前　言

数据库技术不但是计算机科学技术中发展最快的领域之一，也是应用最广的技术之一。数据库管理系统是国家信息基础设施的重要组成部分，是国家信息安全的核心技术之一。数据库系统已成为计算机信息系统与应用系统的核心技术和重要基础。设置这门课程的主要任务是介绍关系数据库的基本概念、数据库标准语言 SQL、数据库的安全性、完整性恢复技术和并发控制，数据库的模仿模型与 E-R 方法、关系数据库设计理论与设计步骤等，从而提高学生对数据库的认识以及在实际中使用数据库的能力。在学习的过程中要特别强调理论和实践的结合，注重学生的实践动手能力，使学生能够在理论学习的过程中把所学的知识应用到实践中去。

目前市场上的同类图书比较多，其主要特点就是：理论知识太强，实践比较少；或者是实践也不少，但是能把理论知识合理地糅和到实践中的教材非常少，往往是理论和实际脱离，不适合应用型本科学生使用。编者通过对部分高校的课程体系进行研究发现，目前本科院校都注重实践能力的培养，而目前的系列教材在实践方面做得并不是特别适用这类高校的学生。所以教育改革的目标是建立一整套与技术应用型人才培养目标一致的教学内容和课程体系。其教学内容在基础理论上要求以应用为目的，以必需、够用为度，在专业内容上要突出针对性和实用性。本教材着重于被广泛采用的较成熟的先进技术，总结了编者长期从事有关教学与科研的经验与体会。

全书共计 12 章。第 1 章数据库基础知识主要介绍数据库系统中各种基本概念、数据库系统的组成与基本模型，使初学者对这门课有一基本认识；第 2 章关系数据库主要介绍关系数据库结构与关系数据库基本操作；第 3 章关系数据库语言 SQL 主要介绍关系数据库标准语言 SQL 的基本内容和使用；第 4 章数据库的安全性与完整性主要介绍数据库安全性与完整性基本概念和相关技术；第 5 章数据库的事务处理与数据恢复主要介绍数据库事务概念与性质，以及事务在并发执行和故障处理、故障恢复的基本应用；第 6 章关系数据库的规范化理论主要介绍关系数据库设计规范化理论；第 7 章数据库设计主要介绍关系数据库在设计过程中的各个环节；第 8 章 SQL Server 2000 关系数据库管理系统主要介绍 SQL Server 2000 的安装，以及主要的数据库操作；第 9 章～第 12 章主要介绍数据库的一些新技术。

本书由梁树军，张建伟任主编，卢中宁任副主编，参加本书编写的有李志强、韩恺、金松河、钱慎一、张保威。其中，第 1、2、7 章由梁树军编写，第 3 章由李志强编写，第 4 章由韩恺、张保威编写，第 5 章由金松河编写，第 6 章由钱慎一编写，第 8 章由韩恺编写，第 9、10 章由卢中宁编写，第 11 章由张建伟编写，第 12 章由金松河、李志强编写。全书由梁树军统稿并定稿，由张建伟教授主审。在本书的编写和出版过程中得到了郑州轻工业学院教务处、中国电力出版社的大力支持和帮助，在此由衷向他们表示感谢!

需要全本请访问 www.ertongbook.com

本书除了可用作高等院校本、专科学生的教材外，也可作为从事计算机应用开发人员在学习数据库技术时的参考书。

由于编写时间仓促，水平有限，本书难免有错误之处，恳请广大读者批评指证。

编　者

2007 年 8 月

目　　录

前　言
第 1 章　数据库基础知识 1
1.1　数据库基本概念 1
1.2　数据库系统 9
1.3　数据模型 17
小结 27
习题 28
第 2 章　关系数据库 29
2.1　关系数据库系统概述 29
2.2　关系数据结构 31
2.3　关系的完整性 35
2.4　关系模型数学理论——关系代数 36
2.5　关系模型数学理论——关系演算 42
2.6　优化查询 44
小结 46
习题 47
第 3 章　关系数据库语言 SQL 48
3.1　SQL 概述 48
3.2　数据定义 50
3.3　数据查询 56
3.4　数据更新 79
3.5　视图 82
小结 85
习题 86
第 4 章　数据库的安全性与完整性 87
4.1　数据库安全性概述 87
4.2　数据库安全性控制 88
4.3　数据库完整性控制机制 96
小结 115
习题 116
第 5 章　数据库的事务处理与数据恢复 117
5.1　事务管理的基本概念 117

5.2 并发控制 119
5.3 数据库恢复 131
小结 138
习题 138
第 6 章 关系数据库的规范化理论 139
6.1 概述 139
6.2 函数依赖 141
6.3 关系模式的规范化 143
6.4 数据依赖的公理系统——ARMSTRONG 公理系统 151
6.5 关系模式分解 153
小结 157
习题 158
第 7 章 数据库设计 159
7.1 数据库设计概述 159
7.2 需求分析 160
7.3 概念结构设计 164
7.4 逻辑结构设计 167
7.5 数据库的物理设计 169
7.6 数据库实施 171
7.7 数据库运行与维护 172
小结 173
习题 173
第 8 章 SQL Server 2000 关系数据库管理系统 174
8.1 SQL Server 2000 简介 174
8.2 SQL Server 2000 的安装 174
8.3 SQL Server 2000 的管理工具 178
8.4 SQL Server 2000 数据库对象的操作 182
8.5 SQL Server 2000 数据库备份 191
8.6 SQL Server 2000 数据库恢复 193
8.7 SQL Server 2000 数据库传输 195
8.8 SQL Server 2000 安全性管理 199
小结 205
习题 206
第 9 章 多媒体数据库与空间数据库 208
9.1 多媒体数据库 208
9.2 空间数据库 214
小结 223
习题 223

