

高等学校教材

# 数据库原理与应用教程

罗志高 苏恒阳 李智波 编著

人民邮电出版社

## 图书在版编目 (CIP) 数据

数据库原理与应用教程/罗志高, 苏恒阳, 李智波编著. —北京: 人民邮电出版社, 2003.6  
(高等学校教材)

ISBN 7-115-11264-9

I. 数... II. 罗... 苏... 李... III. 数据库系统—教材 IV. TP311.13

中国版本图书馆 CIP 数据核字 (2003) 第 036296 号

### 内 容 提 要

本书全面系统地介绍了数据库系统的基本原理、设计和应用技术。内容包括数据库原理、Visual FoxPro 程序设计、Visual Basic 程序设计、Access 程序设计、Delphi 程序设计五大部分, 以奠定数据库理论基础、培养数据库开发能力为目标, 既讲原理又讲设计与应用, 所介绍的技术以实用为本, 并且反映数据库技术的最新进展。

本书内容循序渐进、深入浅出, 每章均给出了适量的习题, 便于读者巩固所学知识。可作为高等院校计算机类专业、信息管理类专业及相关专业本、专科学生的教材, 也可供从事计算机软件工作的科技人员、工程技术人员参阅。

高等学校教材

### 数据库原理与应用教程

- 
- ◆ 编 著 罗志高 苏恒阳 李智波
  - 责任编辑 姚予疆
  - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
  - 邮编 100061 电子函件 315@ptpress.com.cn
  - 网址 <http://www.ptpress.com.cn>
  - 读者热线 010-67180876
  - 北京汉魂图文设计有限公司制作
  - 印刷厂印刷
  - 新华书店总店北京发行所经销
  
  - ◆ 开本: 787×1092 1/16
  - 印张: 21.5
  - 字数: 524 千字 2003 年 月第 1 版
  - 印数: 1—0 000 册 2003 年 月北京第 1 次印刷

ISBN 7-115-11264-9/TP · 3446

定价: 28.00 元

本书如有印装质量问题, 请与本社联系 电话: (010) 67129223

# 前 言

从多年的数据库软件开发和教学实践体会到，像 Visual FoxPro、Visual Basic、Access、Delphi 这些数据库软件，它们都有相同的数据库基本理论，相似的常量、变量以及表达式，相同的程序设计思路；它们所采用的语法结构也是大同小异，都采用了顺序结构、分支结构、循环结构，都有共同特点和设计技巧，所不同的只是使用的程序设计语言而已。所以在阐述数据库原理后，重点讲述一种数据库设计软件，其它数据库只讲总的程序设计部分和设计技巧，同时强调该数据库与已重点讲述的数据库异同，就可以达到融会贯通、举一反三的目的。

本书是我们的一次尝试，目的是给学习数据库的朋友以启发，总结其规律，提高学习效益。由于我们的水平有限，如何更好地表达我们的意思，还需要我们探索，也希望读者多提宝贵意见，使我们的工作做得更好。

本书是结合作者多年采用计算机技术和网络技术进行实验室管理工作的实际，以及多年来一直从事《数据库及应用》、《计算机管理信息系统与应用》课程教学的心得体会，并融合作者开发的《物资设备管理网络查询系统》、《高校物资管理系统》、《实验室综合网络管理系统》、《仪器设备网上查询系统》、《计算机网络多媒体课件》、《微波技术多媒体课件》、《网上考试系统》等多个教学和管理软件的成果编写而成，内容包括数据库原理、Visual FoxPro 程序设计、Visual Basic 程序设计、Access 程序设计、Delphi 程序设计五大部分，以奠定数据库理论基础、培养数据库开发能力为目标，既讲原理又讲设计与应用，所介绍的技术以实用为本，并且反映数据库技术的最新进展。

本书在编写过程中得到了中山大学电子系系主任、博士生导师秦家银教授的关心和支持，得到了中山大学电子系领导和同事们以及丘仲兴老师、陈盛荣老师等的帮助，在此表示衷心的感谢。由于作者水平有限，书中疏漏之处在所难免，敬请读者批评指正。

编著者

# 目 录

第 1 章 数据库原理知识简介 .....	1
1.1 数据库系统概述 .....	1
1.2 数据库的数据结构 .....	10
1.3 数据模型 .....	12
1.4 数据库的存储结构 .....	16
1.5 关系数据库 .....	19
1.6 关系数据库的实现及使用 .....	22
1.7 计算机网络基础简介 .....	25
1.8 客户机 / 服务器简介 .....	32
1.9 SQL 语言概貌 .....	37
1.10 Visual FoxPro 中的 SQL 查询语句 .....	41
1.11 Access、Visual FoxPro、Visual Basic、Delphi 的共同点 .....	45
习题 .....	45
第 2 章 Visual FoxPro 6.0 数据库程序设计 .....	46
2.1 Visual FoxPro 基础 .....	46
2.2 数据库和表 .....	54
2.3 建立 Visual FoxPro 应用程序 .....	62
2.4 Visual FoxPro 程序设计基础 .....	68
2.5 制作表单 .....	78
2.6 报表设计 .....	87
2.7 设计并使用自定义的类 .....	94
2.8 程序调试 .....	98
2.9 应用程序的改进与完善 .....	102
2.10 Visual FoxPro 在实验室设备管理中的应用实例 .....	106
习题 .....	116
第 3 章 Visual Basic 数据库程序设计 .....	117
3.1 Visual Basic 基本概念 .....	117
3.2 Visual Basic 的开发环境 .....	117
3.3 Visual Basic 程序设计的一般过程 .....	123
3.4 Visual Basic 基本语言 .....	124
3.5 Visual Basic 的数据库编程 .....	131
3.6 Visual Basic 6.0 网络数据库编程 .....	136

