第 1 章 数据库基础知识

1.1 计算机应用的发展

自 1945 年第一台计算机 ENIAC 诞生以来，计算机作为人类的信息处理工具已有半个世纪。计算机应用技术的发展大体经历了 3 种计算模式：

- 单主机计算模式(mainframe computing)
- 分布式客户机／服务器计算模式（distributed client/server computing)
- 网络计算模式(network computing)

3 种模式一脉相承，成为计算机应用不同时期的主流计算模式，并决定了计算机硬件体系、软件体系和应用的面貌。

1.1.1 单主机计算模式

从计算机诞生之日至 1985 年，都是以单主机计算模式为主导，其特征是由单台计算机构成一个计算机系统。

早期计算机系统的硬件主要是由 CPU、内 /外存储器，以及若干种外部设备组成；其软件仅为简单的操作系统，高级语言编译器或解释程序，以及库函数和系统支撑软件等。由于计算机内存容量很小，外部设备有限，用户是通过计算机控制面板控制程序运行，程序运行中无法与操作者交互。所以，人们利用计算机主要是进行科学计算，完成重复性的大型计算任务。

IBM/360 分时操作系统的研制成功以及计算机终端的普及，形成了主机一终端的结构模式，用户操作计算机的方式发生了改变。用户可以通过终端直接使用计算机，利用终端与运行的程序交互。每个用户可以认为整个计算机资源都在为自己工作，但实际上主机是在轮流地为每个终端用户服务。操作系统统一地管理和调度软硬件资源，实现作业管理、CPU 与进程管理、存储管理、设备管理和文件管理等。

需要全本请访问 www.ertongbook.com

20 世纪 60 年代是编译技术的成熟期，编译技术、目标程序优化技术得到充分发展，应用程序编写也主要采用高级程序设计语言。为了降低大型计算的机时开销，程序设计的重点是选择好的算法，精心设计程序流程，流程图（flow chart）成为程序设计的主要表示方法。建立数学模型、选择计算方法、设计流程图、编写源程序代码、测试及交付使用是开发程序的一般步骤。

20 世纪 60 年代中期以前，软件通常是规模较小的程序，程序的编写者和使用者往往是同一个（组）人，这种个体化的软件环境养成了软件开发者的软件设计工作只是在头脑中进行的一种默认过程，程序开发的成果只是保存程序清单。随着计算机应用日益普及和深化，计算机软件规模以惊人的速度迅速膨胀。到 20 世纪 70 年代，一个程序可包含几百万行源程序语句。而此时，开发者仍然沿用早期形成的个体化软件开发方法。当程序运行中发现错误时、当用户有新的需求时、当硬件和操作系统更新时，开发者都需要对程序进行修改，而这往往要耗费巨大的人力。更严重的是，许多程序的个体化特性使得程序最终不可维护，软件危机开始出现。

什么是软件危机呢？软件危机是指在计算机软件的开发和维护过程中所遇到的一系列严重问题。为了解决这一系列问题，计算机工作者在如何开发软件、如何维护不断膨胀的已有软件等方面进行探讨，并提出了一系列技术方案。

从 1965 年至 1975 年为结构化程序设计阶段。结构化程序设计方法使程序开发的速度和规模都有大幅度的提高，不同程度上缓解了软件危机及硬软件发展不成比例的压力。通过大型程序的开发，软件人员认识到“分析应重于设计”“设计应重于编码”。

软件设计人员将一个问题分解成相关的模块（module），使模块之间的耦合程度尽可能低，相互之间的影响尽可能小，从而减少一个模块的变动对其他模块的影响。在结构化方法中，功能分解的设计技术采用了直观的数据流图（data flow diagram，DFD），它描述了数据在模块之间的流动。结构化程序设计方法自然地支持了自顶向下逐步细化的设计方法。20 世纪 70 年代的程序最大规模可达几百万条语句。

从 1975 年至 1985 年为单主机计算的软件工程阶段。如何才能降低软件的成本呢？软件只有大量生产、高度复用、易重组、易维护，才能大幅度降低费用。所以，人们提出了“工程化开发软件”的方法，以缩短开发周期、降低软件费用、寻求最佳的工程效果。

软件工程方法采用工程的概念、原理、技术和方法来开发和维护软件，即将经过实践考验并证明是正确的管理技术和当前能够得到的最好的软件技术结合起来，这就是软件工程。软件工程方法主要有生命周期法、快速原型法、面向对象方法、软件构件技术等。