第 10 章　Web 数据库与分布式数据库 224
10.1　概述 224
10.2　Web 数据库 225
10.3　Web 数据库访问技术 227
10.4　分布式数据库 230
小结 240
习题 241
第 11 章　决策支持系统和数据仓库 242
11.1　决策支持系统 242
11.2　数据仓库 248
11.3　数据挖掘 257
小结 261
习题 261
第 12 章　面向对象数据库 262
12.1　面向对象概念与数据库系统 262
12.2　面向对象的基本思想 266
12.3　面向对象数据模型 267
12.4　面向对象数据库管理系统 271
12.5　对象—关系数据库 274
小结 275
习题 276
参考文献 277

需要全本请访问 www.ertongbook.com

第 1 章　数据库基础知识

数据库技术已经成为计算机科学的一个重要分支，它不但是数据管理的最新技术，而且还是计算机科学技术中发展最快的领域之一。许多信息系统都是以数据库为基础建立的，它已成为计算机信息系统的核心技术和重要基础。数据库已经成为人们储存数据、管理信息、共享资源的最先进最常用的技术。

本章主要介绍数据库系统的基本概念，包括数据管理技术的发展过程、数据库系统的基本概念、数据模型及数据库系统的体系结构等。读者从中可以学习到为什么要使用数据库技术以及数据库技术的重要性，本章是学习后面各章节的预备和基础。

1.1　数据库基本概念

在系统地学习数据库技术之前，首先来了解一下数据库中最常用的几个基本概念。

1.1.1　数据与信息

1. 数据

（1）数据的概念。数据（data）是数据库中最基本的存储对象，是现实世界中客观存在的物体在计算机中的抽象表示，是存储在计算机中的符号串。人们对数据的理解大多是狭义的，也就是数字。例如：55、2666、-23、￥458 等，广义地讲，文字、图形、图像、声音、符号等有意义的元素都称为数据。

（2）数据的特性。数据有多种特性，其主要特性如下。

① 表现多样性。数据可以有多种表现形式，除狭义上的数字以外，数据还可以是文字、图形、图像、声音、视频等多种表现形式。正是由于数据表现形式的多样性，才能为数据库的广泛应用提供了有力的基础。

② 数据的可构造性。从结构上看，数据分为结构化数据、半结构化数据和非结构化数据。不同的应用中数据的结构也是不同的，比如互联网中的 Web 数据属于非结构化及半结构化形式。而软件中一般使用的都是结构化数据。结构化数据有型（type）与值（value）之分，数据的型给出了数据的类型，如整型、实型、字符型等；而数据的值是指符合给定类型的数值，比如数据库中是这样来描述数据的：

（2007001，王伟，男，18，计算机系）

即把学生的学号、姓名、性别、年龄、系别组织起来，形成一个记录，这条学生的记录就是描述一个学生的数据，这就是一种结构化的数据，（学号、姓名、性别、年龄、系别）是数据的型，（2007001，王伟，男，18，计算机系）是数据的值。

③ 数据的持久性与挥发性。长期对系统有用的或者需长期保存的数据一般是存放在计算

机的外部存储器（如光盘、硬盘）中，这样的数据称为持久保存的数据，具有持久性。另一部分数据与程序仅有短时间的交互关系，随着程序的结束而消亡，它们一般保存在计算机的内存中，这样的数据称为临时性数据或挥发性数据。

④ 数据的私有性与共享性。从数据的服务对象上看，数据可分为私有数据和共享数据两种。为特定应用程序服务的数据称为私有数据，而为多个应用程序服务的数据称为共享数据。

⑤ 数据的海量存储性。从数据的存储数量上看，数据可分为不同的层次。随着计算机技术的不断发展，计算机硬件的存储容量也在不断增长，这就使得数据的存储量也越来越大，出现了海量存储。

2. 信息

信息就是新的、有用的事实和知识，是现实世界各种事物的特征、形态以及不同事物之间的联系等在人脑海中的抽象反映。对这些经抽象而形成的概念，人们可以认识理解，可以加工传播，可以进行推理，从而达到认识世界、改造世界的目的。特别是在信息技术高速发展的今天，信息已经越来越重要了。例如：交通部门通知“中央大道堵车比较严重，请司机朋友绕行”，司机朋友听到这个信息后就绕道行驶，避免发生堵车现象。现在是市场经济，经销商经过市场调研后，对市场中的一些信息进行分析整理，决定是否减产或改变销售模式等。这些都是信息作用的一种体现，也充分说明了信息在人类社会中有着非常重要的地位。

信息具有四个基本特征：

（1）信息的内容是关于客观事物或思想方面的知识。

（2）信息是有用的，它是人们活动的必需知识。

（3）信息能够在空间上和时间上被传递，在空间上传递信息称为信息通信，在时间上传递信息称为信息存储。

（4）信息需要一定的表示形式，信息与其表现符号不可分离。

3. 信息处理

信息处理是对信息进行收集、保存、传输、加工和利用的过程，最终体现为对数据的处理。除暗示外，几乎所有的信息都可表示为计算机能识别的字符串或位串，这为信息的传播提供了可能，从而使人们对信息、数据的利用和处理实现自动化、网络化和社会化，使整个人类社会成为信息化社会。