3.7	Visual Basic 与 Visual Basic.Net 的区别 .....	152
3.8	Visual Basic 6.0 应用实例 .....	155
	习题 .....	179
<b>第 4 章</b>	<b>Access 2000/2002 数据库程序设计 .....</b>	<b>181</b>
4.1	Access 2000 中文版简介 .....	181
4.2	建立数据库 .....	185
4.3	创建表 .....	187
4.4	查询 .....	190
4.5	报表制作 .....	192
4.6	窗体设计 .....	203
4.7	宏 .....	210
4.8	Access 使用 Visual Basic 编程 .....	218
4.9	Web 发布 .....	242
4.10	OLE (对象链接和嵌入) .....	246
4.11	Access 2000/2002 导入、导出与链接 .....	250
4.12	Access 2000/2002 在网络习题库中的应用 .....	259
	习题 .....	261
<b>第 5 章</b>	<b>Delphi 7.0 数据库程序设计 .....</b>	<b>262</b>
5.1	Delphi 集成开发环境 .....	262
5.2	数据库表的创建 .....	271
5.3	Delphi 7.0 访问数据库的工具 .....	275
5.4	编程实例 .....	321
	习题 .....	334

# 第 1 章 数据库原理知识简介

本章主要介绍数据库和数据库系统的基本概念、数据库技术的发展简介、数据库的结构特点、网络数据库基础和所涉及的计算机网络基础知识。通过对该章的学习，应该对数据库的基本原理有较清晰的认识。

## 1.1 数据库系统概述

### 1.1.1 数据库基本概念

#### 一、信息、数据、数据处理、数据管理

信息是客观世界的事物在人头脑中的反映，是对客观事物状态、特性、特征的描述。数字、文字、图形、声音以及图像是信息的不同表现形式，都是信息的载体，信息就是通过这些载体来传播的。数据是人们用来反映客观世界而记录下来的可以被鉴别的符号，是语言、文字、图形、光、声、色等有意义的组合。这种组合具体地表示出信息的内容，是对事物进行的描述。这里的“符号”不仅仅指数字、字母、文字和其他特殊符号，而且还包括图形、图像、声音等多媒体数据。

信息是现实世界的反映。信息是以数据的形式表示的，即数据是信息的载体，但不是所有的数据都能表示信息，信息是人们消化理解了的数据，同时，信息是抽象的，不随数据设备所决定的数据形式而改变，而数据的表示方式却具有可选择性。

数据处理就是对数据进行收集、整理、存储、检索、统计、维护和传输等一系列活动的总称。数据处理的目的是根据人们的需要，从大量的数据中抽取对于人们来说是特定的有意义、有价值的信息，作为决策和行动的依据。数据管理则是数据处理的中心问题，就是对数据进行分类、组织、编码、存储、检索和维护，是数据处理的中心问题。

数据处理是指将数据转换成信息的过程，如对数据的收集、存储、传播、检索、分类、加工或计算，打印各类报表或输出各种需要的图形。在数据处理的一系列活动中，数据收集、存储、传播、检索、分类等操作是基本环节，这些基本环节统称为数据管理。

#### 二、数据库、数据库系统、数据库管理系统、数据库应用系统

##### 1. 数据库

数据库 (DataBase, DB)，顾名思义，是存放数据的仓库。所谓数据库是长期储存在计算机内的、有组织的、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、描述和储存，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。数据库并不是孤立存在的，它与其他有联系的部分共同组成

数据库系统，人们在实际应用中所面对的通常都是数据库系统。

数据库具有以下几个方面特点：

(1) 共享性。数据共享是数据库系统的目的，是数据统一管理和数据结构化的结果。它包含了以下三个方面的含义：第一，当前所有用户可以存取数据库中的数据；第二，数据易于扩充，这样就可以满足新用户的需求；第三，用户能以灵活的方式，如通过编程语言或终端交互命令来存取、处理数据库中的数据。

(2) 可控冗余度。如果数据被单个用户专用，每个用户都拥有自己的数据，就难免产生数据的重复，这就是数据的冗余。数据冗余不仅造成磁盘空间的浪费，而且也会给管理带来麻烦。实现数据共享后，多个用户共同使用数据库中的数据信息，从而避免了重复，控制了数据的冗余。

(3) 独立性。数据独立性是指数据与应用程序之间的独立，这种独立性分为逻辑独立性和物理独立性，即用户的应用程序和数据的逻辑结构与物理存储方式无关。

(4) 安全性和完整性。数据库系统具有一系列保障数据安全和完整的措施，防止数据的非法使用，以确保数据的正确有效。

## 2. 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整个系统构成，一般由数据库、数据库管理系统（及其开发工具）、应用系统、数据库管理员和用户构成，也就是数据库、计算机硬件系统、计算机软件系统和用户四大部分。数据库系统的四大部分介绍如下：

数据库是结构化的相关数据的集合，是数据库系统管理的对象，它能为多种用户服务并独立于应用程序之外；硬件系统指计算机系统的硬件设备，包括主机、键盘、显示器、软盘驱动器、硬盘、打印机等。复杂的数据库系统一般建立在计算机网络环境下；软件系统能对所有数据库的数据进行管理和维护，是控制管理数据库运行的工具，包括操作系统 (OS, Operating System)、数据库管理系统 (DBMS, DataBase Management System)、高级语言及其编译系统、应用开发工具软件等；用户一般可以分为三种类型：终端用户、应用程序设计员和数据库管理员 (DBA, DataBase Administrator)，也可以分为四种类型：就是数据库管理员、系统分析员、应用程序员和最终用户。

数据库管理员是数据库系统中的最重要的角色，他们决定数据库中数据的内容和结构；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；数据库的改进和重组。系统分析员的主要职责是确定系统的基本功能、数据库的结构和应用程序的设计、软硬件的配置，并组织整个系统的开发。应用程序员则根据系统的功能需求负责设计和编写应用系统的程序模块，并参与对程序模块的调试。用户即数据库的使用者。

