

数据库原理与设计

——基于SQL Server 2012

- ◆ 数据库系统概论
- ◆ 关系模型与关系代数
- ◆ 数据库设计
- ◆ 数据库的规范化
- ◆ SQL语言、T-SQL
- ◆ 存储过程与触发器
- ◆ 数据库管理
- ◆ 系统体系结构



王世民 王雯 刘新亮 编著



清华大学出版社

高等学校计算机应用规划教材

内容简介

本书以数据库系统概论为基础，以SQL Server 2012为平台，全面系统地介绍了数据库系统的基本概念、原理、设计、实现、应用和维护。全书共分10章，主要内容包括：数据库系统概论、数据库系统组成、数据库系统性能优化、数据库系统安全、数据库系统备份与恢复、数据库系统维护、数据库系统应用、数据库系统开发、数据库系统测试、数据库系统部署。

数据库原理与设计

——基于 SQL Server 2012

王世民 王雯 刘新亮 编著

清华大学出版社

北京

清华大学出版社

地址：北京清华大学学研大厦A座

邮编：100084

清华大学出版社

地址：北京清华大学学研大厦A座

邮编：100084

清华大学出版社

地址：北京清华大学学研大厦A座

邮编：100084

清华大学出版社

地址：北京清华大学学研大厦A座

邮编：100084

清华大学出版社

地址：北京清华大学学研大厦A座

邮编：100084

清华大学出版社

产品编号：001788-01

内 容 简 介

本书根据理论与应用相结合的写作思路,用一个简单的数据库应用实例贯穿理论部分和应用部分,使读者清晰认识理论和应用各自解决的问题。在理论方面,详细介绍了数据库系统理论,包括数据模型、关系代数理论、数据库的设计、规范化理论、数据库管理和数据库的体系结构以及常用的 SQL 函数;在应用方面,主要介绍了在 SQL Server 2012 环境下如何完成数据库的建设、数据库的操作、数据库应用设计及与数据库的连接等。全部内容分为 9 章,为了便于读者理解和掌握,每章配有大量习题,有些习题可供读者上机使用。

本书可作为信息管理与信息系统专业、电子商务专业及其他相关专业数据库课程的教材,也可作为其他数据库应用人员的参考书。

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

数据库原理与设计:基于 SQL Server 2012/王世民,王雯,刘新亮 编著. —北京:清华大学出版社,2015
(高等学校计算机应用规划教材)

ISBN 978-7-302-39090-9

I. ①数… II. ①王… ②王… ③刘… III. ①关系数据库系统—高等学校—教材 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2015)第 017152 号

责任编辑:刘金喜

封面设计:牛艳敏

版式设计:思创景点

责任校对:曹 阳

责任印制:何 芊

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

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

课 件 下 载: <http://www.tup.com.cn>, 010-62794504

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:21.75

字 数:502 千字

版 次:2015 年 3 月第 1 版

印 次:2015 年 3 月第 1 次印刷

印 数:1~3500

定 价:35.00 元

前 言

数据库原理课程是信息管理与信息系统专业、电子商务专业本科学生的专业核心课程,本课程主要阐述数据库系统的基本原理、基本方法以及基本技术,理论性很强,学生在学习过程中对抽象的理论知识比较难理解,总有一种与实际脱钩的感觉。为此,在与其他老师探讨的基础上,结合多年教学经验,与另两位教师共同编写了本书。

本书的特点是突出理论与应用的结合,用一个数据库应用实例贯穿各章节,并随着内容的不断深入而完善数据库应用实例的设计。通过实例帮助学生理解抽象的理论知识;通过应用设计提高学生解决实际问题的能力;通过大量习题检查学生对基本知识的掌握程度;通过上机实验培养学生的实践能力,本书以 SQL Server 2012 为训练环境。

本书打破已有数据库原理书的写作思路,以数据库的设计、管理和应用为主线,组织各章的内容。全书共分为 9 章,具体内容如下:

第 1 章和第 2 章介绍数据库系统的基本概念,主要内容有数据模型和关系代数理论。

第 3 章和第 4 章介绍数据库的设计及优化,主要内容有数据库的概念设计、逻辑设计、物理设计和规范化理论。

第 5~7 章介绍数据库的操作和管理,主要内容有数据库的查询操作和更新操作、T-SQL 的编程技术、触发器的创建与使用、存储过程的创建与使用等内容。

第 8 章介绍数据库的连接、数据库的完整性控制、并发性控制、安全性控制和数据库的备份与恢复等。

第 9 章介绍数据库的体系结构,主要内容有常用数据库系统的体系结构、分布式数据库、主动数据库和多媒体数据库的概念等。

本课程建议授课为 38 学时,实验为 26 学时。

本书 PPT 课件可通过 <http://www.tupwk.com.cn/downpage> 下载。

本书由北京工商大学王世民、王雯、刘新亮老师编著并负责全书的统稿。本书在编写期间得到了北京工商大学计算机与信息工程学院领导和老师给予的大力支持和帮助,在此表示衷心的感谢。

由于编著者水平有限,书中难免有不妥之处,恳请读者原谅,并提出宝贵意见。

编著者

2014 年 9 月 1 日于北京

目

第1章 数据库系统概论	1
1.1 数据库系统的应用实例	1
1.1.1 学院教学管理系统	1
1.1.2 超市信息管理系统	3
1.2 数据管理的发展	4
1.2.1 人工管理阶段	4
1.2.2 文件系统阶段	5
1.2.3 数据库系统管理阶段	6
1.3 数据库系统概述	7
1.3.1 数据库系统的组成	7
1.3.2 数据库管理系统	9
1.4 数据模型	11
1.4.1 数据模型的类型和组成	11
1.4.2 概念模型	12
1.4.3 常用的数据模型	14
1.5 数据库系统结构	20
1.5.1 数据抽象过程	20
1.5.2 数据库系统的三级模式结构	21
1.6 本章要点回顾	23
练习题1	24
第2章 关系模型与关系代数	29
2.1 关系概念	29
2.1.1 域	29
2.1.2 笛卡儿积	30
2.1.3 关系的数学定义	31
2.1.4 关系模式、关系数据库模式和关系数据库	32
2.1.5 关系性质	32
2.2 关系的码	33
2.3 关系的完整性	35

录

2.3.1 实体完整性	35
2.3.2 参照完整性	35
2.3.3 用户定义完整性	36
2.4 关系操作、关系运算及关系代数	37
2.4.1 关系操作	37
2.4.2 关系运算	37
2.4.3 关系代数	38
2.4.4 关系运算的安全性和等价性	46
2.5 查询优化	47
2.5.1 关系代数表达式的优化问题	47
2.5.2 关系代数表达式的等价变化原则	48
2.5.3 关系代数表达式的优化策略	49
2.5.4 关系代数表达式的优化算法	50
2.6 本章要点回顾	51
练习题2	51
第3章 数据库设计	61
3.1 数据库的设计过程	62
3.1.1 数据库设计概述	62
3.1.2 需求分析	64
3.2 数据库概念结构设计	69
3.2.1 E-R方法	69
3.2.2 属性和联系概念的扩展	70
3.3 E-R模型的设计	73
3.3.1 概念结构设计方法和步骤	73
3.3.2 局部E-R模型的设计	74