4. 数据处理

不是任何信息都是有用的，数据处理是对各种形式的数据进行收集、储存、加工和传播的一系列活动的总和。其目的是从大量原始数据中抽取、推导出有价值的信息作为行动和决策的依据。数据处理有三个阶段：

（1）手工处理阶段（原始社会至 19 世纪）

工具：算盘、计算尺、微分机、基于齿轮的六位加法器等。

特点：计算工具极低级、精度差、处理能力低、离不开手工。

（2）机械处理阶段（1890—1946 年）

工具：机械计算工具（卡片制表机）。

特点：机械代替了手工操作，数据处理能力、效率有所提高，但受机械性能限制。

（3）电子处理阶段（1946 年至今）

工具：计算机。

特点：速度快、存储量大、传输速度快、逻辑推理能力强、不疲劳、可靠性好、精度高，不但能进行科学计算而且还能进行数据处理。

5. 数据管理

数据管理是指对数据进行分类、组织、编码、存储、检索和维护，它是数据处理的中心问题。在应用需求的推动下，在计算机硬件、软件发展的基础上，它经历了人工管理、文件系统、数据库系统三个阶段。

现在，任何一个企业、单位每天都需要处理大量的数据，而且这些数据的处理都离不开计算机。企业的成功都不能离开准确、及时地获取有关业务的运行数据，数据不仅是管理者进行管理、决策的重要依据，而且数据本身也是被管理的资源。

数据的处理通常比较简单，而数据的管理则比较复杂。因为当今社会是一个信息社会，数据量呈爆炸性增长，而且数据的种类也在不断增多。作为一个管理者，不但要使用数据，而且最重要的是管理数据，数据管理的好坏直接影响企业决策的成功与否。

1.1.2　数据库、数据库管理系统与数据库系统

1. 数据库（Database，DB）

顾名思义，数据库是存放数据的仓库。只不过这个仓库是在计算机的存储设备上，而且数据是按照一定的数据模型组织并存放在外存上的一组相关数据集合，通常这些数据是面向一个组织、企业或部门的。例如学生成绩管理系统中，学生的基本信息、课程信息、成绩信息等都是来自学生成绩管理数据库的。

除了用户可以直接使用的数据外，还有另外一种数据，它们是有关数据库的定义信息的，如数据库的名称、表的定义、数据库用户名及密码、权限等。这些数据用户不会经常性的使用，但是对数据库来说它们非常重要。这些数据通常存放在一个“数据字典（data dictionary）”中。数据字典是数据库管理系统中非常重要的组成部分之一，它是由数据库管理系统自动生成并维护的一组表和视图。数据字典是数据库管理系统工作的依据，数据库管理系统借助于数据字典来理解数据库中数据的组织，并完成对数据库中数据的管理与维护。数据库用户可通过数据字典获取有用的信息，如用户创建了哪些数据库对象，这些对象是如何定义的，这些对象允许哪些用户使用等。但是，数据库用户是不能随便改动数据字典中的内容的。

人们收集并抽取出一个应用所需要的大量数据之后，应将其保存起来供进一步查询和加工处理，以获得更多有用的信息。过去人们把数据存放在文件柜里，当数据越来越多时从大量的文件中查找数据显得十分困难。现在人们借助计算机和数据库科学地保存和管理大量复杂的数据，能方便而充分地利用这些宝贵的信息资源。

严格地讲，数据库是长期存储在计算机内，有组织的、大量的、可共享的数据集合。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，并可为各种用户共享。

总之，数据库数据具有永久存储、有组织和可共享 3 个基本特点。

2. 数据库管理系统（Database Management System，DBMS）

建立了数据库之后，下一个问题就是如何科学地组织和存储数据，如何高效地获取和维护

数据，完成这个任务的是一个系统软件——数据库管理系统（DBMS）。

数据库管理系统（DBMS）是指数据库系统中对数据进行管理的软件系统，它是数据库系统的核心组成部分。数据库系统的一切操作，包括查询、更新及各种控制，都是通过 DBMS 进行的。DBMS 总是基于数据模型，因此可以把它看成是某种数据模型在计算机系统上的具体实现。根据所采用数据模型的不同，DBMS 可以分成网状型、层次型、关系型、面向对象型等。但在不同的计算机系统中，由于缺乏统一的标准，即使同种数据模型的 DBMS，它们在用户接口、系统功能等方面也常常是不同的。

如果用户要对数据库进行操作，由 DBMS 把操作从应用程序带到外部级、概念级，再导向内部级，进而操作存储器中的数据。一个 DBMS 的主要目标是使数据作为一种可管理的资源来处理，DBMS 应使数据易于为各种不同的用户所共享，应该增进数据的安全性、完整性及可用性，并提供高度的数据独立性。

3. *数据库系统*（Database System，DBS）

数据库系统是指在计算机系统中引入数据库后的系统，一般由数据库、数据库管理系统（及其开发工具）、应用系统和数据库管理员构成。应该指出的是，数据库的建立、使用和维护等工作只靠一个 DBMS 是远远不够的，还要有专门的人员来完成，这些人被称为数据库管理员（Date Base Administrator，DBA）。