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

数据库管理系统是对数据库系统进行建立、使用和维护等管理的系统软件，是数据库系统的核心。它处于用户和物理数据库之间，完成把数据库提供给用户进行管理的功能。DBMS 建立于操作系统之上，数据库命令及应用程序的执行都是通过 DBMS 来完成的。用户对数据库数据的任何操作，包括数据库定义、数据查询、数据维护、数据库运行控制等都是在 DBMS 管理下进行的，应用程序只有通过 DBMS 才能和数据库打交道。它的主要功能包括以下几个方面：

### (1) 数据库的定义和建立

数据库的定义称为数据库描述，用 DBMS 提供的数据库描述语言 (DDL, Data Description Language)，从用户的、概念的和物理的三个不同层次定义数据库，并将定义好的数据库模式保存起来，作为存取和管理数据的依据。

### (2) 数据库的操作

数据库的操作是 DBMS 面向用户的功能，主要作用是接收、分析和执行用户对数据库提出的各种数据要求，完成数据处理的任务。

### (3) 数据库的控制

数据库的控制包括数据库的安全性、完整性检查、数据共享的并发控制等，以保证数据库的可靠性和可用性。

### (4) 数据库的维护

数据库的维护包括记录数据库运行工作日记、监视和分析数据库性能。在用户要求或系统变化时，修改和更新数据库。

### (5) 故障恢复

在系统软件发生故障时，恢复数据库。DBMS 的故障恢复功能提供了有效措施，在系统发生故障时，能以最快的速度排除故障，恢复和重新启动系统，从而能把损失降低到最低限度。

数据库管理系统由语言处理程序、运行控制程序和系统建立维护程序等三部分组成。语言处理程序包括数据描述语言 (DDL)、数据操作语言 (DML)、终端命令解释程序和数据库控制命令解释程序等；运行控制程序包括系统总控制程序、存取控制程序、数据防护程序、并发控制程序、完整性控制程序和通信控制程序等；系统建立维护程序包括数据装入程序、工作日记程序、性能检测程序、重新组织程序、系统恢复程序、性能统计分析程序和存储程序密码定义程序等。

## 4. 数据库应用系统

数据库应用系统是一个复杂的系统，它由硬件、操作系统、数据库管理系统、编译系统、用户应用程序和数据库组成，也就是说数据库应用系统 (DBAS) 是在数据库管理系统支持下的一类计算机应用系统。我们通常把数据库系统和各种方便操作的应用程序合起来称为数据库应用系统，它不同于使用普通文件和由文件管理系统支持的应用系统。因此一个数据库应用系统还包括专用的或通用的 DBMS。

该系统中应用程序可以分为功能程序和控制程序两大类。功能程序的主要任务是完成特定的操作要求，如修改、查询、打印等；控制程序又称为菜单程序，它的主要任务是显示功能清单 (菜单)，接受用户的选择，并按其选择调用相应的程序来完成用户的操作意向。

## 1.1.2 数据库技术的发展

随着计算机的普及，以及电子计算机软件和硬件技术的发展，数据处理过程发生了划时代的变革，人们找到了一种能够对大量数据进行方便、有效、准确、快速地管理的技术，这就是数据库技术。采用数据库技术管理数据就是把相关的数据集中存放在一个或多个数据库文件中，用户通过数据库管理软件来使用数据库中的数据。

### 一、数据管理的发展历史

数据管理是对数据进行分类、存储、检索以及维护等方面的操作。数据管理技术发展到今天，经历了人工管理阶段、文件系统管理阶段、数据库系统阶段和高级数据库技术阶段这四个阶段。

#### 1. 人工管理阶段

人工管理阶段是数据管理的初始阶段，可以说是管理最低级的阶段。这一阶段主要指 20 世纪 50 年代以前 (当然根据各国的管理水平不同而不同)，当时计算机的发展刚刚起步，而且主要应用于科学计算中，受到客观条件的限制，没有相应的操作系统、也没有相应的软件对数据进行管理。当时采用的数据处理方式是批

处理方式。当时数据管理主要表现为：

(1) 数据与程序之间存在相互依赖性。程序往往是和数据结合成为一个有机的整体，这时数据往往作为程序不可缺少的一部分而存在(虽然有时为了方便,在形式上数据与程序是分开的),在逻辑上两者是相互依赖的。程序的设计往往考虑最多的是数据的传输问题,而数据存储结构的变化也要影响到程序中与数据读取有关部分的执行。在这种条件下,通常是将数据与程序结合在一起,运行程序的同时将其装入存储空间中,而程序结束时又同时释放空间。

(2) 数据不具备共享性,即一个程序的数据不能被另一个程序所使用。当程序需要使用其他程序中的部分数据时,只有通过再次输入这些数据才能使程序得以运行,这样增加了数据的重复性,而且数据几乎没有通用性,这是这种方式最大的缺陷,所有这些对系统资源来说是一种极大的浪费。

(3) 没有文件的概念。数据是由应用程序设计者自行设计安排的。

## 2. 文件系统管理阶段

这一阶段主要是指 20 世纪 50 年代中期到 60 年代中期。这一阶段计算机得到了很大程度的发展,计算机已经超出应用于科学计算的范畴,开始进入管理系统中。在计算机硬件方面,已经出现了许多直接存储设备,为新的发展提供了硬件上的支持,大大方便了这一阶段的数据管理。这期间数据管理的特点主要体现为:

(1) 数据开始具有独立保存性。相对于以前的管理,这可以说是很大的进步。数据可以独立于程序而保存在存储介质上,并且可以根据需要随时进行修改、查询、插入等操作,并可以让多个用户使用,具有了一定的共享性,这就给数据的使用者提供了极大的方便。与前一阶段相比,这是一个巨大的进步。