分析传统产业的发展，其基本模式都是标准零部件（构件）生产和基于标准构件的

产品生产（组装）相结合。标准零部件生产业的独立存在和发展是产业形成规模经济的前提，计算机硬件产业的成功发展就是基于这种模式，它充分证明了这种模式的可行性和正确性。软件产业要想快速发展并形成规模经济，也必须采用这种模式。

总而言之，单主机计算模式的计算机系统有其特点：在多用户操作系统的支持下，各个用户通过终端设备访问同一台计算机，实现资源共享。特别是主机上的数据库系统提供了信息管理和信息资源共享的先进技术，为计算机应用开辟了广阔的领域。单主机计算模式的管理与维护也比较简单，因为系统管理任务集中在一台计算机上。

但是，单主机模式也存在一些缺点。例如，当在线用户数增加时，要求主机必须有大量的系统资源作支持。为了适应用户处理的需要，系统往往就必须升级，更换功能更强的计算机设备，此时就显得主机系统的灵活性差。到了 20 世纪 80 年代，网络技术趋于成熟，个人计算机蓬勃发展，微机局域网得到了广泛应用。用户通过网络共享计算机资源，其投资远远低于主机一次性投资。因此到 20 世纪 90 年代初期时，客户机／服务器计算模式开始兴起，并在不同领域中得到广泛应用。

1.1.2 分布式客户机／服务器计算模式

PC 机的出现和网络技术的成熟，促使计算机应用以惊人的速度发展。虽然与大型机相比，PC 机的资源有限，但在网络技术支持下，应用程序不仅可用本机资源，还可通过网络方便地共享网上其他计算机资源。利用 PC 机、局域网以及网络上各种计算机资源，也能实现大型机上的很多应用。在这种背景下，形成了分布式客户机／服务器的计算模式。

在企业中，各项业务活动之间的协调是非常重要的，各项业务的处理不能孤立进行，它们之间要相互交流或共享资源。因此，企业的计算机应用模式有必要扩展到局域网的计算环境。在这种环境中，企业内部的各种计算机连接在一起，相互通信，资源共享。

客户机／服务器模型与人类生产、交流、生活中各种群体协作活动模型天然一致，例如民航售票系统、银行日常业务系统、计算机网络辅助教学系统等。客户机／服务器的计算模式使计算机应用从完成某种功能软件的层次上升到系统层次。

企业网络计算机可以按客户机／服务器（client/server）的概念建立应用模式，建立分布式处理的企业计算机应用环境。网络上的各种设备可各司其责：基层的工作站设备群可由各种微机组成，运行客户程序模块；部门级、企业级的各种不同规模的计算机（该计算机必须支持多用户多任务）可作为服务器使用，运行各种不同的服务器程序模块（如电子邮件服务器、文件服务器、数据库服务器等等）。

目前，客户机／服务器模式仍然是 Internet 上最为普遍的应用模式。Internet 包容

了人类文明共有的信息宝藏，例如有大型联网图书馆、学术文献库、网上杂志、网上新闻、网上博物馆等。交易、购物、学习等活动等都可以在网上进行。如果说 Internet 在相当长的时间内还只局限于科研学术界，那么从 20 世纪 90 年代起，随着环球信息网 WWW(world wide web) 的发明和广泛应用，Internet 现在已是“飞人寻常百姓家”了。Internet 的重要性归根到底是两个方面：一方面它提供了人与人之间快速的通信服务，从这个意义上说，它使世界变小了；另一方面它提供了网上巨大的信息资源，所以它又使每个人的世界变大了。正是天涯咫尺，咫尺天涯。网上的资源包括丰富的硬软件资源和信息资源，而 Internet 的应用服务也正是为用户提供了各种工具，使用户能方便地获取和利用这些资源。随着 Internet 的商业化和电子商务的发展，Internet 的应用愈来愈广泛。

Internet 向用户提供了 3 种最基本的服务，它们是：

1. 远程登录服务（Telnet）

Telnet 允许用户的主机通过 Internet 登录到远程主机。实际上，在 UNIX 工作站或 PC 机的 MS-DOS 界面(或 Windows 的运行窗口）键人命令‘ telnet 远程主机域名”后，会出现远程主机的登录信息，在提示符后键入用户名和口令就可以登录到远程主机了。这样一来，用户的主机就好像变成了连在远程主机的终端。 Telnet 是一个虚终端协议，它使用户的主机被模拟成终端，并和远程主机进行通信。注意：主机域名就是主机在 Internet 上惟一的名字，当然也可以指定 IP 地址。