在一般不引起混淆的情况下，人们常常把数据库系统简称为数据库。数据库系统如图 1-1 所示，数据库系统在计算机系统中的地位如图 1-2 所示。

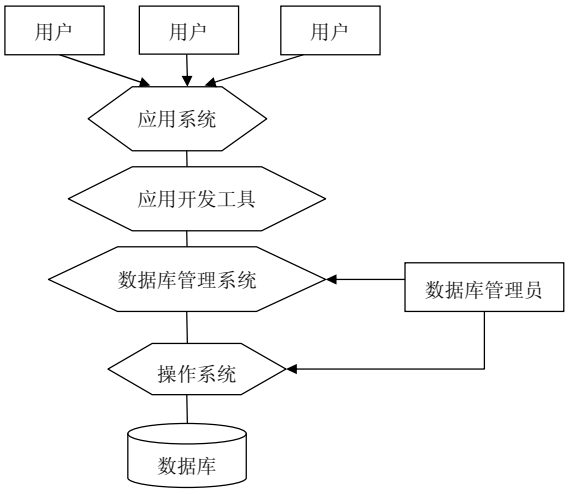


图 1-1　数据库系统

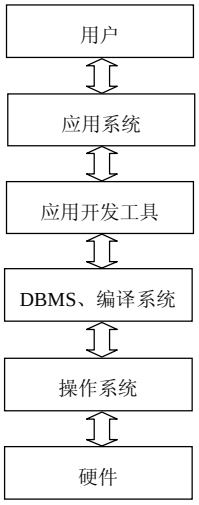


图 1-2　数据库在计算机系统中的地位

1.1.3　数据库技术的产生与发展

使用计算机之后，数据处理的速度及规模都是过去人工或机械方式无法比拟的。随着数据处理量的不断增加，数据管理技术应运而生，其演变过程随着计算机硬件或软件的发展而不断变化，例如以一个学校的教务处对学生、课程和成绩的管理为例，在没有使用计算机的时候，学校教务处的工作人员将有关学生的信息抄写在一张张的卡片上，为了方便查找，将同一个系、

同一个年级、同一个班级的学生的卡片存放在相邻的地方，并对不同的系、年级和班级做上标签。每门课程的信息也是抄写在卡片上，将同一个专业的卡片放在一起并做上标签。每个学期末将同一个班的各门课的成绩单收集起来存放在档案中。

当查找一个学生的信息时，如果知道他所在的系和班级，按照建好的标签可以很快找到该学生的卡片。如果只知道他的姓名，那只有在所有的学生卡片中一个一个查找，需要花费很多时间。

当计算一个学生某个学期的平均成绩时，首先在档案中找到该学生所在班级这个学期的所有成绩单，从中找出该学生各门课程的成绩，再算平均成绩。统计某一门课的成绩分布时也要进行类似的处理和计算。

总的来说，数据库技术的发展经历了人工管理阶段、文件系统阶段、数据库系统阶段和高级数据库阶段。

1. 人工管理阶段

计算机没有应用到数据管理领域之前，数据管理的工作是由人工完成的。这种数据处理方式经历了很长一段时间。

20 世纪 50 年代中期以前，计算机主要用于科学计算。当时的硬件状况是外存只有纸带、卡片、磁带，没有磁盘等直接存取的存储设备；软件状况是没有操作系统，没有管理数据的软件，因此，称这样的数据管理方式为人工管理数据。人工管理数据具有如下特点：

（1）数据不保存。由于当时计算机主要用于科学计算，一般不需要将数据长期保存，只是在计算某一课题时将数据输入，用完就撤走。

如果要用计算机统计分析全校每一门课的成绩，就要编写统计分析程序，在运行该程序时读入相应的学生选修课程成绩单等数据，计算完成后数据和程序都撤走，不再在计算机中保存。

（2）应用程序管理数据。数据需要由应用程序自己管理，没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构，而且要设计物理结构，包括存储结构、存取方法、输入方式等，因此，这增加了程序员的负担。

（3）数据不共享。数据是面向应用的，一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时，由于必须各自定义，无法互相利用、互相参照，因此程序与程序之间有大量的冗余数据。例如教务处要统计分析某一门课的成绩和统计分析某一个学生的成绩，要编写两个程序，尽管都要使用学生选修课程成绩单，但是每个程序要分别定义两个成绩单数据，分别输入，分别使用，如图 1-3 所示。

（4）数据不具有独立性。数据的逻辑结构或物理结构改变后，必须对应用程序做相应的修改，这就进一步加重了程序员的负担。例如，学生成绩由 5 级记分制改为百分制时，上面两个统计分析程序都要修改。

在人工管理阶段，程序与数据之间的一一对应关系如图 1-4 所示。

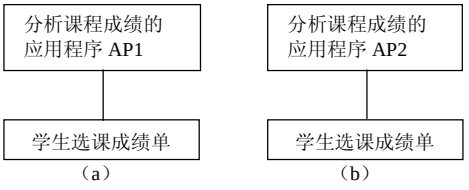


图 1-3　两个应用程序使用同一数据

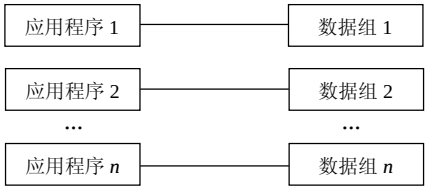


图 1-4　人工管理阶段应用程序与数据之间的对应关系

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期，硬件方面已有了磁盘、磁鼓等直接存取存储设备；软件方面，操作系统中已经有了专门的数据管理软件——文件系统。这时可以把相关的数据组织成一个文件存放在计算机中，在需要的时候只要提供文件名，计算机就能从文件系统中找出所要的文件，把文件中存储的数据提供给用户进行处理。