(2) 文件系统提供了程序与数据之间的存取方法。文件管理系统作为程序和数据之间的一个接口,可以通过文件管理系统建立程序和数据之间的联系,而不必去关心数据实际的物理位置。这些为设计者提供了方便。

(3) 文件类型多样化。出现了各种类型的文件,对数据的访问方式也呈现多样化,可以顺序访问,也可以随机访问。

## 3. 数据库系统阶段

文件系统管理阶段尽管在数据管理方面有了较大的进步,但在许多方面仍存在许多不足。主要的表现就是数据冗余,即仍然会出现同样的数据段在不同的文件中存取这一情况,数据之间的联系性差,即数据在各文件之间缺乏联系,因此造成了数据联系的困难。

在 20 世纪 60 年代中后期,随着计算机技术的进一步发展,计算机在管理方面的规模日趋扩大,数据的处理量急剧增加。为了满足这种要求,提高效率,人们开始研究开发数据库系统。数据库系统是位于用户和操作系统之上的一层数据管理软件。

数据库系统管理数据,出现了以下特点:

(1) 数据库中的数据是结构化的。在文件系统中,数据在整体上看是无结构的,即不同文件中记录类型之间没有联系,它仅关心记录内部数据项之间的联系,而对数据记录之间的联系就考虑得很少;而数据库不仅考虑数据项之间的联系,还要考虑记录类型之间的联系。

(2) 数据库中的数据是面向系统的,不是面向某个具体应用的,这样在实际中就减小了数据冗余,从而实现了数据共享。

(3) 数据库系统比文件系统有较高的数据独立性。数据库系统的结构分为 3 级:用户的数据逻辑结构、整体数据的逻辑结构和数据的物理结构。当整体数据的逻辑结构或数据的物理结构发生变化时,应用是不变的。数据的独立性是通过数据库系统在数据的物理结构与整体数据的逻辑结构、整体数据的逻辑结构与用户的数据逻辑结构之间提供的映像实现的。

(4) 数据库系统为用户提供了方便的接口。用户不仅可以通过数据库系统提供的查询语言交互式命令来操纵数据库,也可以通过程序方式编程来操纵数据库,这样就拓宽了数据库的应用面。

(5) 数据库系统还具有如下的控制功能:

**数据的完整性。**数据的完整性在数据库的应用中是很重要的,为了保证数据库的正确性,要使用数据库系统提供的存取方法设计一些完整性规则,对数据值之间的联系进行校验。

**数据的安全性。**在实际的应用中,并非每个应用都可以存取数据库中的全部数据。它可能仅仅是对数据库中的一部分数据进行了操作,因此需要保护数据库以防止不合法的使用,避免数据的丢失、被窃取,这样看来数据的安全性就是十分重要的。这就是所谓的数据安全问题。

**并发控制。**当多个用户同时存取、修改数据库中的数据时,可能会发生相互干扰,使数据库中数据的完整性受到破坏,而导致数据的不一致性。数据库的并发控制防止了这种现象的发生,提高了数据库的利用率。

**数据库的恢复。**有的时候会出现软硬件的故障,这时数据库系统应具有恢复能力,能把数据库恢复到最近某个时刻的正确状态。

#### 4. 高级数据库技术阶段

20世纪70年代中期以来,随着计算机技术的不断发展,出现了分布数据库、面向对象数据库和智能知识数据库等,这些技术通常被称为高级数据库技术。

## 二、数据库技术的发展特点

### 1. 数据库技术的新进展

数据库技术与其他学科的内容相结合,是新一代数据库技术的一个显著特征,涌现出各种新型的数据库,例如:数据库技术与分布处理技术相结合,出现了分布式数据库;数据库技术与并行处理技术相结合,出现了并行数据库;数据库技术与人工智能相结合,出现了演绎数据库、知识库和主动数据库;数据库技术与多媒体处理技术相结合,出现了多媒体数据库;数据库技术与模糊技术相结合,出现了模糊数据库等等。下面以几个新型数据库为例,描述数据库技术是如何吸收、结合其他计算机技术从而形成了数据库领域的众多分支和研究课题,极大地丰富和发展了数据库技术的。

#### (1) 分布式数据库(Distributed DataBase)

把集中式数据库用网络连接起来,使分散在各个场地上集中式数据库可以被网络上的用户通过远程登录加以访问,或者通过网络传递数据库数据,这就是分布式数据库。分布式数据库的主要特点是:

**数据的物理分布性。**数据库中的数据不是集中存储在一个地区的一台计算机上,而是分布在不同场地的计算机上,且每个计算机拥有相同的等级。

**数据的逻辑整体性。**数据库虽然在物理上是分布的,但这些数据并不是互不相关的,它们在逻辑上是相互联系的整体。

**数据的分布独立性(也称分布透明性)。**分布式数据库中除了数据的物理独立性、数据的逻辑独立性外,还有数据的分布独立性。从用户的视角来看,整个数据库仍然是一个集中的数据库,用户不必关心数据的分布,也不必关心数据物理位置分布的细节,更不必关心数据副本的一致性,分布的实现完全由系统来完成。系统的操作者所看到的是一个整体的类似子集式的数据库。

**场地自治和协调。**系统中的每个结点都具有独立性,能执行局部的应用请求。同时每个结点又是整个系统的一部分,可通过网络处理全局的应用请求。

**数据的冗余。**在这点上分布式数据库与集中式数据库是不同的,分布式数据库中应存在适当冗余以提

高系统处理的效率和可靠性。因此，数据复制技术是分布式数据库的一项很重要的技术。

从上面我们不难看出，分布式数据库在物理上是分布的，在逻辑上是统一的，这是分布式数据库最根本的特点，也是它与其他类型的数据库相区分的标志。

目前，分布式数据库开始进入实用阶段，但现有的分布式数据库技术尚不能解决异构数据和系统的许多问题。虽然已有很多数据库研究单位在进行异构 DBMS 集成问题的探索，并且已有一些系统宣称在一定程度上实现了异构系统的互操作，但是异构分布式数据库技术还远未成熟，有待进一步研究。