2. 电子邮件服务 E-mail

用户可以在 Internet 上发送和接收电子邮件，电子邮件服务是 Internet 上使用最广泛的服务之一。

3. 文件传输服务(FTP)

FTP 允许用户在 Internet 上将一台主机上的文件复制到另一台主机。 FTP 将文件分成两类：ASCII 文本文件和二进制文件。用 FTP 既可传 ASCII 文本文件，也可用二进制方式传送任何其他文件，包括程序、图形、图像、语音等。FTP 在以二进制方式传输文件时，是原封不动地对源文件进行复制，不对文件格式进行任何转换。在 UNIX 工作站或 PC 机上运行 ftp 程序，指定好远程主机（文件服务器）域名或 IP 地址、并按提示要求键入用户名和口令后，就可以访问该远程主机上的文件系统。

有两类 FTP 服务 用户 FTP 和匿名 FTP。只有在远程主机上有账号的用户才能使用‘ 用户 FTP’，访问该主机上的文件。而匿名 FTP 是对公众的一种文件复制服务。

任何人都可以访问提供匿名 FTP 服务的主机，用户都使用 anonymous 作为用户标识名，而口令按惯例是用户的 E-mail 地址。匿名 FTP 一般限制用户只能访问文件服务器上的一部分内容。

这 3 种最基本的 Internet 服务的成功，激励了新的网络应用。环球信息网（简称 WWW 或 Web）服务是目前使用最广泛的 Internet 服务。Web 服务以客户机／服务器方式实现。在客户机上运行着 WWW 的客户机程序（即浏览器），该程序的主要功能是向用户提供友好的查询界面，将用户的信息请求转换成查询命令传送到 Internet 上相应的 WWW 服务器进行处理，然后将查询到的结果呈现给用户。WWW 服务器运行着 WWW 服务器程序（Web Server）。当 Web Server 接收到来自某一客户机的请求后，便进行网页的搜索并将得到的结果返回给客户机，再由客户机程序将这些数据转换成相应的形式显示给用户。Web Server 是信息的存储地和提供者，它不仅提供自身的信息服务，并且还“指引”存放在其他服务器上的信息。这样，全球范围的信息服务器便相互交织在一起，形成遍布世界的万维网。Internet 的贡献是向人们显示出信息共享的可能性和现实性，并重新定义了信息的共享方式。

在这里，要特别说一下数据库服务器的应用。网络的数据库服务提供了共享数据的存储、查询、管理和恢复等多方面的服务。在数据库服务中，客户机的任务是接收用户的请求，并将这些请求按一定格式发送到服务器，客户机还对服务器返回的数据进行处理，并按规定形式呈现给客户。数据库服务器用来分析用户请求，实施对数据库的访问与控制，并将处理结果返回给客户端。在这种模式下，网上传送的只是请求和少量的结果数据，其网络通信负担比基于文件传送（文件服务器）要少得多。

1.1.3 网络计算模式

网络计算模式是以万维网（WWW）的出现、主页文档标准化和 Java 语言诞生这三大技术的应用为主要标志。网络计算模式是由客户机/服务器方式发展而来，但并不是对客户机/服务器技术的改进与补充。

在 20 世纪 80 年代到 90 年代，随着 PC 机的发展，客户机／服务器体系结构开始逐渐流行起来。有人曾预言：“不久的将来，大型主机一终端方式的计算机系统将消亡”。实际怎样呢？大型主机计算机系统非但没有从市场上消失，相反还按自身的规律以合理的速度不断增长。靠大量 PC 机通过网络连接起来，并在此基础上建立企业的信息管理系统（MIS）并不那么简单。利用客户机/服务器体系结构把客户机、服务器、网络设备以及相应的软件产品集成，以此建立企业网络计算应用环境，在某些应用场合下要达到满意的性能要求，其费用大大高于主机系统。

由于 WWW 技术的最大特点是不受操作系统和硬件的制约，可以很容易地构筑大

需要全本请访问 www.ertongbook.com

型网络。所以，Internet 的 WWW 服务器和浏览器技术在企业信息系统中迅速普及开来。针对上述情况，各家软件开发商纷纷推出新型的体系结构，将 WWW 技术、客户机／服务器技术、主机系统技术以及数据库技术融合在一起，建立开放式的企业网络计算环境。

