

Access 2010数据库 应用技术案例教程

- ◆ 数据库系统概述
- ◆ 走进Access
- ◆ 数据库和表
- ◆ 查询
- ◆ 窗体
- ◆ 报表
- ◆ 宏
- ◆ VBA程序设计
- ◆ ADO数据库编程
- ◆ 数据库综合操作案例
- ◆ 数据库应用系统开发



主 编 刘 垣

副主编 林敦欣 连贻捷 林铭德

张波尔 刘 琰



清华大学出版社

Access 2010数据库应用技术案例教程

主要内容

本书全面讲述了Access 2010数据库应用技术的理论知识和案例应用。全书共11章，其内容主要包括：数据库领域的基础核心内容，Access的发展历程、集成环境、安全性和数据库对象的应用，软件工程基础和小型数据库应用系统的开发。

读者对象

本书可作为全国计算机等级考试二级Access数据库程序设计考试、福建省高校计算机应用水平二级Access 2010关系数据库考试的备考用书，也可作为数据库应用技术的专门用书，还可作为广大计算机爱好者的Access培训用书或参考用书。

本书特色

本书将Access 2010置于互联网和Windows的大环境中介绍，注重计算思维的培养，以大学教务管理数据库案例贯穿始终，选用思维导图引领各章学习。配套的《Access 2010数据库应用技术案例教程学习指导》(ISBN: 978-7-302-49443-0)各章提供了多种类型的实验案例，适合翻转课堂教学。

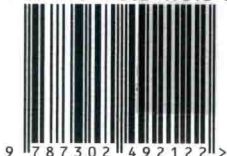
本书提供配套的电子教案和案例素材，下载网址见内容简介。

清华社官方微信号



扫 我 有 惊 喜

ISBN 978-7-302-49212-2



9 787302 492122 >

定价：58.00元

高等学校计算机应用规划教材

Access 2010 数据库 应用技术案例教程

主 编 刘 垣

副主编 林敦欣 连贻捷 林铭德

张波尔 刘 琰

清华大学出版社

北 京

内 容 简 介

本书第1章为数据库系统概述,综述了数据库系统的概念与结构、数据模型、关系数据库、数据库设计的方法和步骤等数据库领域的基础核心内容,第2章至第10章依次介绍了 Access 的发展历程、集成环境、安全性和6种标准数据库对象的基础应用与综合应用,第11章介绍了软件工程基础和几个小型数据库应用系统。本书将 Access 2010 置于互联网和 Windows 的大环境中介绍,注重计算思维的培养,以大学教务管理数据库案例贯穿始终,选用思维导图引领各章学习。配套的《Access 2010 数据库应用技术案例教程学习指导》(ISBN: 978-7-302-49443-0)中各章提供多种类型的实验案例,适合翻转课堂教学。

本书涵盖《全国计算机等级考试二级 Access 数据库程序设计考试大纲(2016 年版)》《福建省高校计算机应用水平二级数据库应用技术 Access 2010 关系数据库考试大纲(2017 版)》的全部知识点,可以作为备考用书,也可以作为数据库应用技术的专门用书,亦适合作为广大计算机爱好者的 Access 培训用书或参考用书。

本书各章对应的素材和电子教案可以通过 <http://www.tupwk.com.cn/downpage> 下载。

图书在版编目(CIP)数据

Access 2010 数据库应用技术案例教程 / 刘垣 主编. —北京:清华大学出版社, 2018
(高等学校计算机应用规划教材)

ISBN 978-7-302-49212-2

I. ①A… II. ①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(2017)第 331795 号

责任编辑:王 定

版式设计:思创景点

封面设计:孔祥峰

责任校对:曹 阳

责任印制:杨 艳

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

印装者:三河市君旺印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:20.75

字 数:480 千字

版 次:2018 年 2 月第 1 版

印 次:2018 年 2 月第 1 次印刷

印 数:1~3000

定 价:58.00 元

产品编号:077744-01

前 言

“工欲善其事，必先利其器。”在当今“互联网+大数据”时代，Access 可谓桌面关系型数据库系统的“利器”，为非计算机专业用户解决工作、学习和生活中的数据存储与数据分析问题提供了一种便捷方式。伴随着经济社会发展进入“新常态”，我国大学生就业、创业所面临的形势与任务也呈现出新的特点。利用 Access 快速开发小型数据库应用程序，亦有助于大学生创新创业管理，锦上添花。

微软公司自 1992 年 11 月首次推出 Access 以来，25 年间，经过了 11 个版本的变迁。从 Access 2007 版本开始，数据库文件格式一直沿用至今，微软公司对将来下一个版本的 Access 也没有打算改变这种数据库文件格式。从 2018 年起，福建省高校计算机应用水平 Access 考试将改版，从 2003 版升至 2010 版，升级为与全国计算机等级考试一样的版本。

为编写此书，我们花费大量时间收集国内外有关 Access 研究的资讯，分析本书受众的需求，研讨本书的体例，结合多年教学实践经验，以面向应用、面向创新为准则，设计了一个大学生熟悉、也适于翻转课堂教学的大学教务管理数据库案例，以任务驱动的方式将整个案例分解、贯穿于全书的各章节，使学习者处在一个从感性认识到知识理解，再到实践应用的学习环境中，由浅入深、循序渐进地掌握 Access 数据库的基础知识。

本书涵盖《全国计算机等级考试二级 Access 数据库程序设计考试大纲(2016 版)》和《福建省高校计算机应用水平二级数据库应用技术 Access 2010 关系数据库考试大纲(2017 版)》的核心内容。每一章的开篇给出了“学习目标”“学习方法”“学习指南”“思维导图”的要点，引导学习者通过思维导图理清章节知识脉络。

本书虽然以 Access 2010 为主要对象，但也兼顾了 Access 低版本和最新版。我们将 Access 置于互联网和 Windows 的大环境中介绍，涉及 Windows 的控制面板、注册表、组策略等程序功能的应用，希望有助于学习者运用计算思维，从宏观整体把控 Access，懂得运用数据库技术解决一些实际问题的基本路径，并由此激发学习者结合自身专业需求在数据库领域继续深入探索。