## (2) 并行数据库 (Parallel DataBase)

并行数据库是在并行机上运行的具有并行处理能力的数据库系统。并行数据库是数据库技术与并行处理技术相结合的产物。近年来，随着在应用领域数据库规模的急剧膨胀，数据库工作的负荷日益加重，对数据库性能的要求也越来越高，如何采用新技术提高数据库性能已成为数据库研究领域的一个迫切要求。目前利用多处理机并行处理产生的规模效益来提高性能的并行处理技术的迅速发展，以及微处理机技术的进步为数据库研究领域提供了一个极好的条件，使数据库领域开始将并行处理技术与数据库技术相结合以提高系统效率，从而形成了并行数据库这一新兴的数据库技术。

并行处理技术与数据库技术相结合，具有潜在的可行性。因为数据库模型，特别是关系数据库模型，在设计本身就有极大的并行可能性。关系数据库模型中，数据库是元组的集合，数据库操作实际是集合操作，许多情况下可分解为一系列对子集的操作，由于这些子集操作不具有数据相关性，因而具有潜在的并行性。

从硬件结构来看，目前主要有四种流行的并行数据库硬件结构：

Shared-Everything (SE)，即所有处理机共享同一全局内存和所有磁盘；

Shared-Disk (SD)，即每个处理机拥有自己的内存，但共享所有磁盘；

Shared-Memory (SM)，即由多个处理器共享主存储器和多个磁盘；

Shared-Nothing (SN)，即每个处理机都是一个独立的整体，拥有自己的内存和磁盘。

目前，并行数据库的研究主要有以下几方面：

实现数据操作的并行算法的研究。

如何实现对数据库数据的并行操作是并行数据库研究的热点，近年来，研究大多集中于这一领域并取得了很多有实效的成果。并行算法研究最多的是 Join 并行算法，它对传统的嵌套循环算法、排序合并算法和 hash 算法都进行了并行化。实验表明，这些并行算法的使用对改善关系运算的效率很有效。

并行数据库物理存储结构的研究。

目前，这方面的工作主要围绕着数据库文件的划分及其在多处理机或多磁盘之间的分布（以下简称数据的划分与分布）进行。实践证明，数据分布的好坏将直接影响数据库的并行度。目前这方面的工作主要集中在一维划分分布问题上。常用的有值域划分、轮转划分和 hash 划分三种方法。

为有效支持多维区域的查询和在非划分属性上具有查询条件的查询操作，需要研究多维数据划分和分布方法。动态多维数据划分和分布问题，以及现有顺序文件结构（如 B 树、Grid 文件）的并行化问题是需要进一步研究的重要问题。

并行数据库查询优化。

优化问题始终是影响数据库性能的一个重要因素，在原有问题还没有完全解决的情况下，并行技术的引入又带来了一些新的问题。传统的静态优化在并行数据库中存在难以克服的缺点，从而导致了新的动态优化技术的出现。动态优化虽然具有显著的优点，但它的实现相对来说要困难得多。

并行数据库查询优化的研究刚刚开始，取得的成果也非常有限，距离并行数据库系统设计与实现的要求相差很远，有大量的工作有待于开展。例如，多个 Join 操作的并行查询优化算法的研究、数据操作算法多参

数复杂性模型、并行查询计划的复杂性模型、查询优化算法本身的并行化研究等。

并行数据库系统是最近刚刚兴起的数据库研究领域，国内外尚无真正的并行数据库系统投入运行。最近一些著名的数据库厂商开始在数据库产品中增加并行处理能力，并在并行计算机上运行。他们只是使用并行数据流方法对原有系统加以简单的扩充，既没有使用并行数据操作算法，也没有并行数据查询优化的能力，因而都不是真正的并行数据库系统。目前，并行数据库的原型系统有 ARIBRE、BUBBA、GAMMA、TERADAT 及 XPRS 等。尽管它的研究有待于进一步地发展和实践，但不可否认，在注重效率的当今社会，并行数据库将成为数据库学科的一个非常重要的分支。

### (3) 主动数据库 (Active DataBase)

主动数据库是相对于传统数据库的被动性而言的。许多实际的应用领域中，如计算机集成制造系统、管理信息系统、办公室自动化等系统中常常希望数据库系统在紧急情况下，能根据数据库的当前状态，主动适时地做出反应，执行某些操作，向用户提供有关信息。但目前传统数据库系统是一种被动的系统，它只能被动地按照用户给出的明确请求执行相应的数据操作，很难充分适应这些应用的主动要求，因此在传统数据库基础上，结合人工智能技术和面向对象技术提出了主动数据库。

主动数据库设计的主要目标是提供对紧急情况及时反应并作出响应的能力，同时提高数据库管理系统的模块化程度。主动数据库通常采用的方法是在传统数据库系统中嵌入 ECA (即事件—条件—动作) 规则，在某一事件发生时引发数据库管理系统去检测数据库当前状态，看是否满足设定的条件，若条件满足，便触发规定动作的执行。为了有效地支持 ECA 规则，主动数据库的研究主要集中于解决以下问题：

主动数据库的数据模型和执行模型，即如何扩充传统的数据库模型，使之适应于主动数据库的要求。

执行模型，即 ECA 规则的处理和执行方式，是对传统数据库系统事务模型的发展和扩充。

条件检测，是主动数据库系统实现的关键技术之一，由于条件的复杂性，如何高效地对条件求值对提高系统效率有很大的影响。

事务调度。与传统数据库系统中的数据调度不同，它不仅要满足并发环境下的可串行化要求，而且要满足对事务时间方面的要求。

体系结构。目前主动数据库的体系结构大多是在传统数据库管理系统的基础上，扩充事务管理部件和对象管理部件以支持执行模型和知识模型，并增加事件检测部件、条件检测部件和规则管理部件。