3.3.3 局部 E-R 模型设计案例	75	4.4.5 第四范式	120
3.3.4 全局 E-R 模型的设计	78	4.5 本章要点回顾	121
3.3.5 全局 E-R 模型设计案例	80	练习题 4	122
3.3.6 优化全局 E-R 模型	81		
3.4 数据库逻辑结构设计	82	第 5 章 SQL 语言	131
3.4.1 逻辑结构设计步骤	82	5.1 SQL 概述	131
3.4.2 E-R 模型向关系模型转换	83	5.1.1 SQL 的发展历程	131
3.4.3 优化关系模式	86	5.1.2 SQL 数据库的体系结构	132
3.4.4 逻辑结构设计案例	86	5.1.3 SQL 的特点与组成	133
3.5 数据库物理结构设计	87	5.2 SQL Server 2012 入门知识	134
3.5.1 存储和文件结构	87	5.2.1 SQL Server 2012 的软硬件 环境	134
3.5.2 数据存储	88	5.2.2 SQL Server 2012 概述	135
3.5.3 文件结构	89	5.2.3 SSMS 概述	137
3.5.4 索引与散列	90	5.3 数据库的创建	139
3.5.5 SQL 中的索引定义	91	5.3.1 SQL Server 2012 数据库 概述	139
3.5.6 评价物理结构	92	5.3.2 创建用户数据库	141
3.6 本章要点回顾	92	5.3.3 数据库的修改与删除	145
练习题 3	93	5.4 表的创建	150
第 4 章 数据库的规范化	101	5.4.1 SQL Server 2012 表概述	150
4.1 什么是好的数据库设计	101	5.4.2 SQL Server 2012 表的创建	152
4.2 函数依赖理论	103	5.4.3 SQL Server 2012 表结构的 修改	157
4.2.1 函数依赖的定义	103	5.4.4 输入数据	159
4.2.2 完全函数依赖与部分函数 依赖	106	5.4.5 表的删除	161
4.2.3 传递函数依赖	107	5.5 视图的创建	162
4.2.4 键(码)	107	5.5.1 SQL Server 2012 视图 概述	162
4.3 关系模式的分解特性	109	5.5.2 SQL Server 2012 视图的 创建	163
4.3.1 模式分解存在的问题	109	5.5.3 SQL Server 2012 视图的 修改	164
4.3.2 无损连接	110	5.5.4 视图的应用	165
4.3.3 无损连接测试	111	5.6 简单查询	166
4.3.4 保持函数依赖测试	113	5.6.1 SELECT 子句和 FROM 子句	168
4.4 关系的规范化	114		
4.4.1 第一范式	115		
4.4.2 第二范式	116		
4.4.3 第三范式	117		
4.4.4 BCNF	119		

5.6.2 WHERE 子句	171	6.2.1 常量	210
5.6.3 GROUP BY 子句	175	6.2.2 数据类型	211
5.6.4 ORDER BY 子句	176	6.2.3 变量	215
5.7 聚集函数	177	6.3 运算符与表达式	218
5.7.1 COUNT 函数	177	6.4 T-SQL 流控制语句	220
5.7.2 SUM 函数	178	6.4.1 begin...end 语句块	220
5.7.3 AVG 函数	179	6.4.2 分支语句	221
5.7.4 MIN 和 MAX 函数	179	6.4.3 循环	222
5.8 空值	180	6.4.4 try...catch 语句块	224
5.9 连接查询	180	6.4.5 PRINT 语句	225
5.9.1 自连接查询	181	6.5 函数	225
5.9.2 内连接	181	6.5.1 数学函数	225
5.9.3 外连接	183	6.5.2 字符串函数	227
5.9.4 交叉连接	185	6.5.3 日期函数	228
5.10 集合运算	185	6.5.4 聚合函数	229
5.10.1 UNION 运算符	186	6.5.5 逻辑函数	230
5.10.2 INTERSECT 运算符	186	6.5.6 数据类型转换函数	231
5.10.3 EXCEPT 运算符	187	6.5.7 排序规则函数	231
5.11 嵌套查询	188	6.5.8 安全函数	232
5.11.1 单值比较嵌套查询	189	6.5.9 行集函数	234
5.11.2 带[NOT] IN 的嵌套 查询	190	6.5.10 图像文本函数	235
5.11.3 多值比较嵌套查询	192	6.5.11 触发器函数	236
5.11.4 带[NOT] EXISTS 的嵌套 查询	193	6.5.12 系统函数	237
5.12 更新操作	194	6.6 游标	239
5.12.1 插入数据	194	6.6.1 游标的类型	239
5.12.2 修改数据	196	6.6.2 声明游标	240
5.12.3 删除数据	198	6.6.3 使用游标	242
5.13 本章要点回顾	199	6.7 本章要点回顾	244
练习题 5	201	练习题 6	244
第 6 章 T-SQL	209	第 7 章 存储过程与触发器	245
6.1 概述	209	7.1 存储过程概述	245
6.1.1 SQL 与 T-SQL 的区别	209	7.1.1 存储过程的优点	245
6.1.2 T-SQL 的组成	209	7.1.2 存储过程的类型	246
6.2 常量与变量	210	7.2 创建存储过程	246
		7.2.1 T-SQL 命令创建存储过程	246
		7.2.2 修改存储过程	248