本书第 1 章为数据库系统概述，综述了数据库系统的概念与结构、数据模型、关系数据库、数据库设计与步骤等数据库领域的基础核心内容，初学数据库应用技术的读者可以先泛读本章，待学完后续各章再精读，并完成第 1 章的实验案例。第 2 章至第 10 章依次介绍了 Access 的发展历程、集成环境、安全性和 6 种标准数据库对象的基础应用与综合应用。第 11 章介绍了软件工程基础和利用 Access 建立的几个小型数据库应用系统。通过全书的学习，读者能够掌握关系数据库系统的基本理论和 Access 的基本操作技能，具备利用 Access 开发小型数据库应用系统的能力。

本书案例在 64 位 Win7/Win10+Access 2010(32 位)环境中调试通过，标题带“*”的部

分为选学内容。

我们以历史和发展的视角编写《Access 2010 数据库应用技术案例教程》，尽可能体现数据库领域新的发展动向、计算机教育新的研究成果，设计既贴近生产生活又体现 Access 特色的典型案例，力求能反映当代 Access 应用的技术水平。我们将尝试案例教程与超星泛雅网络教学平台配合使用，以探索将应用技术用于大学计算机基础教与学的新模式。

本书编写人员都是高校计算机教学一线教师，对 Access 数据库教学做过较深入的探索，案例教程中的许多内容是作者依据多年教学讲授和实践总结而形成的。其中，刘垣编写第 1 章、第 2 章、第 10 章、第 11 章，林敦欣编写第 5 章和第 7 章，连贻捷编写第 8 和第 9 章，林铭德编写第 6 章，张波尔编写第 4 章，刘琰编写第 3 章。全书由刘垣审改和统稿。此外，参与本书编写的还有许锐、邝凌宏、郭李华、肖琳、王晨阳、林好、徐沛然、温馨和苏备迎等人，在此一并表示衷心感谢！

由于作者水平有限，书中难免存在疏漏和不足，殷切希望广大读者提出宝贵的批评意见和建议。联系电子邮箱：hnlwangd@163.com，电话 010-62794504。

作 者

2017 年 9 月于榕城

目 录

第 1 章 数据库系统概述	1
1.1 数据库技术	3
1.1.1 数据管理技术的产生与发展	3
1.1.2 基本术语	7
1.1.3 数据库系统的结构	11
*1.1.4 图灵奖得主	13
1.2 数据模型	15
1.2.1 数据模型的分层	15
1.2.2 概念模型的表示方法: 实体- 联系方法	20
1.2.3 关系模型的基本术语及性质	23
1.2.4 关系运算	26
1.2.5 关系的完整性	27
1.3 数据库设计	29
1.3.1 数据库系统的需求分析	29
1.3.2 概念结构设计	30
1.3.3 逻辑结构设计	30
1.3.4 物理结构设计	34
1.3.5 数据库的实施	35
1.3.6 数据库的运行和维护	35
1.4 本章小结	35
1.5 思考与练习	36
1.5.1 选择题	36
1.5.2 填空题	37
1.5.3 简答题	37
第 2 章 走进 Access	38
2.1 Access 概述	40
2.1.1 Access 的发展历程	40
2.1.2 Access 的特点	43
2.2 Access 2010 的安装	46
2.2.1 Access 的多版本安装	46

2.2.2 Access 的单版本安装	49
2.3 Access 2010 的启动与退出	50
2.3.1 Access 2010 的启动	50
2.3.2 Access 2010 的退出	51
2.4 Access 2010 集成环境	52
2.4.1 Access 2010 的主窗口	53
2.4.2 Access 2010 的帮助	61
2.5 Access 数据对象的导入与导出	62
2.5.1 Access 2010 的数据库对象	63
2.5.2 Access 数据对象的导入	64
2.5.3 Access 数据对象的导出	65
2.6 Access 2010 的安全性	67
2.6.1 信任中心	67
2.6.2 禁用模式	69
2.6.3 数据库的打包、签名和分发	70
2.6.4 数据库的压缩和修复	72
2.6.5 数据库的加密与解密	73
2.6.6 其他安全措施	74
2.7 本章小结	75
2.8 思考与练习	75
2.8.1 选择题	75
2.8.2 填空题	76
2.8.3 简答题	76
第 3 章 数据库和表	77
3.1 创建数据库	79
3.1.1 建立新数据库	79
3.1.2 数据库的基本操作	81
3.1.3 管理数据库	83
3.2 创建表	85
3.2.1 使用表模板创建数据表	85
3.2.2 使用表设计创建数据表	86
3.2.3 通过输入数据创建表	88

3.2.4 使用 SharePoint 列表创建表	89	4.4 选择查询	124
3.2.5 通过获取外部数据创建表	90	4.4.1 不带条件的选择查询	124
3.2.6 主键的设置	91	4.4.2 条件表达式	125
3.3 数据类型与字段属性	92	4.4.3 带条件的选择查询	126
3.3.1 数据类型	92	4.4.4 在查询中使用计算	127
3.3.2 字段属性	93	4.5 交叉表查询	128
3.4 建立表之间的关系	101	4.6 参数查询	129
3.4.1 创建表关系	101	4.7 操作查询	130
3.4.2 查看与编辑表关系	103	4.7.1 追加查询	130
3.4.3 实施参照完整性	104	4.7.2 更新查询	131
3.4.4 设置级联选项	104	4.7.3 删除查询	131
3.5 编辑数据表	104	4.7.4 生成表查询	132
3.5.1 向表中添加与修改记录	104	4.8 SQL 查询	132
3.5.2 选定与删除记录	106	4.8.1 SQL 概述	132
3.5.3 数据的查找与替换	106	4.8.2 Select 语句格式	133
3.5.4 数据的排序与筛选	107	4.8.3 Select 语句示例	133
3.5.5 行汇总统计	111	4.8.4 SQL 数据更新命令	134
3.5.6 表的复制、删除与重命名	111	4.9 本章小结	135
3.6 设置数据表格式	112	4.10 思考与练习	136
3.6.1 设置表的行高和列宽	112	4.10.1 选择题	136
3.6.2 设置字体格式	113	4.10.2 填空题	137
3.6.3 隐藏和显示字段	114	4.10.3 简答题	137
3.6.4 冻结和取消冻结	114	第 5 章 窗体	138
3.7 本章小结	115	5.1 窗体概述	140
3.8 思考与练习	115	5.1.1 窗体的功能	140
3.8.1 选择题	115	5.1.2 窗体的类型	140
3.8.2 填空题	116	5.1.3 窗体的视图	142
3.8.3 简答题	116	5.1.4 窗体的构成	143
第 4 章 查询	117	5.2 创建窗体	144
4.1 查询的基本概念	118	5.2.1 自动创建窗体	145
4.1.1 查询的功能	118	5.2.2 使用向导创建窗体	147
4.1.2 查询的类型	119	5.2.3 使用“空白窗体”按钮创建窗体	149
4.1.3 查询视图	120	5.2.4 使用设计视图创建窗体	150
4.2 使用向导创建查询	120	5.2.5 创建主/子窗体	150
4.2.1 使用“简单查询向导”	120	5.3 设计窗体	151
4.2.2 使用“交叉表查询向导”	121	5.3.1 窗体设计视图的组成与主要功能	152
4.2.3 其他向导查询	123	5.3.2 为窗体设置数据源	152
4.3 使用“查询设计视图”创建查询	123		