系统效率。系统效率是主动数据库研究领域中的一个重要问题，是设计各种算法和选择体系结构时应主要考虑的设计目标。

主动数据库是目前数据库技术中一个活跃的研究领域，近年来的研究已取得了很大的成果。但许多概念尚不成熟，不少技术问题还有待进一步研究解决。

综上所述，当数据库技术与其他计算机技术结合后，不仅大大丰富并提高了数据库的功能、性能和应用领域，而且大大促进了数据库的概念和技术的发展。

数据库技术和其他相关技术相结合产生了众多新型的数据库系统，在新一代数据库大家族中，它们是重要的组成成员。应该指出，这些系统不是相互孤立的概念和系统，它们是互相关联的。例如，虽然分布式数据库系统强调了分布式的数据库结构以及分布处理功能，但它们支持的数据模型可以是关系模型、扩展关系模型、OO 模型或者某一种特定数据模型。另外主动数据库系统虽然强调的是数据库反应能力上的主动性、快速性和智能化的特性，但是其数据模型有的是在关系模型中加入事件驱动的主动成分，有的是研究用 OO 模型实现主动数据库，而它的系统结构可以是集中式的，也可以是分布式的。所以不同的数据库系统是互相关联的，而具体到某一应用系统中的数据库系统常常兼有以上多种数据库系统的技术特性，而不是简单的割裂单一数据库系统。

## 2. 面向应用领域的数据库新技术

数据库技术被开发出来，最终还是要应用到特定的领域中，这样就出现了工程数据库、地理数据库、统计数据库、科学数据库、空间数据库等多种数据库，使数据库领域中新的技术内容层出不穷。

### (1) 工程数据库 (Engineering DataBase)

工程数据库的特点是能存储和管理各种工程图形，并能为工程设计提供各种服务。它适用于 CAD / CAM、计算机集成制造 (CIM) 等通称为 CAX 的工程应用领域。传统数据库处理的是简单的对象和相对比较规范化的数据，而对具有复杂结构和内涵的工程对象以及工程领域中的大量“非经典”应用，它们是无能为力的。工程数据库正是针对传统数据库的这个缺陷而采取相应改进后的数据库，它针对工程应用领域的需求，对工程对象进行处理，并提供相应的管理功能及良好的设计环境。为了支持工程数据库，相应地开发了工程数据库管理系统，基于工程数据库中数据结构复杂、相互联系紧密、数据存储量大的特点，和传统数据库管理系统不同的是，工程数据库管理系统应具有以下功能：

支持复杂多样的工程数据的存储和集成管理。

支持复杂对象（如图形数据）的表示和处理。

支持变长结构数据实体的处理。

支持多种工程应用程序。

支持模式的动态修改和扩展。

支持设计过程中多个不同数据库版本的存储和管理。

支持工程长事务和嵌套事务的处理和恢复。

从传统的数据模型来看，因为它不能满足 CAX 应用对数据模型的要求，因此在工程数据库的设计过程中，需要运用当前数据库研究中的一些新的模型技术，如扩展的关系模型、语义模型、面向对象的数据模型。虽然在目前的工程数据库研究中已取得了很大的成绩，但要全面达到应用所要求的目标仍有待进一步深入研究。

### (2) 统计数据库 (Statistical DataBase)

统计数据是人类对现实社会各行各业、科技教育、国情国力的进行大量调查后获取的数据，是人类社会活动结果的实际反映，是信息行业的重要内容。采用数据库技术实现对统计数据的管理，对于充分发挥统计信息的作用具有决定性的意义。

统计数据虽然具有层次型特点，但它也不完全是层次型结构，而且统计数据也有关系型特点，但关系型也不完全能满足需要。统计数据的第一个特点就是多维性，这也是最基本的特点；其次，在实际的应用中，统计数据要求在一定时间（年度、季度、月度）期末产生大量数据，因此入库时总是定时地大批量加载。经过各种条件下的查询以及一定的加工处理，通常又要输出一系列结果报表。这就是统计数据的“大进大出”特点；第三，统计数据的另一个最基本的属性是时间属性，即任何统计量都离不开时间因素，而且经常需要研究时间序列值，所以统计数据查询又有时间向量性的特点；第四，由于不同的用户的观察角度是不同的，因此统计数据查询出来后常有转置的要求。例如若干指标的时间序列值，考虑指标之间的比例关系时常以时间为主栏、指标为宾栏；而考虑时间上的增长量、增长率时，又常以时间为宾栏、指标为主栏。统计数据还有其他的一些特点，但是它的基本特性就是多维结构特性。

统计数据库是一种用来对统计数据进行存储、统计（如求数据的平均值、最大值、最小值、总和等等）和分析的数据库系统。用户在应用统计数据库时，主要是应用统计数据库中提供的统计数字，而不是某一个体的具体数据。在统计数据库中，数据可以分为两类：微数据 (Micro Data) 和宏数据 (Macro Data)。这两者的定义也是相对的，微数据描述的是个体或事件的信息，而宏数据是综合统计数据，它可以直接来自应用领域，也可以是微数据的综合分析结果。

由于统计数据库具有一系列自身的特点，一般关系型数据库还不能完全满足它的需求。因此，如何使用 RDBMS 建立统计数据库，是一项具有特定技术的工作。

### (3) 空间数据库 (Spatial DataBase)

空间数据库是以描述空间位置和点、线、面、体特征的拓扑结构的位置数据及描述这些特征的性能的属性数据为对象的数据库。其中的位置数据为空间数据，属性数据为非空间数据。空间数据是用于表示空间物体的位置、形状、大小和分布特征等信息的数据，用于描述所有二维、三维和多维分布的关于区域的信息，它不仅具有表示物体本身的空间位置及状态的信息，还具有表示物体的空间关系的信息；非空间信息主要包含表示专题属性和质量描述的数据，用于表示物体的本质特征，以区别地理实体，对地理实体进行语义定义。