7.2.3	删除存储过程	249
7.2.4	查看存储过程	249
7.2.5	重命名存储过程	249
7.3	使用存储过程	250
7.4	触发器	255
7.4.1	概述	255
7.4.2	创建触发器	258
7.4.3	管理触发器	261
7.5	本章要点回顾	262
	练习题 7	263
第 8 章	数据库管理	265
8.1	数据库的安全性	265
8.1.1	安全性控制	266
8.1.2	安全性级别	267
8.1.3	权限	268
8.1.4	SQL 中安全性的控制	269
8.1.5	视图	271
8.1.6	数据加密法	271
8.1.7	自然环境的安全性	273
8.2	并发处理与并发控制	273
8.2.1	事务处理	273
8.2.2	事务的概念	274
8.2.3	事务的性质	274
8.2.4	并发处理与并发控制	275
8.2.5	并发操作带来的问题	275
8.2.6	封锁	278
8.2.7	活锁与死锁	280
8.2.8	封锁的粒度	282
8.3	数据库的备份与恢复	282
8.3.1	故障分类	283

8.3.2	恢复的原理与实现的方法	283
8.3.3	检查点机制	287
8.3.4	故障的恢复对策	288
8.4	数据库完整性的控制	288
8.4.1	完整性规则的组成	289
8.4.2	完整性约束的分类	289
8.4.3	SQL 中的完整性约束	290
8.5	本章要点回顾	296
	练习题 8	297
第 9 章	系统体系结构	305
9.1	数据库体系结构	305
9.1.1	客户/服务器体系结构	306
9.1.2	并行系统	307
9.1.3	分布式系统	308
9.2	并行数据库	309
9.2.1	概述	309
9.2.2	并行系统设计	310
9.3	分布式数据库	310
9.3.1	概述	311
9.3.2	分布式数据库的系统结构	313
9.3.3	数据分片	315
9.3.4	分布式数据库管理系统	316
9.3.5	分布式查询处理	317
9.4	主动数据库	318
9.5	多媒体数据库	320
9.6	本章要点回顾	321
	练习题 9	322

实验	323
-----------	------------

参考文献	339
-------------	------------

第 1 章 数据库系统概论

【学习目标】

学完本章后，你应该能够：

- 理解数据管理的发展阶段及各阶段的特点；
- 理解数据库系统的组成及数据库管理员(DBA)的职责；
- 掌握数据库管理系统(DBMS)的组成及功能；
- 理解数据模型的概念；
- 识记实体、实体集、属性、联系、码等基本术语；
- 理解层次模型、网状模型、关系模型的特点；
- 掌握数据库系统的体系结构。

【本章内容概要】

本章沿着数据管理发展阶段—数据库系统组成—数据库管理系统—数据模型—数据库体系结构这条概念主线介绍本章内容。首先由两个应用实例说明数据管理的重要性；如何进行管理，则介绍不同的数据管理方法，其中最完善的管理方法是数据库管理，由此引出数据库概念；数据库系统由数据库、软件、硬件和用户四部分组成，其中核心组成是数据库管理系统(DBMS)；接着介绍 DBMS 的组成及主要功能；数据库管理特点之一是数据结构化，即数据库采用数据模型表示复杂的数据结构，其中介绍数据模型概念及常用数据模型；最后介绍数据库体系结构。

1.1 数据库系统的应用实例

1.1.1 学院教学管理系统

学校作为教学组织单位，需要按照一定的教育方针，运用先进的管理手段，组织、协调、指挥各方面的工作，以高效率、高质量地完成各项教学工作，达到国家所制定的教育目标。学校的管理工作一般分为教学管理、行政管理和后勤管理，其中教学管理尤为重要，教学管理的主要工作包括学生学籍管理、考试管理、教学计划管理、教师任课管理、学生选课管理、教室管理、排课调课管理、教材管理、教学成果管理、学科建设、毕业实习和毕业论文管理等。为了对上述管理工作进行信息化管理，需要保存系、教研室、班级、学生、课程、教师、教材等基本信息，一般以二维表的形式保存，如表 1-1~表 1-5 所示。