5.3.3	窗体的常用属性与事件	153	7.1.1	什么是宏	198
5.3.4	在窗体中添加控件的方法	155	7.1.2	宏的类型	199
5.3.5	常用控件及其功能	155	7.1.3	宏的设计视图	199
5.3.6	常用控件的使用	158	7.1.4	常用的宏操作	200
5.4	修饰窗体	165	7.1.5	宏的结构	201
5.4.1	主题的应用	165	7.2	独立宏的创建与运行	202
5.4.2	条件格式的使用	166	7.2.1	创建独立宏	202
5.4.3	窗体的布局及格式调整	166	7.2.2	独立宏的运行	203
5.5	定制用户入口界面	167	7.2.3	自动运行宏	204
5.5.1	创建切换窗体	167	7.2.4	宏的调试	205
5.5.2	创建导航窗体	168	7.3	嵌入宏的创建与运行	206
5.5.3	设置启动窗体	169	7.3.1	创建嵌入宏	206
5.6	本章小结	169	7.3.2	通过事件触发宏	206
5.7	思考与练习	170	*7.4	数据宏的创建与运行	207
5.7.1	选择题	170	7.4.1	创建数据宏	207
5.7.2	简答题	170	7.4.2	数据宏的运行	209
第6章	报表	171	7.5	宏的应用	209
6.1	认识报表	172	7.5.1	自定义菜单简介	209
6.1.1	报表的类型	172	7.5.2	自定义功能区菜单的创建	210
6.1.2	报表的组成与报表视图	173	7.5.3	自定义快捷菜单的创建	212
6.2	快捷创建报表	176	7.6	本章小结	213
6.2.1	创建基本报表	176	7.7	思考与练习	214
6.2.2	使用空报表创建报表	177	7.7.1	选择题	214
6.2.3	使用报表向导创建报表	178	7.7.2	简答题	214
6.3	报表设计	181	第8章	VBA 程序设计	215
6.3.1	设计报表的布局 and 背景	181	8.1	程序设计概述	216
6.3.2	使用报表设计视图创建 报表	184	8.1.1	程序与程序设计语言	216
6.3.3	报表的排序、分组和计算	186	8.1.2	VBA 概述	218
6.4	创建其他报表	191	8.2	数据类型、表达式和内部 函数	221
6.4.1	创建图表报表	191	8.2.1	数据类型	221
6.4.2	创建标签报表	193	8.2.2	常量	223
6.5	导出报表	194	8.2.3	变量	224
6.6	本章小结	195	8.2.4	数组	226
6.7	思考与练习	196	8.2.5	运算符和表达式	228
6.7.1	选择题	196	8.2.6	VBA 内部函数	231
6.7.2	简答题	196	8.3	程序控制结构	234
第7章	宏	197	8.3.1	VBA 基本语句	234
7.1	宏的概述	198	8.3.2	顺序结构	236

8.3.3 分支结构	238	10.11 思考与练习	292
8.3.4 循环结构	245	10.11.1 基本操作题	292
8.3.5 过程与函数	254	10.11.2 简单应用题	293
8.4 本章小结	256	10.11.3 综合应用题	293
8.5 思考与练习	256	第 11 章 数据库应用系统开发	294
8.5.1 选择题	256	11.1 软件工程基础	296
8.5.2 填空题	258	11.1.1 基本概念	296
8.5.3 简答题	258	11.1.2 软件测试	297
8.5.4 程序设计题	258	11.1.3 程序的调试	298
第 9 章 ADO 数据库编程	259	11.2 Access 数据库开发特色	298
9.1 ADO 概述	260	11.2.1 支持两种数据库	298
9.1.1 数据库引擎和接口	260	11.2.2 Access 自带分析工具	299
9.1.2 ADO	262	11.2.3 Access 支持的字符集	300
9.2 ADO 主要对象	262	11.2.4 小型数据库应用系统 开发	301
9.2.1 ADO 对象模型	262	11.3 驾驶人科目一模拟考试系统	301
9.2.2 Connection 对象	263	11.3.1 系统功能简介	303
9.2.3 Recordset 对象	265	11.3.2 系统 VBA 源代码简介	305
9.2.4 Command 对象	270	11.3.3 系统的维护与升级	307
9.3 ADO 在 Access 中的应用	271	11.4 客户管理系统	308
9.3.1 ADO 编程方法	271	11.4.1 系统功能简介	308
9.3.2 ADO 编程实例	273	11.4.2 系统 VBA 源代码简介	308
9.4 本章小结	277	11.5 商贸进销存管理系统	312
9.5 思考与练习	277	11.5.1 系统功能简介	312
9.5.1 选择题	277	11.5.2 盟威平台实现系统功能	312
9.5.2 填空题	278	11.6 采购报销管理系统	315
9.5.3 简答题	278	11.6.1 系统功能简介	315
9.5.4 设计题	278	11.6.2 盟威平台实现系统功能	316
第 10 章 数据库综合操作案例	279	11.7 精准扶贫信息化平台	317
10.1 综合案例 1	281	11.7.1 系统功能简介	317
10.2 综合案例 2	282	11.7.2 Office 中国 Access 通用 平台实现系统功能	317
10.3 综合案例 3	283	11.8 本章小结	319
10.4 综合案例 4	284	11.9 思考与练习	319
10.5 综合案例 5	285	11.9.1 选择题	319
10.6 综合案例 6	287	11.9.2 填空题	320
10.7 综合案例 7	288	11.9.3 简答题	320
10.8 综合案例 8	289	参考文献	321
10.9 综合案例 9	290		
10.10 本章小结	292		