例如，在学校教务处对学生学籍的管理中，为了改变查找、计算工作量大及花费时间长的被动局面，将学生卡片、课程卡片和学生学习成绩单中的内容存放到文件 student、文件 course、文件 study 中，并对每个文件编写一组程序用于数据维护，包括增加、删除、修改一条记录和查询一条记录。在此基础上根据需要编写一些查询和报表打印程序，例如，根据学生的姓名、学生编号查找学生的信息，统计某学期某个学生的平均成绩，统计某门课的平均成绩等。

管理系统投入应用后，教务处的工作效率大大提高。例如，每学期末将各门课的考试成绩输入计算机以后，可以很快计算出学生的平均成绩，并打印出需要补考的学生的名单。

但是，由于数据的组织仍然是面向程序的，所以存在大量的数据冗余，经过一段时间的使用后，教务人员发现有时必须修改程序和文件结构才能适应工作的需要。例如，学校领导要求统计不同生源地的学生成绩，因为原来的学籍管理软件中没有实现这个功能，所以，必须编写一段程序来实现这个功能。又如，当需要在 student 文件中增加“个人网址”属性时，这涉及到改变文件的结构，需要若干步骤才能完成。

第一步，建立一个新文件 student-new，其结构是在 student 的结构中加入“个人网址”这一项。

第二步，编写一个程序将文件 student 中的数据转存到 student-new 中。

第三步，删除文件 student。

第四步，将文件 student-new 重命名为 student。

这项工作到此并没有结束，因为文件中保存的是数据，不保存数据的结构，数据结构是在程序中定义的，文件 student 的结构写到了所有使用它的程序中，必须一一修改这些程序以适应新的文件结构，否则程序运行时就会出错。

从这个例子可以看到用文件系统管理数据的优点和不足。一般地讲，用文件系统管理数据具有如下特点。

（1）数据可以长期保存。数据可以组织成文件长期保存在计算机中反复使用。

（2）由文件系统管理数据。文件系统把数据组织成内部有结构的记录，实现“按文件名访问，按记录进行存取”的管理技术。

文件系统使应用程序与数据之间有了初步的独立性，程序员不必过多地考虑数据存储的物理细节。例如，文件系统中可以有顺序结构文件、索引结构文件、Hash 等。数据在存储上的不同不会影响程序的处理逻辑。如果数据的存储结构发生改变，应用程序的改变很小，节省了程序的维护工作量。但是，文件系统仍存在以下缺点。

（1）数据共享性差，冗余度大。在文件系统中，一个（或一组）文件基本上对应于一个应用（程序），即文件是面向应用的。当不同的应用（程序）使用部分相同的数据时，也必须建立各自的文件，而不能共享相同的数据。因此数据的冗余度大，浪费存储空间。同时由于相同数据的重复存储、各自管理，容易造成数据的不一致性，给数据的修改和维护带来了困难。

需要全本请访问 www.ertongbook.com

（2）数据独立性差。文件系统中的文件是为某一特定应用服务的，文件的逻辑结构对该应用来说是优化的，因此要相对现有的数据再增加一些新的应用会很困难，系统不容易扩充。一旦数据的逻辑结构发生改变，就必须修改应用程序，修改文件结构的定义。因此数据与程序之间仍缺乏独立性。文件系统阶段程序与数据之间的关系如图 1-5 所示。

3. 数据库系统阶段

20 世纪 60 年代后期以来，计算机用于管理的规模越来越大，应用越来越广泛，数据量急剧增长，同时多种应用、多种语言互相覆盖的共享数据集合的要求也越来越强烈。

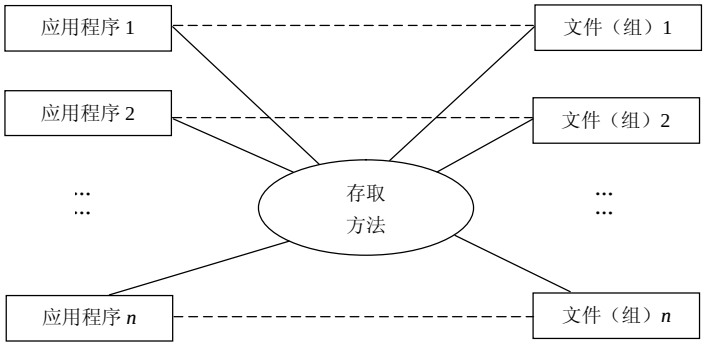


图 1-5　应用程序与数据之间的对应关系图

这时，硬件已有大容量磁盘，硬件价格下降；而软件价格则上升，为编制和维护系统软件及应用程序所需的成本相对增加。在这种背景下，以文件系统作为数据管理手段已经不能满足应用的需求，于是为解决多用户、多应用共享数据的需求，使数据为尽可能多地应用服务，数据库技术便应运而生，出现了统一管理数据的专用软件系统——数据库管理系统。

用数据库系统来管理数据比文件系统具有明显的优点，从文件系统到数据库系统，标志着数据管理技术的飞跃。

例如，对上面的教务管理应用，学校决定采用数据库技术，购买了一个关系数据库管理系统（RDBMS），在这个 RDBMS 之上建立一个应用系统。将教务处和学生工作处保存的学生数据进行综合设计，供全校各院系的教师和教务人员共享访问和使用这些数据。