空间数据库的目的是利用数据库技术实现空间数据的有效存储、管理和检索，为各空间数据库用户所使用。目前，空间数据库的研究主要集中于空间关系与数据结构的形式化定义、空间数据的表示与组织、空间数据查询语言、空间数据库管理系统。

众所周知，任何一项技术的出现，都离不开市场的推动以及相应的硬件技术的发展，数据库技术也是如此，推动数据库技术前进的原动力是应用需求和硬件平台的发展。正是这些特定领域的特定需求的提出，才产生了特种数据库系统的研究，从而推动了新一代数据库技术的产生和发展。而新一代数据库技术也首先在这些特种数据库中发挥了作用，得到了应用。而特种数据库技术也会相应地吸取那些已经存在的数据库的成型的技术，发展自身的技术。

## 1.2 数据库的数据结构

传统的数据库有三类：层次数据库、关系数据库和网状数据库。它们分别采用树、图和线性表三种不同数据结构。

### 一、层次数据库结构

早期层次数据库结构的典型代表是 IBM 公司的 IMS (Information Management System)，例如学校实验室组织的层次数据结构模型，如图 1-1 所示。

层次数据库结构是使用链表结构方式组织数据的，如图 1-2 所示。

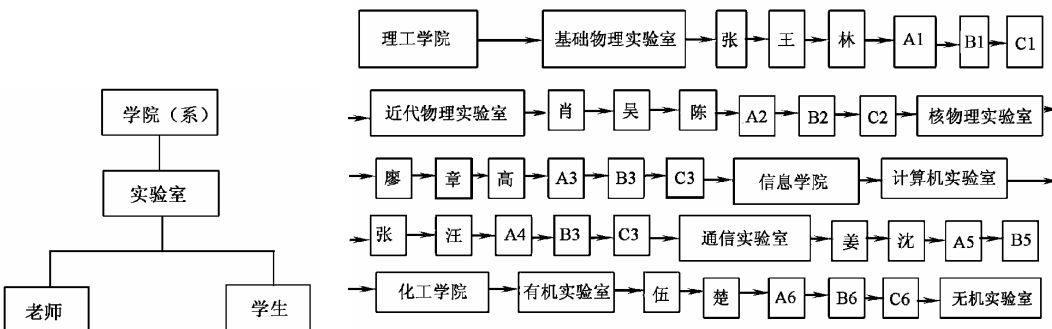


图 1-1 学校实验室组织的层次数据结构模型

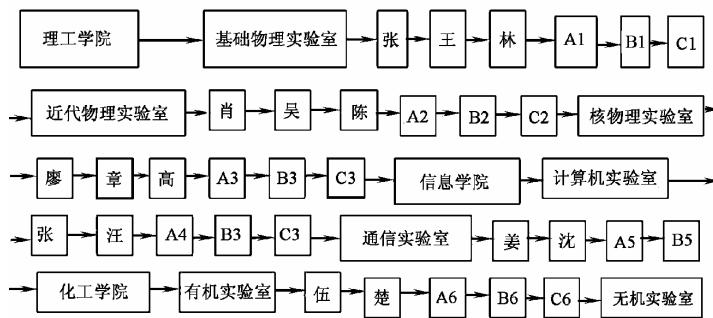


图 1-2 链表式结构的层次数据库

## 二、关系数据库结构

关系数据库采用线性表结构组织数据，对前述同样问题，它采用学院、实验室、教师、学生等四个线性表来组织有关数据，如表 1-1 ~ 表 1-3 所示。

表 1-1 学院数据表

学院代码	学院名	地 址	电 话	学院实验室
9410	理工学院	物理楼	84221234	基础物理、近代物理、核物理
9310	信息学院	信息楼	84225678	通信、自控、计算机、微波
9200	化工学院	化学楼	84229876	有机、无机、高分子、同位数

表 1-2 教师数据表

校编号	姓 名	性 别	实验室	职 称
94102311	张三	男	基础物理	教授
94102321	王琳	女	基础物理	工程师
94102331	林小明	男	基础物理	助理工程师
94102351	肖婴	女	近代物理	工程师
92002371	吴小华	男	近代物理	高级工程师
93102361	陈丽	女	近代物理	助理工程师

表 1-3 学生数据表

学 号	姓 名	性 别	实 验 室
20001019030	A1	男	基础物理
20001019034	B1	女	基础物理
20002019014	A2	女	近代物理
20002019064	C2	男	近代物理
20002020064	A5	女	通 信
20002020032	B6	男	自 控
20002020014	A7	男	微 波
20002021014	A8	女	有机化学
20002021054	B9	女	无机化学

注：1. 关系数据库采用的每一表称为一个关系。

2. 表的每一行称为一条记录，代表一个实体。

3. 每一列称为字段或数据项，代表实体一个属性。

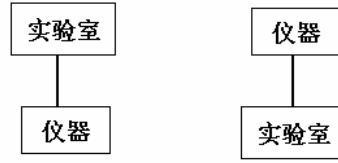
## 三、网状数据库结构

网状数据库采用链表结构组织数据，其典型代表是 DBTG 数据库。

在 DBTG 中，将不同实体集的数据分别在不同的存储区域内存放，针对不同应用需求建立不同链表描述实体集与实体集的联系，然后根据不同应用需求，使用不同的链表联系。图 1-3 描述了两个“系”：“实验室 - 仪器”及“仪器 - 实验室”，前者用于查一个实验室有哪些仪器的问题，后者用于查一种仪器在哪些实验室的问题。