张小强	男
-----	---

张小强	男
-----	---

齐力	男
----	---

齐力	男
----	---

程序设计	
------	--

程序设计	
------	--

三	四	五	六	七	八	九	十	十一	十二	十三	十四	十五	十六	十七	十八	十九	二十	二十一	二十二	二十三	二十四	二十五	二十六	二十七	二十八	二十九	三十	三十一	三十二	三十三	三十四	三十五	三十六	三十七	三十八	三十九	四十	四十一	四十二	四十三	四十四	四十五	四十六	四十七	四十八	四十九	五十	五十一	五十二	五十三	五十四	五十五	五十六	五十七	五十八	五十九	六十	六十一	六十二	六十三	六十四	六十五	六十六	六十七	六十八	六十九	七十	七十一	七十二	七十三	七十四	七十五	七十六	七十七	七十八	七十九	八十	八十一	八十二	八十三	八十四	八十五	八十六	八十七	八十八	八十九	九十	九十一	九十二	九十三	九十四	九十五	九十六	九十七	九十八	九十九	一百
---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

三	四	五	六	七	八	九	十	十一	十二	十三	十四	十五	十六	十七	十八	十九	二十	二十一	二十二	二十三	二十四	二十五	二十六	二十七	二十八	二十九	三十	三十一	三十二	三十三	三十四	三十五	三十六	三十七	三十八	三十九	四十	四十一	四十二	四十三	四十四	四十五	四十六	四十七	四十八	四十九	五十	五十一	五十二	五十三	五十四	五十五	五十六	五十七	五十八	五十九	六十	六十一	六十二	六十三	六十四	六十五	六十六	六十七	六十八	六十九	七十	七十一	七十二	七十三	七十四	七十五	七十六	七十七	七十八	七十九	八十	八十一	八十二	八十三	八十四	八十五	八十六	八十七	八十八	八十九	九十	九十一	九十二	九十三	九十四	九十五	九十六	九十七	九十八	九十九	一百
---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

表 1-5 班级

班 号	班 名	班主任编号
C51	信管	01002
C52	电商	02001
C53	电商	02003
C61	信管	01001
C62	电商	02002

1.1.2 超市信息管理系统

连锁超市起步于 20 世纪 90 年代初期,现已成为全国零售商业的一种重要形态。随着连锁超市的高速发展,其经营管理变得日益复杂,日常所需处理的数据量渐渐庞大,商业运转的中间环节也越来越多,原先主要依靠人工管理和简单计算机管理的方法,已无法适应超市规模的不断扩大。为此许多大中型超市采用了计算机信息管理系统,将其连锁超市的 3 个主要环节(门店、配送中心、超市总部)有机联网,实施全面的计算机管理,以实现超市从采购商品、入库、出库到商品的营销等各个环节的管理,可为商家提供全面、快捷的信息管理,降低超市的运营成本。

超市的进货管理、商品交易、库存管理、人员管理等管理工作都需要基础数据的支持,如供应商、商品、价格、库存、销售、现金账、支票账和人员等信息,如表 1-6~表 1-9 所示。主要解决核算和管理、数据的传送、管理数据的提取分析、基础信息的集中维护等。

表 1-6 商品

商品编码	商品名	生产厂家	生产日期	保质期	库存量

表 1-7 供应商

供应商编码	供应商名	地 址	电 话

表 1-8 顾客

顾客编码	顾客名	联系地址	电话
50010			180
10020			520
10030			820
10040			180

表 1-9 员工

员工号	员工名	性别	出生日期	来店工作日期	工种

由此可见,为了实现管理工作的信息化,需要管理大量的数据,目前大多采用数据库系统进行管理。为实现这种管理首先要设计数据的存储结构,即数据库的存储结构,还要设计数据的操作方法,即各种应用程序。另外,数据库系统还要保证数据存储的安全性,防止非法访问或越权访问数据库;还要保证系统的可靠性,若系统故障或者系统崩溃,要有办法恢复。这些问题如何解决将在本书后续章节进行介绍。

数据管理是指数据收集、存储、分类、加工、检索、传输等操作,随着计算机软硬件技术的发展,数据管理技术的发展大致经历了人工管理、文件系统和数据库系统管理3个发展阶段。

1.2 数据管理的发展

1.2.1 人工管理阶段

20世纪50年代中期以前,计算机除硬件外,没有软件对数据进行管理。若要使用计算机进行数据管理,设计人员不但要考虑应用程序的设计和逻辑定义及组织,而且还必须考虑数据在存储设备内的存储方式和地址。所以这个阶段的特点是:

(1) 数据不保存。因计算机主要用于科学计算,不要求保存数据。所以每次计算时,首先将程序和数据输入计算机主存,然后进行计算,最后将结果输出,计算机不保存程序和数据。

(2) 程序员的工作量大,程序难维护。程序员编写程序时还要考虑数据的物理存储,使得程序员的工作量很大。另外,当数据的物理存储改变时,则必须重新编写程序,程序难以维护。

(3) 数据面向程序。每个程序都有自己的一组数据，程序与数据融为一体，相互依赖。各程序之间的数据不能共享，所以数据重复存储。数据与程序之间的关系如图 1-1 所示。



图 1-1 数据与程序之间的关系

1.2.2 文件系统阶段

20 世纪 50 年代中后期至 60 年代中期，计算机有了磁盘等存储设备，软件有了操作系统和专门用于管理数据的应用软件。这时的计算机除了用于科学计算外，还被大量用来进行数据处理。这个阶段的特点是：

(1) 数据可以长期保存。因计算机大量用于数据处理，所以数据需要长期保存在外存中，并且数据是按一定的规则组织成为一个文件进行保存的。

(2) 数据的物理结构和逻辑结构有了区别。程序员不用考虑数据的物理位置，只需用文件名进行数据的读取。

(3) 文件形式多样化。为了适应不同的应用，文件有多种类型，如顺序文件、索引文件、链接文件等。

(4) 程序和数据之间有了一定的独立性。应用程序通过文件系统对数据文件中的数据进行读取和处理，程序员不需考虑数据的物理存储，程序和数据之间具有一定的独立性。

尽管文件系统有上述优点，但是文件仍然存在如下缺陷：

(1) 文件是面向特定用途组织的，而程序是根据文件设计出来的，两者相互依赖，缺乏灵活性和通用性。

(2) 数据冗余。在文件系统中，用户一般都是将完成一项应用所需要的数据存储在一个文件中。可能多个应用会涉及某些相同的数据项，这就造成同一个数据项重复存储的现象，这种现象称为数据冗余。数据冗余是文件系统的固有缺欠，很难克服。

(3) 文件与文件之间缺乏联系，不能反映现实世界事物之间的内在联系。

文件系统中程序与数据之间的关系如图 1-2 所示。

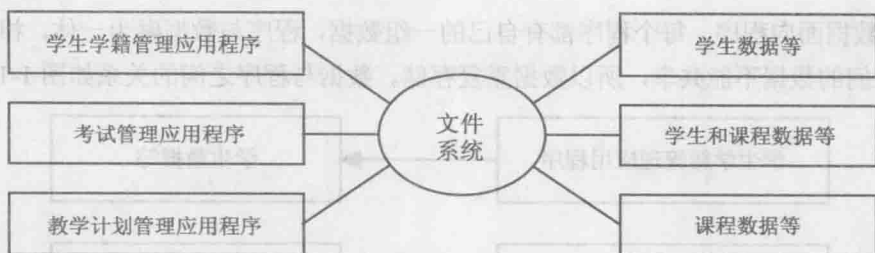


图 1-2 文件系统中程序与数据之间的关系

1.2.3 数据库系统管理阶段

20 世纪 60 年代末期开始,随着计算机软硬件技术的发展,人们研究了一种新的数据管理方法,即数据库系统。数据库系统克服了文件系统的缺陷,对数据的管理更加有效。它的基本思想是对所有数据实行统一的、集中的、独立的管理,使数据存储独立于使用数据的程序,实现数据共享。在数据库系统中用于集中管理独立存放数据的软件叫做数据库管理系统(DataBase Management System, DBMS),它是用户应用程序与数据之间的接口,如图 1-3 所示。

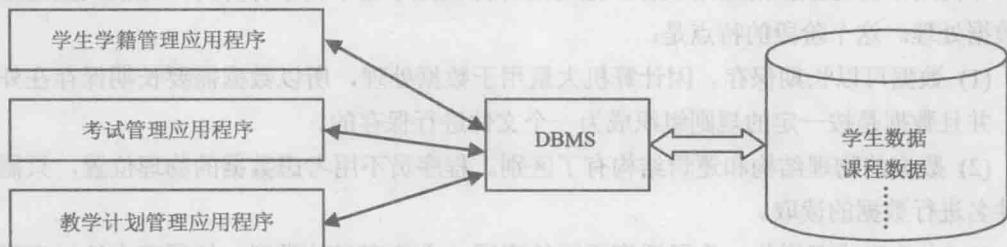


图 1-3 数据库系统管理方式

这个阶段的特点是:

- (1) 数据共享。数据库中的数据是面向所有用户的数据需求组织的,数据可以共享。
- (2) 数据独立性。数据独立性是指应用程序与数据之间相互独立,互不依赖。这在很大程度上减少了应用程序的设计与维护的工作量。
- (3) 数据结构化。数据库采用数据模型表示复杂的数据结构。数据模型不仅描述数据本身的特征,还要描述数据之间的联系。这也是有别于文件系统的—一个重要特征。
- (4) 可控数据冗余。由于上述 3 个特点使得数据不必重复存储,这不仅可以节省存储空间,而且可以从根本上保证数据的一致性。当数据完全不重复时,冗余度为零,但有时为了提高查询速度,也可保留一些重复数据,这种冗余是由用户控制的,称为可控冗余度。
- (5) 统一数据管理和控制功能。数据库中数据的共享一般是并发的,即多个用户同时存取数据库中的数据。因此, DBMS 必须提供以下 4 个方面的数据控制功能:

- 数据安全性控制。数据安全性控制是指采取一定的安全保密措施保护数据,以防止非法用户使用数据,造成数据的泄露、篡改或破坏。DBMS 必须为用户提供保

护数据安全的功能。

- 数据完整性控制。数据的完整性是指数据的正确性、有效性和相容性。所谓正确性,是指数据的合法性;所谓有效性,是指数据是否在定义的有效范围内;所谓相容性,是指表示同一个事实的两个数据应相同,不一致就是不相容。DBMS 必须提供数据完整性控制,防止错误的数据进入数据库系统而造成无效的操作。
- 并发控制。并发控制是指系统对并发事务进行合理调度,避免并发事务之间的相互干扰而得到错误的结果或使得数据库的完整性遭到破坏。因此,DBMS 必须提供并发控制以对所有用户的并发操作进行控制和协调。
- 数据恢复。数据恢复是指当系统发生故障造成数据库中部分或全部数据丢失时,DBMS 必须具有将数据库从错误状态恢复到故障发生前的某一已知的正确状态的功能。

从文件系统到数据库系统是信息处理领域的重大变化。在文件系统阶段,以加工数据的程序为中心,程序处于主导地位;在数据库系统管理阶段,以共享的数据库为中心,这样既有利于数据的集中管理,又有利于应用程序的设计与维护。

目前,各行各业都建立了以数据库为中心的管理信息系统。在学院教学管理系统中,学生、课程、教师、教材等基本信息都保存在数据库中;在超市信息管理系统中,供应商、商品、价格、库存、销售、现金账、支票账和人员等信息也保存在数据库中。如何完成数据库的设计和应用程序的设计将在后续章节中介绍。

1.3 数据库系统概述

1.3.1 数据库系统的组成

数据库系统(Database System, DBS)由数据库、软件、硬件和用户四部分组成。

1. 数据库(Database, DB)

数据库是指长期存储在计算机内、有组织的、统一管理的相关数据的集合。数据库中的数据按一定的数据模型组织和存储,可共享并具有较小的冗余度,数据之间相互联系而又有较高的独立性。

2. 软件

数据库系统的软件包括数据库管理系统(DBMS)、操作系统(OS)、数据库应用系统等。其中,DBMS 是数据库系统的核心软件,它是负责数据存取、维护和管理的软件,是用户使用数据库的接口,任何一个应用程序如果要访问数据库,必须通过 DBMS。

3. 硬件

硬件是指存储数据库和运行数据库管理系统的硬件资源。它包括物理存储数据库的存储设备(如磁盘等)、CPU、控制器、内存、I/O 通道和外部设备等。

数据库一般存储在外存中,当运行时才将其调入内存。随着数据库存储数据量的增大,要求有足够大的内存,以存储操作系统、DBMS、应用程序和数据缓冲区;同时也要求有足够大的外存和较高的通道能力。

4. 用户

用户是指一组使用数据库的不同人员,他们向数据库提出存储、维护和查询数据等各类请求。数据库系统中主要有 4 类用户:数据库管理员、系统分析与设计人员、应用程序员和终端用户(也称最终用户)。