第1章 数据库系统概述

学习目标

1. 掌握数据库系统的基础知识。
2. 掌握关系数据库的基本原理。
3. 掌握数据库应用系统设计方法。
4. 了解数据管理技术的发展历史和趋势。
5. 了解数据库系统领域的杰出人物。

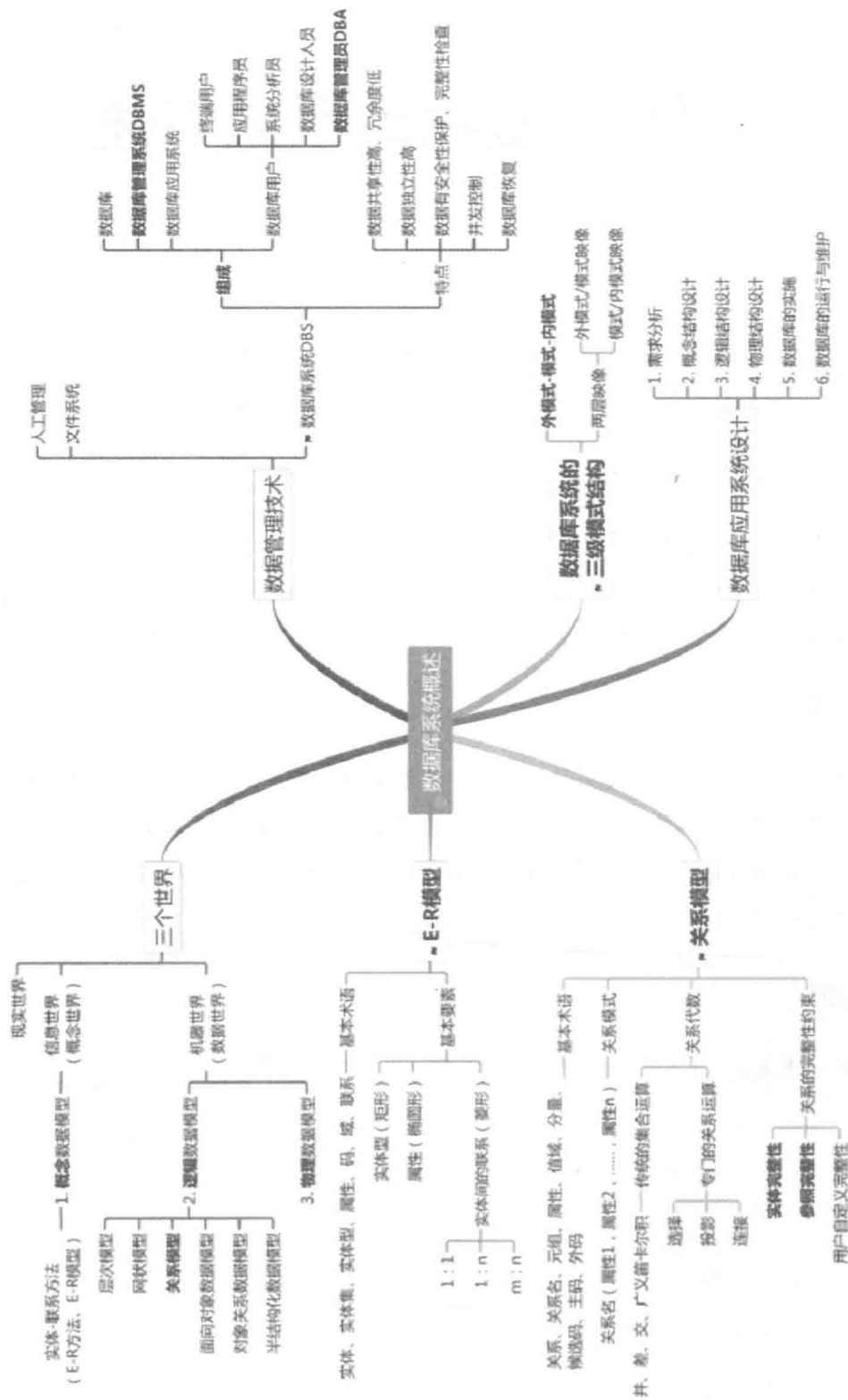
学习方法

本章内容理论性较强，涵盖的知识面广，涉及的概念多，学习者不容易理解掌握。建议初学者循序渐进地学习本章，通过比对、类比等方式识记数据库基本术语和基本原理，仔细领会各案例的前提条件和解决方法，勤思考、多动手练习和归纳总结，依据本章的思维导图理清知识脉络。

学习指南

本章的重点是 1.1.2 节、1.1.3 节和 1.2 节，难点是 1.2 节。

思维导图



1.1 数据库技术

数据库是数据管理的有效技术，是计算机科学的重要分支，是现代大数据管理和分析的基石。数据库技术诞生于 20 世纪 60 年代末期，历经数代演变，已造就了四位图灵奖获得者：Charles W. Bachman(1973 年)、Edgar F. Codd(1981 年)、James Nicholas Gray(1998 年)和 Michael Stonebraker(2014 年)。

数据库技术有着扎实的理论基础和广泛的应用领域。农、林、牧、渔业，采矿业，制造业，电力、热力、燃气及水生产和供应业，建筑业，批发和零售业，交通运输、仓储和邮政业，住宿和餐饮业，信息传输、软件和信息技术服务业，金融业，房地产业，租赁和商务服务业，科学研究和技术服务业，教育等国民经济行业都有数据库应用的印迹。人们的日常生活也无时无刻不享受着数据库应用的便利，例如，网络购买虚实物品、各种智能终端便捷消费、各类新媒体及其服务、虚拟现实的体验等。

