的过程,它们之间的区别主要体现