(1) 数据库管理员(Database Administrator, DBA)

DBA 是全面负责数据库系统的管理、维护和正常使用的人员。DBA 的主要职责有:

- 设计数据库的结构和内容。DBA 要参与数据库设计的全过程,并与最终用户、应用程序员、系统分析与设计人员密切配合,共同完成数据库的设计。
- 定义数据库的存储结构和存取方式。以获得数据存储空间的高利用率和较高的存取效率为目的,完成数据库存储结构和存取方式的定义。
- 定义数据的安全性和完整性。DBA 通过授权机制规定各个用户访问数据库的权限,以确保数据库的安全性。通过定义完整性约束,以保证数据的完整性。
- 数据库的日常维护。DBA 需要定期备份数据库,以防故障发生时数据丢失。DBA 要监控数据库的运行,确保数据库的性能不会因某些用户延时操作而下降。
- 改进和重新构造数据库系统。DBA 要不断地改进数据库设计和重新构造数据库的物理组织,以适应实际应用需求的变化和提高系统的性能。

DBA 除了要掌握一定的数据处理、数据库技术之外,还应有处理好人际关系的素质、能力。在一个企事业中,特别是一个规模较大的数据库,不可能由一两个人来完成管理工作,所以 DBA 常指数据库管理部门。

开发数据库系统时,一开始就应设置 DBA 的职位或相应的机构,以明确 DBA 的职责、权限。

(2) 系统分析与设计人员(System Analyst and Designer)

系统分析与设计人员是指负责需求分析、数据库设计和系统功能设计的人员。他们与用户和 DBA 配合,对用户的需求进行调查,并对应用系统进行分析,编写系统分析报告,确定系统的软硬件配置,设计数据库和系统功能结构。在很多情况下,数据库设计人员通常由 DBA 担任。

(3) 应用程序员(Application Programmer)

应用程序员是指负责设计和编写应用程序的人员,他们还负责程序的安装与调试。

(4) 终端用户(End User)

终端用户是指从计算机联机终端存取数据库的人员。他们通过数据库系统提供的终端

命令或表格语言或菜单驱动等交互式对话方式来存取数据库中的数据。终端命令简单、易懂、易学、易用，特别适合于非计算机专业人员使用。

1.3.2 数据库管理系统

1. DBMS 的基本功能

(1) 数据定义功能

DBMS 提供数据定义语言(DDL)来定义数据库结构、定义数据库中数据之间的联系、定义数据完整性约束条件和保证完整性的触发机制等。

(2) 数据操作功能

DBMS 提供数据操作语言(DML)来完成对数据库中数据的查询、插入、删除、修改等操作，其中查询是最重要和最常用的一类操作，所以有些系统把 DML 称为查询语言。

DML 语言有两类：一类是交互式命令语言，语法简单，可独立使用，称为自主型；另一类是把数据库存取语言嵌入在主语言(如 C 语言)中使用，它不能独立使用，称为宿主型的。目前，很多 DBMS 具有自主型和宿主型两类 DML，允许用户选择使用，如 SQL 语言。

(3) 数据库运行的管理功能

DBMS 提供数据控制功能，完成对数据库的安全性控制、完整性控制和并发控制，以确保对数据库的运行进行有效的控制和管理。

(4) 数据库维护功能

DBMS 通过实用程序来完成初始数据的装入、数据库的转储、系统目录的自动维护、备份、恢复及性能监控等维护功能。

2. DBMS 的组成

DBMS 一般是由许多“系统程序”所组成的一个集合，不同的 DBMS 其组成有所不同，但一般都由以下三部分组成。

(1) 语言编译处理程序

语言编译处理程序主要包括以下程序：

- 数据定义语言 DDL 翻译程序。
- 数据操作语言 DML 处理程序。将应用程序中的 DML 语句转换成可执行语句。
- 终端命令解释程序。解释执行每一个终端命令。
- 数据库控制命令解释程序。解释执行每一个控制命令。

(2) 系统运行控制程序

系统运行控制程序主要包括以下程序：

- 系统总控程序。是 DBMS 运行程序的核心，它控制和协调各程序的运行。
- 存取控制程序。用于检查用户或应用程序是否合法。
- 并发控制程序。协调各应用程序对数据库的操作，保证数据的一致性。