1.1.1 数据管理技术的产生与发展

数据库技术是应数据管理任务的需要而产生的。数据管理是指对数据进行分类、组织、编码、存储、检索、维护和应用，它是数据处理的中心问题。数据处理是指对各种数据进行采集、存储、检索、加工、传播和应用等一系列活动的总和。随着应用需求的推动和计算机软硬件的发展，数据管理技术经历了人工管理、文件系统和数据库系统三个阶段。这三个阶段产生的背景与特点如表 1-1 所示。

表 1-1 数据管理三个阶段的比较

三个阶段 背景与特点		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、数据管理	大规模数据管理
	硬件背景	无直接存取存储设备	磁带、磁鼓、磁盘	大容量磁盘、磁盘阵列
	软件背景	没有操作系统(OS)	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理、批处理	联机实时处理、批处理、分布处理
特点	数据的管理者	用户(程序员)	文件系统	数据库管理系统(DBMS)
	数据面向的对象	某一应用程序	某一应用	现实世界(一个部门、企业等)
	数据的共享程度	无共享，冗余度极大	共享性差，冗余度大	共享性高，冗余度小
	数据的独立性	不独立，完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构，整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由 DBMS 提供数据安全性、完整性、并发控制和恢复能力

1. 人工管理阶段

数据管理技术的人工管理阶段主要指 20 世纪 50 年代中期以前,数据需要由应用程序定义和管理,一个数据集只能对应一个应用程序。数据无共享,冗余度极大;数据不独立,完全依赖于程序。这个阶段的计算机很简陋,主要应用于科学计算。计算机硬件状况是:没有直接存取存储设备;软件状况是:没有完整的操作系统,没有管理数据的专门软件;数据处理方式是批处理。批处理方式是对需要处理的数据不做立即处理,待积累到一定程度、一定时间,再成批地进行处理。

Herman Hollerith(1860.02.29—1929.11.17)被广泛视为现代机器数据处理之父。为解决美国人口普查统计繁难、花费时间长的问题, Hollarith 根据自动提花织布机原理,利用穿孔卡片记录美国人口普查信息,发明了穿孔卡片制表机(punch card tabulating machine)。这种穿孔卡片制表机是机电式计数装置,安装有一组盛满水银的小杯,已穿孔的卡片放置在这些水银杯上。卡片上方有几排精心设置的金属探针,探针连接在电路的一端,水银杯连接在电路的另一端。只要某根探针遇到卡片上有孔的位置,便会自动掉落下去,与水银接触即接通电路,启动计数装置前进一个刻度。Hollerith 的穿孔卡片表达了二进制思想:有孔处能接通电路计数,代表该调查项目为“有”,即 1;无孔处不能接通电路计数,表示该调查项目为“无”,即 0。

1890 年美国人口普查首次选用 Hollerith 的穿孔卡片制表机,获得巨大成功。在此后的计算机系统里,用穿孔卡片输入数据的方法一直沿用到 20 世纪 70 年代,数据处理也发展成为计算机的主要功能之一。Hollerith 发明的制表机和穿孔卡片如图 1-1 和图 1-2 所示,制表机由两部分构成:整理箱(sorting box)由制表机控制,分拣机(sorter)是后期发展起来的独立机器。Hollerith 于 1896 年创办的公司 Tabulating Machine Company 后来成为 IBM(International Business Machines Corporation)的前身之一。1900—1950 年,穿孔卡片是商务数据存储和检索的基本形式,IBM 是组合和排序穿孔卡片设备以及利用穿孔卡片数据打印报表的主要供应商。



图 1-1 Hollerith 和穿孔卡片制表机

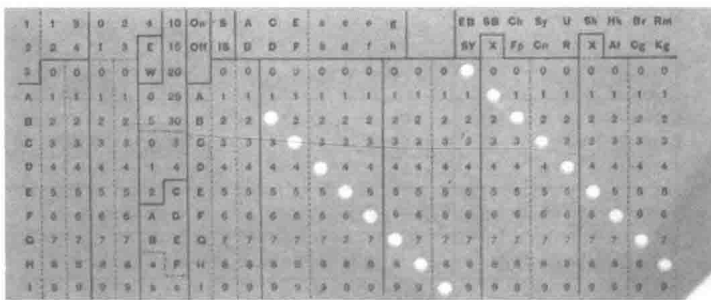


图 1-2 穿孔卡片

2. 文件系统阶段

文件系统阶段主要指 20 世纪 50 年代末到 20 世纪 60 年代中期,利用文件系统管理数据。对于一个特定的应用,数据被集中组织存放在多个数据文件或文件组中,并针对该文件组开发特定的应用程序。数据的共享性差,冗余度大;数据独立性差。

这个阶段的计算机已有操作系统,在操作系统基础之上建立的文件系统已经成熟并广泛应用。计算机除了应用于科学计算,也开始应用于数据管理。数据处理方式不仅有批处

理, 还有联机实时处理。实时处理是指需要对收集到的数据立即进行处理, 并及时反馈。

在这个阶段, 磁带被主要用于数据存储。数据处理是从一个或多个磁带上顺序读取数据, 再将数据写回到新的磁带上。数据也可以由一叠穿孔卡片输入, 再输出到打印机上。

以美国在 20 世纪 60 年代初制定的阿波罗计划(Apollo Program/Project Apollo)为例, 阿波罗飞船由约 200 万个零部件组成, 它们分散在世界各地制造生产。为了掌握计划进度及协调工程进展, 阿波罗计划的主要合约者 Rockwell 公司曾研制开发了一个基于磁带的零部件生产计算机管理系统, 系统共用了 18 盘磁带, 虽然可以工作, 但效率极低, 18 盘磁带中 60%是冗余数据, 维护十分困难。这个系统曾一度成为美国实现阿波罗计划的重大障碍之一。

3. 数据库系统阶段

自 20 世纪 60 年代末期以来, 计算机管理的对象规模越来越大, 应用范围越来越广, 数据量激增, 多种语言、多种应用互相覆盖的共享数据集合的要求越来越强烈。这时计算机硬件价格在下降, 软件价格在上升, 为编制和维护系统软件及应用程序所需的成本相对增加。文件系统作为数据管理手段已经不能满足应用的需求。为解决多用户、多应用共享数据的需求, 使数据为尽可能多的应用服务, 数据库技术应运而生并不断发展壮大。从文件系统到数据库系统标志着数据管理技术的飞跃。对数据库系统特点的详细介绍参见 1.1.2 节。