对应一个实体的数据称为记录，其数据结构称为记录的每一个实体集与实体集的联系整体称为“系”，其数据在一个“系”中可以有許多链，一条链链接的所有数据为“系值”。链的起始元素称为系主记录，其他称为成员是根，成员记录是叶子，构成一棵二级树。

采用链表结构优点是效率高，缺点是结构较复杂、作缺少灵活性；采用线性表结构并以顺序文件形式存放简单、数据维护容易、容易实施、有很强的适应性和灵活的主要形式，但其缺点是效率较低。



“实验室—仪器”系与“仪器—实验室”系

图 1-3 链表联系

录型。面向应用结构称为系型。元素的集合称记录。系主记录

维护不方便、操作优点是结构较灵活性，是目前采

## 1.3 数据模型

### 一、数据模型的概念

设计数据库系统时，一般先用图或表的形式抽象地反映数据彼此之间的关系，称为建立数据模型。常用的数据模型一般可分为两类，一是语义数据模型，如实体—联系模型（E-R 模型）、面向对象模型等；二是经典数据模型，如层次模型、网状模型和关系模型。

语义数据模型强调语义表达能力，建模容易方便，概念简单、清晰，易于理解，是现实世界到信息世界的第一层抽象，是用户和数据库设计人员之间进行交流的语言；经典数据模型用于机器世界，一般和实际数据库对应，例如层次模型、网状模型、关系模型分别和层次数据库、网状数据库和关系数据库对应，可在机器上实现。这类模型有更严格的形式化定义，常需加上一些限制或规定。设计数据库系统时通常利用第一类模型作初步设计，之后按一定方法转换为第二类模型，再进一步设计全系统的数据库结构。通常包括数据结构、数据操作和完整性约束三部分内容。

数据结构描述的是数据库数据的组成、特性及其相互间的联系，是对系统静态特性的描述。在数据库系统中通常按数据结构的类型来命名数据模型，例如层次结构、网状结构和关系结构的模型分别命名为层次模型、网状模型和关系模型。

数据操作指对数据库中各种对象的实例允许执行的操作的集合，包括操作及有关的操作规则，是对系统动态特性的描述。主要有检索和维护（包括录入、删除、修改）等两大类操作。

数据的约束条件指数据完整性规则的集合，它是给定数据模型中数据及其联系所具有的制约和依存规则，用以限定符合数据模型的数据库状态及其变化，以保证数据的完整性。

### 二、数据之间的联系

#### 1. 一对一联系（1:1）

若对于实体集 A 中每一个实体，实体集 B 中至多只有一个实体与之联系；反之对于实体集 B 中每一个实体，实体集 A 中也至多只有一个实体与之联系，则称实体集 A 与实体集 B 之间具有一对一联系，记为 1:1。

#### 2. 一对多联系（1:N）

若对于实体集 A 中的每一个实体，实体集 B 中有 n 个实体（n ≥ 0）与之联系；而对于实体集 B 中的每一个实体，实体集 A 中至多只有一个实体与之联系，则称实体集 A 与实体集 B 有一对多的联系，记为 1:N。

### 3. 多对多联系 (M:N)

若对于实体集 A 中的每一个实体, 实体集 B 中有 n 个实体 ( $n > 0$ ) 与之联系; 反过来对于实体 B 中的每一个实体, 实体集 A 中有 m 个实体 ( $m > 0$ ) 与之联系, 则称实体集 A 与实体集 B 之间有多对多联系, 记为 M:N。

### 三、实体—联系模型 (Entity-Relationship Model)

E-R 模型是 P.P.S.Chen 于 1976 年提出的一种概念模型, 是用 E-R 图来描述一个系统中的数据及其之间关系。

现实世界任何可相互区别的事务, 不论是实际存在的东西, 还是概念性的东西, 或是事务与事物之间的联系, 统称为实体。实体所具有的性质, 统称为属性。实体和属性都有“型”与“值”之分, 型是概念的内涵, 而值是概念的实例。同一类型的实体集合称为实体集。在 E-R 图中, 用长方形表示实体集, 在长方形框内写上实体名; 用菱形表示实体间联系, 菱形框内写上联系名; 用无向边把菱形和有关实体相连接, 在无向边旁标上联系的类型, 如 1、M 或 N; 用椭圆形表示实体或联系的属性。如用椭圆形将它与一个相应实体间以无向边相连, 或以表格形式, 表示方法为: 实体名 (属性 1, 属性 2...)

学校实验室教学系统的 E-R 图如图 1-4 所示。

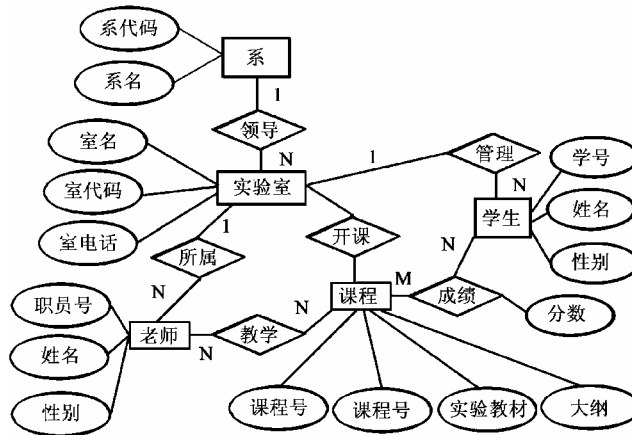


图 1-4 学校实验室教学系统的 E-R 图

要注意以下几个问题:

- (1) 某些联系也具有属性。
- (2) 对于三个实体 M:N:P 的联系的老师、学生、实验课程间的联系可用图 1-5 描述。
- (3) E-R 图中用以表现一个实体内部的部分成员和另一部分成员间的联系称为自回路, 如图 1-6 所示。