这样，在系统中建立 3 个关系：学生基本表 student、课程基本表 course 和学习基本表 sc。在数据库系统中只要用 DDL 语言向 RDBMS 提交 CREATE TABLE 语句就可以了，例如建立关系 student：

```
CREATE TABLE student
                (Sno       CHAR(10) NOT  NULL  UNIQUE,
                 Sname     CHAR(10),
                 Sex       CHAR(2),
                 Age       INT,
                 Classno     CHAR(6),
                 Dept      CHAR(12));
```

这条语句在数据库中建立了一个关系 student，用来保存学生的信息，更重要的是 RDBMS 将 student 的结构也保存到数据库的数据字典中。

向关系中增加一个元组（记录）、删除一个元组、修改一个元组相应地用 RDBMS 提供的

语句 INSERT、DELETE、UPDATE 来完成。这些语句如何操作磁盘上的数据是由 RDBMS 来完成的，程序员不用编写专门的程序，从而节省了程序员大量的时间和精力。

因为关系 student 的结构和数据都由 RDBMS 管理，在教务管理应用中，向关系 student 添加“Email”属性时，既不需要编写转储数据的程序，也不用修改那些使用了关系 student 的程序。

从这个例子可以看出，数据库系统阶段应用程序与数据之间的对应关系如图 1-6 所示。

由于数据库是以数据为中心组织数据、减少数据的冗余，提供更高的数据共享能力，同时要求程序和数据具有较高的独立性，当数据的逻辑结构改变时，不涉及数据的物理结构，也不影响应用程序，这样就降低了应用程序研制与维护的费用。

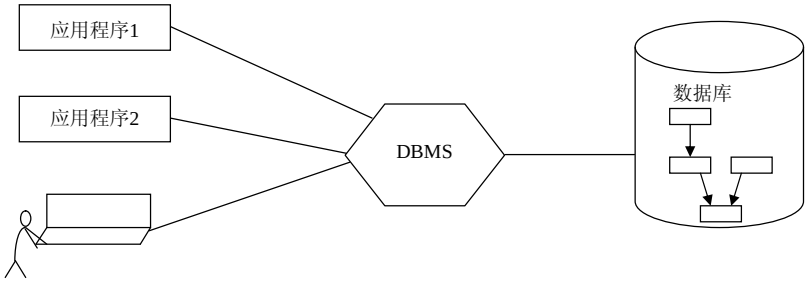


图 1-6　应用程序与数据之间的对应关系图

4. 高级数据库阶段

这一段的主要标志是 20 世纪 80 年代的分布式数据库系统、90 年代的对象数据库系统和 21 世纪初的网络数据库系统的出现。

（1）分布式数据库系统。在这一阶段以前的数据库系统是集中式的。在文件系统阶段，数据分散在各个文件中，文件之间缺乏联系。集中式数据库把数据库集中在一个数据库中进行管理，减少了数据冗余的不一致性，而且数据联系比文件系统强得多。但集中式系统也有弱点：一是随着数据量增加，系统相当庞大，操作复杂，开销大；二是数据集中存储，大量的通信都要通过主机，造成拥挤现象。随着小型计算机和微型计算机的普及、计算机网络软件和远程通信的发展，分布式数据库系统崛起了。

分布式数据库系统主要有以下三个特点：

① 数据库的数据物理上分布在各个场地，但逻辑上是一个整体。

② 各个场地既可以执行局部应用（访问本地 DB），也可以执行全局应用（访问异地 DB）。

③ 各地的计算机由数据通信网络相联系。本地计算机单独不能胜任的处理任务，可以通过通信网络取得其他 DB 和计算机的支持。

分布式数据库系统兼顾了集中管理和分布处理两个方面，因而具有良好的性能。

（2）对象数据库系统。在数据处理领域，关系数据库的使用已相当普遍、相当出色。但是，现实世界存在着许多具有更复杂数据结构的实际应用领域，已有的层次、网状、关系三种数据模型对这些应用领域都显得力不从心。例如多媒体数据、多维表格数据、CAD 数据等应用问题，需要更高级的数据库技术来表达，以便于管理、构造和维护大容量的持久数据，并使它们能与大型复杂程序紧密结合。对象数据库正是适应这种形势发展起来的，它是面向对象的程序设计技术与数据技术结合的产物。

对象数据库系统主要有以下两个特点：

① 对象数据库模型能完整地描述现实世界的数据结构，能表达数据间嵌套、递归的联系。

② 具有面向对象技术的封装性（把数据与操作定义在一起）和继承性（继承数据结构和操作）的特点，提高了软件的可重用性。

（3）网络数据库系统。随着 C/S（客户机/服务器）结构的出现，人们可以最有效地使用计算机资源。但在网络环境中，如何隐藏各种复杂性，这就要使用中间件。中间件是网络环境中保证不同的操作系统、通信协议和 DBMS 之间进行对话、互操作的软件系统。其中涉及到数据访问的中间件，就是 20 世纪 90 年代提出的 ODBC 技术和 JDBC 技术。

现在，计算机网络已成为信息化社会中十分重要的一类基础设施。随着广域网（WAN）的发展，信息高速公路已发展成为采用通信手段将地理位置分散的、各自具备自主功能的若干台计算机和数据库系统有机地连接起来组成因特网（Internet），用于实现通信交往、资源共享或协调工作等目标。这个目标在 20 世纪末已经实现，正在对社会的发展起着极大的推进作用。