(1) 20 世纪 60 年代末至 20 世纪 70 年代

20 世纪 60 年代末硬盘的广泛使用改变了数据处理的方式, 硬盘允许直接对数据进行访问, 数据摆脱了磁带和卡片组顺序访问的限制。人们可以创建网状数据库和层次型数据库, 可以将表和树这样的数据结构保存在磁盘上。程序员可以构建和操作这些数据结构。

1970 年 Edgar F. Codd 发表了一篇具有里程碑意义的论文, 定义了关系模型以及在关系模型中查询数据的非过程化方法, 由此诞生了关系型数据库。

(2) 20 世纪 80 年代

尽管关系模型简单, 能够对程序员屏蔽所有实现的细节, 在学术界备受重视, 但它最初被认为性能不好, 没有实际的应用价值。关系型数据库在性能上还不能和当时已有的网状数据库和层次型数据库相提并论。这种情况直到 IBM 的 System R 出现才得以改变。与此同时, 加州大学伯克利分校(University of California, Berkeley)专家 Michael Stonebraker 主持开发了 INGRES(interactive graphics and retrieval system)系统, 它后来发展成商品化的关系数据库系统。

20 世纪 80 年代初期, 关系型数据库已经可以在性能上与网状数据库和层次型数据库进行竞争。关系数据库凭借其简单易用, 最后完全取代了网状数据库和层次型数据库。

在这个时期, 人们还对并行数据库和分布式数据库进行了很多探究, 对面向对象数据库也有初步的研究。

(3) 20 世纪 90 年代

许多数据库厂商推出了并行数据库产品, 并开始在数据库中加入对“对象-关系”的支持。决策支持和查询再度成为数据库的一个主要应用领域。分析大容量数据的工具有了长足进步。

随着互联网的爆炸式发展, 数据库技术有了更加广泛的应用。在这个阶段, 数据库系统必须支持高速的事务处理, 支持对数据的 Web 接口, 而且还要有很高的可靠性和 7×24 小时的可用性。

(4) 21 世纪第一个十年

万维网联盟(World Wide Web Consortium, W3C)于 1998 年 2 月推荐的 XML(eXtensible Markup Language)兴起, 与之相关联的 XQuery 查询语言成为新的数据库技术。虽然 XML 被广泛应用于数据交换和一些复杂数据类型的存储, 但关系数据库仍然是构成大多数大型数据库应用系统的核心。为减少系统管理开销的自主计算/自动管理技术得到了成长; 开源数据库系统(如 PostgreSQL 和 MySQL)的应用也显著增长。

用于数据分析的专门的数据库增速惊人, 特别是将一张二维表的每列高效地存储为一个单独的数组的列存储, 以及为非常大的数据集的分析而设计的高度并行的数据库系统。有几个新颖的分布式数据存储系统被构建出来, 以应对庞大的 Web 节点(如 Amazon、Facebook、Google、Microsoft 和 Yahoo!)的数据管理需求。数据挖掘技术被广泛部署应用, 如基于 Web 的产品推荐系统和 Web 页面上的相关广告自动布放。在管理和分析流数据(如股票市场报价数据、计算机网络监测数据)方面也取得重要进展。

(5) 21 世纪第二个十年

以互联网大数据应用为背景发展起来的分布式非关系型的数据管理系统 NoSQL(Not Only SQL), 融合了 NoSQL 系统和传统数据库事务管理功能的 NewSQL, 分析型 NoSQL 技术的主要代表 MapReduce 技术……各类技术的互相借鉴、融合和发展, 成为数据管理技术的发展趋势。

国际电信联盟(International Telecommunication Union, ITU)于 2017 年 9 月发布 *The State of Broadband 2017*(《2017 年宽带状况》)报告: 现在全球已有 48% 的人使用网络, 估计到 2017 年底, 发展中国家的互联网普及率将达到 41.3%, 中国互联网用户人数仍位居全球第一。我们每个人既是大数据信息的接受者也是产生者, 大数据通常被认为是 PB 或更高数量级的数据($1\text{YB}=1024\text{ZB}=1024^8=2^{80}$ 字节, $1\text{ZB}=1024\text{EB}=1024^7=2^{70}$ 字节, $1\text{EB}=1024\text{PB}=1024^6=2^{60}$ 字节, $1\text{PB}=1024\text{TB}=1024^5=2^{50}$ 字节, $1\text{TB}=1024\text{GB}=1024^4=2^{40}$ 字节, $1\text{GB}=1024\text{MB}=1024^3=2^{30}$ 字节, $1\text{MB}=1024\text{KB}=1024^2=2^{20}$ 字节, $1\text{KB}=1024\text{Byte}=1024=2^{10}$ 字节)。大数据已经从概念落到实地, 具有体量大、结构多样、时效性强的特点。大数据处理促进了新型计算架构和智能算法的云计算的发展, 也将已诞生 60 多年的人工智能(Artificial Intelligence, AI)带入新阶段。

2016 年 10 月美国总统行政办公室(Executive Office of the President)联合美国国家科学技术委员会(National Science and Technology Council, NSTC)共同发布了 *Preparing For The Future of Artificial Intelligence* 研究报告, 对人工智能的发展现状、应用领域、目前存在的问题进行了阐述, 并向美国政府及相关机构提出了相应的发展建议和对策。

2017 年 7 月 20 日国务院印发《新一代人工智能发展规划》, 这是我国第一部在人工智能领域进行系统部署的规划, 提出了面向 2030 年我国人工智能的发展规划。

在大数据、移动互联网、超级计算、传感网、脑科学等新理论新技术以及经济社会发展强烈需求的共同驱动下, 人工智能呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征。新一代人工智能相关学科发展、理论建模、技术创新、软硬件升级等整体推进, 正在引发链式突破, 推动经济社会各领域从数字化、网络化向智能化加速跃升。人工智能的发展也有不确定性。这种影响面广泛的颠覆性技术, 可能会改变我们的就业结构、冲击现有的法律与社会伦理、侵犯个人隐私、挑战国际关系准则等。

迄今图灵奖唯一的华人获奖者、清华大学姚期智教授(Andrew Chi-Chih Yao)在2017年8月公开表示:过去十年的人工智能发展,在算法、计算力和大数据三要素中的后两者已达到一个相当惊人的地步。这三个核心要素中,最有可能在下一波人工智能浪潮中跃迁的是算法。因为数据量归根结底还是有限的,计算力也会有一个差不多的极限,但算法的空间还很大。