1. WWW 技术

WWW 是一个基于超文本（hypertext）方式的信息查询工具。WWW 技术最初由欧洲共同体核研究所(CERN)在 1990 年开发，它将位于全世界 Internet 上各站点的相关数据信息编织在一起，并提供一种界面友好的信息查询接口，用户可方便地获取信息。WWW 为用户带来的是世界范围的超文本服务，用户只要操纵鼠标器，就可以通过 Internet 网从全世界任何地方调来所要的文本、图像、活动影像、声音等信息。

此外，WWW 浏览器可为用户提供传统的 Internet 服务：如 Telnet 虚拟终端）、FTP 文件传输服务）、Gopher(分布式信息查询工具)等。

WWW 使用超文本标记语言 HTML(hyper text markup language)来描述网络上的资源，并以 HTML 数据文件的形式存放在服务器中。 HTML 数据文件为纯文本文件格式，由标准的 ASCII 码构成，并用超链接（hyper link）来表明信息之间的关联。HTML 利用 URL 表示超链接，可在文本内指向其他网络资源。 URL 可指向网络文件(file)、超文本传输协议（HTTP）文件传输服务（FTP)、Gopher 服务、远程登录(Telnet）以及电子新闻（News)等信息资源。

CGI(common gateway interface)为公共网关接口，它为 WWW 服务器定义了一种与其他应用程序共享信息的方法。当服务器接收到来自某一客户机的请求、并被要求启动一个网关程序（通常称为 CGI 脚本）时，服务器将有关请求的信息综合到一个环境变量中，交给 CGI 程序。CGI 脚本程序将检查这些环境变量，以试图找到那些为响应请求所必需的信息。 CGI 脚本负责处理由服务器请求引起的动态响应所必需的所有任务。CGI 的主要用途在于使用户能够编写浏览器界面中的交互程序。 CGI 标准非常简单，任何一种可访问 UNIX 环境变量的编程语言均可支持 CGI.

在 WWW 环境下访问数据库的处理过程如下：

①用户在浏览器界面通过填充由 HTML 创建的表格来执行查询和数据请求。

② CGI—Bin 脚本程序将输入到表格中的信息提取出来，并将它组织成有效的数据库操纵语句，并由 CGI 脚本程序将其发送到数据库服务器。

③ CGI 脚本程序将接收由数据库服务器返回的结果，并将结果以 HTML 格式回送到客户端浏览器上，显示给用户阅读。

CGI 为 Web 服务器和其他应用程序共享信息提供了可能。用户可通过 Web 页

面，交互式填入查询表(from)，通过 Web 存取数据库中的数据，并在浏览器上实时地显示结果。CGI 方式的实现存在性能问题，当每次访问时都需重新启动一个 CGI 程序实例，以进程方式运行，因而占用资源较多，效率较低，且难于解决连续的事务处理。因此 Web 浏览器厂商和数据库厂商竞相出台中间件以取代 CGI，目前已有的 Web_DB 的产品结构如图 1-1 所示。

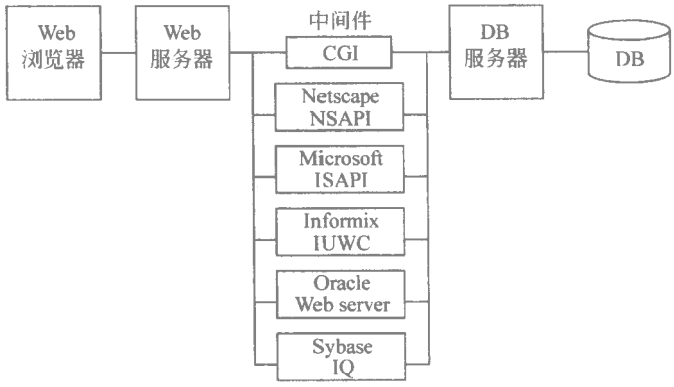


图 1-1 Web_DB 产品结构

2. Java 语言

Java 和 Web 都始于 20 世纪 90 年代初。Web 从诞生之日到流行，其间共用了 3 年时间。而从此时到 Java 在 Internet 上流行，又经过了两年时间。Java 语言在 1995 年被美国著名杂志《PC Magazine》评为十大优秀科技产品。Java 为 Internet 和 WWW 服务开辟了一个崭新时代。