1.2 数 据 库 系 统

1.2.1 数据库系统的组成

前面已经介绍了数据库系统一般由数据库、数据库管理系统（及其开发工具）、应用系统和数据库管理员构成。下面分别介绍这几部分的内容。

1. 硬件平台及数据库

硬件系统主要指计算机的各个组成部分。鉴于数据库应用系统的需求，特别要求数据库主机或数据库服务器外存要足够大，I/O 存取效率要高，主机的吞吐量大、作业处理能力强。对于分布式数据库而言，计算机网络也是基础环境。硬件平台主要包括：

（1）要有足够大的内存，存放操作系统和 DBMS 的核心模块、数据库缓冲区和应用程序。

（2）有足够大的磁盘等直接存取设备存放数据库，有足够的光盘、磁盘、磁带等作为数据备份介质。

（3）要求连接系统的网络有较高的数据传送率。

（4）有较强处理能力的中央处理器（CPU）来保证数据处理的速度。

2. 软件

数据库系统的软件主要包括：

（1）DBMS。DBMS 是为数据库的建立、使用和维护配置的软件。

（2）支持 DBMS 运行的操作系统。

（3）与数据库通信的高级程序语言及编译系统。

（4）为特定应用环境开发的数据库应用系统。

3. 数据库管理员及相关人员

数据库有关人员包括数据库管理员（DBA）、系统分析员、应用程序员和普通用户，其各自职责分别如下：

（1）数据库管理员（Database Administrator，DBA）。数据库管理员负责管理和监控数据库系统，负责为用户解决应用中出现的系统问题。为了保证数据库能够高效正常地运行，大型数据库系统都设有专人负责数据库系统的管理和维护。其主要职责如下：

① 决定数据库中的信息内容和结构。数据库中要存放哪些信息，DBA 要参与决策。因此，DBA 必须参加数据库设计的全过程，并与用户、应用程序员、系统分析员密切合作共同协商，做好数据库设计工作。

② 决定数据库的存储结构和存取策略。

③ 监控数据库的运行（系统运行是否正常，系统效率如何），及时处理数据库系统运行过程中出现的问题。例如，系统发生故障时，数据库会因此遭到破坏，DBA 必须在最短的时间内把数据库恢复到正确状态。

④ 安全性管理，通过对系统的权限设置、完整性控制设置来保证系统的安全性。DBA 要负责确定各个用户对数据库的存取权限、数据的保密级别和完整性约束条件。

⑤ 日常维护，如定期对数据库中的数据进行备份、维护日志文件等。

⑥ 对数据库有关文档进行管理。

（2）系统分析员和数据库设计人员。系统分析员负责应用系统的需求分析和规范说明，和用户及 DBA 相结合，确定系统的硬件、软件配置，并参与数据库系统概要设计。

（3）应用程序员。应用程序员是负责设计、开发应用系统功能模块的软件编程人员，他们根据数据库结构编写特定的应用程序，并进行调试和安装。

（4）用户。这里的用户是指最终用户。最终用户通过应用程序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、图形显示、报表等。

1.2.2 数据库系统的结构

数据库系统的结构可以从不同的层次或角度来考察。

从数据库管理系统角度看，数据库系统通常分为三级模式，这是数据库管理系统内部的体系结构。

从数据库最终用户角度看，数据库系统的结构分为单用户结构、主从式结构、分布式结构、客户/服务器结构等。这是数据库系统外部的体系结构。

数据库系统结构是数据库的一个总的框架。尽管实际的数据库系统软件产品多种多样，支持不同的数据模型，使用不同的数据库语言，建立在不同的操作系统之上，但绝大多数数据库系统在总的体系结构上都具有三级模式的结构特征。学习数据库的三级模式将有助于理解数据库设计及应用中的一些基本概念。

数据库的三级模式为外模式、概念模式和内模式，如图 1-7 所示。

1. 数据库的三级模式

（1）概念模式。概念模式也称模式，是对数据库中全局数据逻辑结构的描述，是全体用户公共的数据视图。这种描述是一种抽象描述，不涉及具体硬件环境与平台，也与具体软件环境无关。

概念模式主要描述数据的概念记录类型及其关系，还包括数据间的一些语义约束，对它的描述可用 DBMS 中的 DDL 定义。

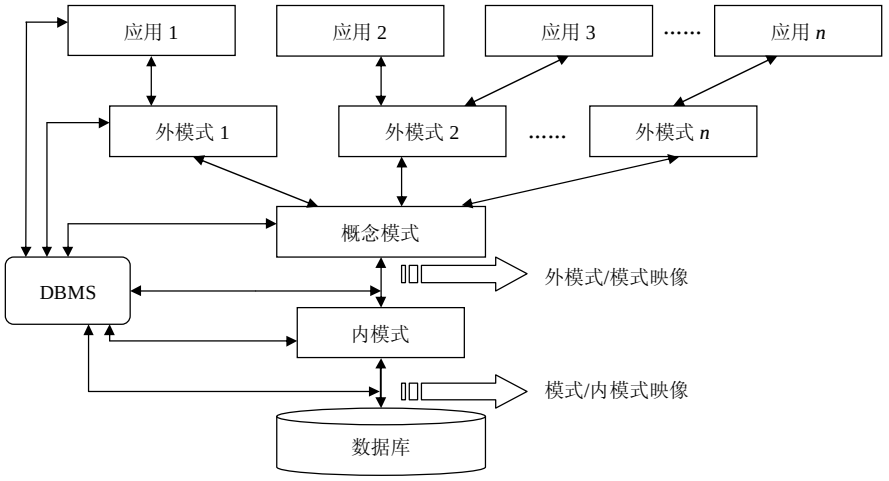


图 1-7　数据库系统结构的三级模式

（2）外模式（external schema）。外模式也称子模式（subschema）或用户模式，它是数据库用户（包括应用程序员和最终用户）能够看见和使用的局部数据的逻辑结构和特征的描述，是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示。