4. 数据管理在中国

欧洲文艺复兴之后,西方国家在科学领域占据垄断性优势。近代中国战事连绵不断,国力羸弱。新中国成立后的1956年,周恩来(1898.03.05—1976.01.08)总理领导制定《1956—1967年科学技术发展远景规划纲要》,同年8月成立中国科学院计算技术研究所筹委会,由1950年回国的数学家华罗庚(1910.11.12—1985.06.12)先生任主任委员。在独立自主、自力更生思想指导下,1958年我国第一台电子管计算机103机诞生,1964年以后我国开始推出一批晶体管计算机,代表机型有中科院计算所的“109乙”“109丙”,15所和738厂的“108乙”和320机,军事工程学院的411B等。

20世纪70年代末,以萨师煊(1922.12.27—2010.07.11)先生为代表的老一辈科学家以强烈的责任心和敏锐的学术洞察力,率先在国内开展数据库技术的教学与研究工作。萨师煊先生生于福建福州,我国数据库学科的奠基人之一,数据库学术活动的积极倡导者和组织者。1978年首次在中国人民大学开设数据库课程,之后又将自己的讲稿汇集成《数据库系统简介》和《数据库方法》,在当时的《电子计算机参考资料》上公开发表供他人免费使用。这是我国最早的数据库学术论文。随后,他又发表了许多涉及数据模型、数据库设计、数据库管理系统实现和关系数据库理论等诸多方面的论文。1983年,他与弟子王珊合作编写国内第一部系统阐明数据库理论和技术的教材《数据库系统概论》,2014年已出版第五版。

近些年,参加国际数据库领域三大顶级会议ACM SIGMOD/PODS、VLDB和ICDE的华人越来越多。非官方的大型千人活动“中国数据库技术大会”已举办八届,从每届年会的主题可洞见我国数据管理的关注点。2010年第一届大会的主题是“数据库与商业智能企业应用最佳实践”,第二届主题是“数据库架构设计”,第三届主题是“数据库技术企业应用最佳实践”,第四届主题是“大数据·数据库架构与优化·数据治理与分析”,第五届、第六届主题是“大数据技术探索和价值发现”,第七届主题是“数据定义未来”,2017年第八届大会以“数据驱动·价值发现”为主题,汇集来自政府、互联网、电子商务、金融、电信、行业协会等20多个领域的技术专家,共同探讨Oracle、MySQL、NoSQL、云端数据库、智能数据平台、区块链、数据可视化、深度学习等领域的前瞻性热点话题与技术。

1.1.2 基本术语

数据、数据库、数据库管理系统和数据库系统是与数据库技术密切相关的基本术语。

1. 数据(data)

当今世界,数据正在以前所未有的速度成为各个领域价值创造的核心驱动力。数据是数据库中存储的基本对象,是描述事物的符号记录。早期的计算机系统主要用于科学计算,处理的数据是整数、实数等数值型数据;现代计算机系统存储和处理的对象十分广泛,除了数值型数据,还可以是非数值型数据,例如文本、音频、视频、图形和图像等。非数值型数据经过数字化后存入计算机。

日常生活中,人们通常采用无结构的自然语言描述事物。例如,一个学生的基本情况

可描述为：陈榕刚，男，团员，2000年3月12日出生，福建生源，专业编号 M01 等。

日常数据管理中，学生的基本情况通常如表 1-2 所示。

表 1-2 学生基本情况表

姓名	性别	政治面貌	出生日期	生源地	专业编号
陈榕刚	男	团员	2000 年 3 月 12 日	福建	M01
⋮	⋮	⋮	⋮	⋮	⋮

数据与其语义是不可分的。数据的表现形式需要经过解释才能完全表达其内容，数据的含义(即语义)需要经过解释才能被正确理解。例如，“2000年3月12日”这个数据可能是某人的生日，也可能是卒日，还可能是某商品的出厂日期等。在表 1-2 中，其语义已由其所所在列的表头栏目名称解释，即出生日期。

记录是计算机中表示和存储数据的一种格式或一种方法。将一个学生的姓名、性别、政治面貌、出生日期、生源地和专业编号等数据组织在一起便构成一条记录，用于描述一个学生的基本情况。这样的数据是有结构的，表格描述的数据称为结构化数据。

数据不能等同信息(information)。信息论奠基人 C. E. Shannon(1916.04.30—2001.02.24)从通信理论出发，用数学方法定义信息。Shannon 认为：信息是用来消除随机不确定性的东西。控制论创始人 Norbert Wiener(1894.11.26—1964.03.18)认为：信息是人们在适应外部世界，并使这种适应反作用于外部世界的过程中，同外部世界进行互相交换的内容名称。我国信息学专家钟义信教授认为：信息是事物的存在方式或运动状态，以这种方式或状态进行直接或间接的表述。美国信息资源管理专家 F. W. Horton 认为：信息是为了满足用户决策的需要而经过加工处理的数据。人们从各自研究的角度出发，对信息有不同的理解和定义，随着时间的推移，信息也将被赋予新的含义。我们可以认为：信息以数据为载体，是具有一定含义的、经过加工处理的数据，是客观事物存在方式和运动状态的反映，对人类决策有帮助和有价值的信息。例如，我们经常通过天气预报了解天气情况，气象台主要依据事先勘测采集的一组气象数据，如气压、云层、温度、湿度、风力等进行整理加工和综合分析(与经验值比较、统计、运算、推断)，从而预报天气的阴、晴、刮风、下雨、下雪等。这里的气压、云层、温度、湿度、风力等称为数据，经过整理加工和综合分析的结果(天气预报)称为信息。

2. 数据库(DataBase, DB)

数据库从字面上可以理解为存放大量“数据”的“仓库”，只不过这个仓库遵循一定的数学理论创建，位于计算机存储设备中，数据按特定的格式存放。

数据库是长期存储在计算机内，有组织且可共享的大量数据的集合。数据库中的数据按一定的数据模型组织、描述和存储，具有较小的冗余度、较高的数据独立性和易扩展性。数据库中不仅存放数据，而且存放数据与数据之间的联系。