外模式通常是模式的子集，一个模式可以有多个外模式。由于它是各个用户的数据视图，如果不同的用户在应用需求、看待数据的方式、对数据保密的要求等方面存在差异，则其外模式描述就可能不同。即使对模式中的同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同。另外，同一外模式也可以为某一用户的多个应用系统所使用，但一个应用程序只能使用一个外模式。

外模式是保证数据库安全性的一个有力措施，每个用户只能看到和访问所对应的外模式中的数据，数据库中的其他数据是看不到的。

DBMS 提供子模式描述语言（子模式 DDL）来严格地定义子模式。

（3）内模式（internal schema）。内模式也称存储模式（storage schema），一个模式只有一个内模式。它是数据物理结构和存储方式的描述，它定义所有的内部记录类型、索引和文件的组织形式，以及数据控制方面的细节。

内部记录并不涉及到物理记录，也不涉及到设备的约束。比内模式更接近于物理存储和访问的那些软件机制是操作系统的一部分（即文件系统），例如，从磁盘读数据或写数据到磁盘上的操作等。

DBMS 提供内模式描述语言（内模式 DDL，或者存储模式 DDL）来严格地定义内模式。

2. 数据库的二级映像

数据库系统的模式、内模式、外模式之间有很大的差别，为了实现用户和数据之间的透明化，DBMS 提供了外模式/模式映像和模式/内模式映像两层映像。

有了这两层映像，用户就能逻辑抽象地处理数据，而不必关心数据在计算机中的具体表示方式和存储方式。

正是这两层映像保证了数据库系统中的数据能够具有较高的逻辑独立性和物理独立性。

映像实质上是一种对应关系，是指映像双方如何进行数据转换，并定义转换规则。这样就能使数据独立性得到保证。

（1）外模式/模式映像。数据库的每一个外模式都有一个外模式/模式映像，它定义了该外模式与模式之间的对应关系，外模式/模式映像一般是在外模式中描述的。

模式描述的是数据的全局逻辑结构，外模式描述的是数据的局部逻辑结构。对应于同一个模式可以有任意多个外模式。对于每一个外模式，数据库系统都有一个外模式/模式映像，它定义了该外模式与模式之间的对应关系。这些映像通常包含在各自外模式的描述中。

如果模式需要进行修改，例如，数据重新定义、增加新的关系、新的属性、改变属性的数据类型等，那么只需对各个外模式/模式的映像做相应的修改，使外模式尽量保持不变，而应用程序一般是依据外模式编写的，因此应用程序也不必修改，从而保证了数据与程序的逻辑独立性，这就是数据的逻辑独立性。

（2）模式/内模式映像。模式/内模式映像是唯一的，因为数据库只有一个模式和内模式。它存在于模式和内模式之间，由于两级模式之间的数据结构可能不一致，甚至可能差别很大。模式/内模式映像定义了模式和内模式之间的对应关系，即数据全局逻辑结构与存储结构之间的对应关系。模式/内模式映像一般是在模式中描述的，当数据库的存储结构改变时（例如，采用了另外一种存储结构），由数据库管理员对模式/内模式映像进行相应改变，可以使模式保持不变，因此应用程序也不必改变。这就保证了数据与程序的物理独立性，简称数据的物理独立性。

在数据库的三级模式结构中，数据库模式即全局逻辑结构是数据库的中心与关键。它独立于数据库的其他层次，因此设计数据库模式结构应首先确定数据库的逻辑模式。

数据库的内模式依赖于它的全局逻辑结构，但独立于数据库的用户视图即外模式，也独立于具体的存储设备。它是将全局逻辑结构中所定义的数据结构及其联系按照一定的物理存储策略进行组织，以实现达到较好的时间与空间效率的目的。

数据库的外模式面向具体的应用程序，它定义在逻辑模式之上，但独立于存储模式和存储设备。当应用需求发生较大变化，相应外模式不能满足其视图要求时，该外模式就得做相应改动，所以设计外模式时应充分考虑到应用的扩充性。

特定的应用程序是在外模式描述的数据结构上编制的，它依赖于特定的外模式，与数据库的模式和存储结构独立。不同的应用程序有时可以共用同一个外模式。数据库的二级映像保证了数据库外模式的稳定性，从而从底层保证了应用程序的稳定性，除非应用需求本身发生变化，否则应用程序一般不需要修改。

数据与程序之间的独立性使得数据的定义和描述可以从应用程序中分离出去。另外，由于数据的存取由 DBMS 管理，用户不必考虑存取路径等细节，从而简化了应用程序的编制，大大减少了应用程序的维护和修改工作。

1.2.3 数据库管理系统的功能与组成

数据库管理系统是位于用户和操作系统之间的一层数据管理软件。数据库管理系统和操作系统一样是计算机的基础软件，也是一个大型复杂的软件系统。

1. 数据库管理系统（DBMS）的功能

（1）数据定义功能。DBMS 提供数据定义语言（Data Definition Language，DDL）用户通

需要全本请访问 www.ertongbook.com