随着 DB 的发展，出现了数据仓库(Data Warehouse, DW)。数据仓库之父 William H. (Bill) Inmon 在经典论著 *Building the Data Warehouse* 第 4 版中将数据仓库定义为“A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions.” (DW 是一个面向主题、集成、非易失性和随时间变化

的集合,用于支持管理层的决策。)

3. 数据库管理系统(DataBase Management System, DBMS)

DBMS 是数据库系统的核心组成部分,是位于用户与操作系统之间的一层数据管理软件,它提供一个可以方便且高效地存取、管理和控制数据库信息的环境。当今主流的数据库管理系统是关系数据库管理系统(Relational DataBase Management System, RDBMS)。常见的 DBMS 有 Oracle、SQL Server、DB2、Informix、Sybase、PostgreSQL 和 MySQL 等。Access 是一个基于 Windows 的桌面关系数据库管理系统。

DBMS 的主要功能:

(1) 数据定义、组织、存储和管理功能

DBMS 为用户提供了数据定义语言(Data Definition Language, DDL),方便对数据库中数据对象的组成和结构进行定义。DBMS 要确定以何种文件结构和存取方式在存储级别上组织数据,以提高存取效率。

(2) 数据操纵功能

DBMS 还为用户提供了数据操纵语言(Data Manipulation Language, DML)以操纵数据,实现对数据库的查询、插入、删除和修改等操作。

DDL 和 DML 不是两种分离的语言,它们简单地构成了单一的数据库语言的不同部分,例如,关系数据库的标准语言 SQL(Structured Query Language)。

(3) 数据库的建立和维护功能

数据库的建立、运用和维护由 DBMS 统一管理和控制,以保证事务(transaction)的正确运行、数据的安全与完整、多用户对数据的并发使用、发生故障后的系统恢复等。

对于用户而言,事务是具有完整逻辑意义的数据库操作序列的集合。对于 DBMS 而言,事务是一个读写操作序列,这些操作要么都执行,要么都不执行,是一个不可分割的逻辑工作单元。为保证事务能安全并发执行,事务具有原子性(Atomic)、一致性(Consistency)、隔离性(Isolation)和持久性(Durability)四个特性,统称为 ACID 特性。在关系数据库中,一个事务可以是一条或一组 SQL 语句。

4. 数据库系统(DataBase System, DBS)

DBS 是由数据库(DB)、数据库管理系统(DBMS)、数据库应用系统和用户组成的存储、管理、处理和维护数据的系统。DBS 的组成如图 1-3 所示。

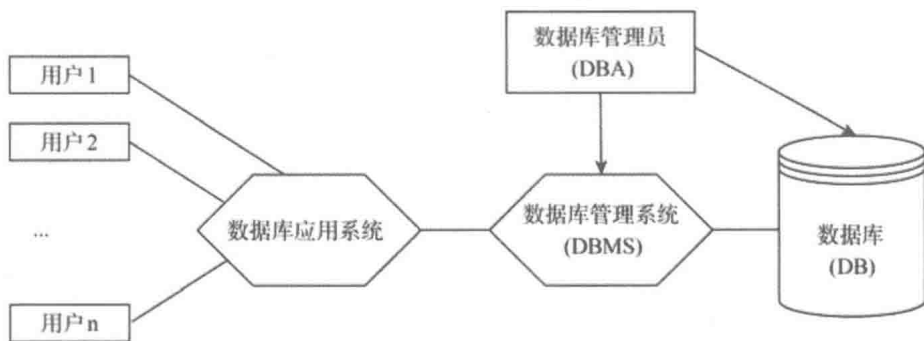


图 1-3 数据库系统的组成

DBS 中的用户有终端用户、应用程序员、系统分析员、数据库设计人员和数据库管理员等多种, 分别承担不同的任务。

- 终端用户通过应用系统的用户接口使用数据库。常用的接口方式有浏览器、菜单驱动、表格操作、报表书写和图形显示等。
- 应用程序员负责设计和编写应用系统的程序模块, 并进行调试和安装。
- 系统分析员负责应用系统的需求分析和规范说明, 确定系统的硬软件配置, 并参与数据库系统的概要设计。
- 数据库设计人员负责数据库中数据的确定、数据库中各级模式的设计。
- 数据库管理员(DataBase Administrator, DBA)负责全面管理和控制数据库系统, 其主要职责是: 决定数据库中的信息内容和结构、决定数据库的存储结构和存取策略、定义数据的安全性要求和完整性约束条件、监控数据库的使用和运行、推动数据库的改进和重组重构等。

DBS 的主要特点:

(1) 数据的共享性高、冗余度低

共享是指数据库中的相关数据能够被多个不同的用户使用, 这些用户可以存取同一种数据并将它用于不同的目的。

数据共享可以大大减少数据冗余, 节约存储空间, 还能避免数据之间的不相容性和不一致性。

数据冗余是指同一数据在数据库中存储了多个副本, 它可能引起如下问题。

- 冗余存储: 相同数据被重复存储, 导致浪费大量存储空间。
- 更新异常: 若重复数据的一个副本被修改, 所有副本都必须同时进行同样的修改。因此在更新数据时, 为了维护数据库的完整性, 系统要付出很大的代价, 否则有可能发生数据不一致的危险。
- 删除异常: 删除某些数据时可能丢失其他与之有关联的数据。
- 插入异常: 只有当一些数据事先已存放在数据库中时, 另一些数据才能存入该数据库中。

(2) 数据独立性高

数据独立性是数据与程序间的互不依赖性, 即数据库中的数据独立于应用程序而不依赖于应用程序。

数据独立性一般分为逻辑独立性和物理独立性两种。

逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。当数据的逻辑结构改变时, 用户的应用程序可以不改变。

物理独立性是指用户的应用程序与数据库中数据的物理存储是相互独立的。DBMS 负责管理数据库中数据的存储, 用户编写应用程序时只需要处理数据的逻辑结构, 当数据的物理存储改变时, 用户的应用程序不用改变。

(3) 数据有安全性保护、完整性检查

数据的安全性是指保护数据以防止非法使用造成的数据泄密和破坏。每个用户只能按规定对某些数据以某些方式进行使用和处理。

数据的完整性是指数据的正确性、有效性和相容性。完整性检查将数据控制在有效的