Java 是面向对象的程序设计语言，它特别适用于 Internet 应用程序的开发。通过 HTML 的特定标记 APP，程序员可将称为 Applet 的 Java 程序包含在 WWW 的网页中。Applet 小程序由浏览器从服务器上取回并执行。Java 语言使 WWW 的交互性更加丰富。但 Java 技术更广泛的应用是在服务器端。例如，目前广为流行的 JSP 技术就能够快速地开发基于 Web 和数据库的各种应用系统。

网络计算模式深刻地改变了计算机的应用面貌，每个客户机都成为资源巨大的世界计算机的终端。客户机端的客户程序越来越简单，应用程序的功能主要依赖于服务器上的软件。

基于网络计算模式的软件开发要求开发者要掌握传统软件开发技术，并熟悉多媒体、网络、数据库、人工智能、通信技术、集成技术以及各种软件工具环境等。

1.2 数据库的系统结构与组成

数据库技术所研究的问题是如何科学地组织和存储数据，如何高效地获取和处理数据。数据库技术已成为计算机领域中最重要的技术之一，是软件学科中一个独立的分支。数据库的出现使得计算机应用渗透到各个领域及各行各业。事务处理系统、管理信息系统、办公自动化系统、决策支持系统等都是使用了数据库技术的计算机应用系统。数据库系统是计算机信息系统的基础和主要组成部分。

近年来，由于网络和多媒体技术的飞速发展，使得数据库在功能、性能等方面得到了进一步的改善。数据库技术由集中式向分布式方向发展，数据库应用系统的开发及运行方式也从原来的主机／终端结构发展为客户机／服务器结构、并进一步发展到当前的多层结构。数据库中的数据也从原来的数值数据、字符数据等扩展到包括语音、图像、图形等多种数据类型。数据库所支持的数据模型也从原来的单一数据模型发展为多种数据模型共存，如关系数据模型和面向对象数据模型共存于同一系统。

1.2.1 信息管理技术的演变

在数据处理领域中，有两个最基本的概念——数据和信息。

数据是描述客观事物的一组文字、数字或符号，它是客观事物的反映和记录。

信息是潜在于数据中的意义，它反映了客观世界中各种事物的状态与特征，它能增长人的知识，影响接受者的行为。众所周知，物质、能源和信息是组成社会的三大要素，而信息已成为社会发展的基础和主导力量，这是社会信息化的必然结果，是人类社会进步的标志。

数据处理是对各种形式的数据进行收集、存储、加工和传播的一系列活动的总和，其目的是从大量原始的数据中抽取出对人类有价值的信息，以作为行动和决策的依据。

在使用计算机之后，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和软件的发展，大体经历了以下 3 个阶段：

人工管理阶段

20 世纪 50 年代中期之前，计算机的软硬件均不完善。硬件存储设备只有磁带、卡片和纸带，软件方面还没有操作系统，当时的计算机主要用于科学计算。这个阶段由于还没有软件系统对数据进行管理，程序员在程序中不仅要规定数据的逻辑结构，还要设计其物理结构，包括存储结构、存取方法、输入输出方式等。当数据的物理组织或存储

设备改变时，用户程序就必须重新编制。由于数据的组织面向应用，不同的计算程序之间不能共享数据，使得不同的应用之间存在大量的重复数据，很难维护应用程序之间数据的一致性。

这一阶段的主要特征可归纳为如下几点：

- 计算机中没有支持数据管理的软件。
- 在程序中要规定数据的逻辑结构和物理结构，数据与程序不独立。
- 数据组织面向应用，数据不能共享，数据重复。
- 数据处理采用批处理方式。

人工管理阶段的处理方式如图 1-2 所示。

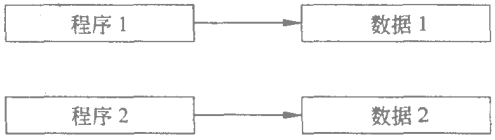


图 1-2　人工管理阶段的数据处理模式

2. 文件系统阶段

这一阶段的主要标志是计算机中有了专门管理数据的软件——操作系统。

20 世纪 50 年代中期至 60 年代中期，由于计算机大容量存储设备（如磁盘）的出现，推动了软件技术的发展，而操作系统的出现标志着数据管理步入一个新的阶段。在文件系统阶段，数据以文件为单位存储在外存，且由操作系统统一管理。操作系统为用户使用文件提供了友好的界面。文件的逻辑结构与物理结构脱钩，程序和数据分离，使数据与程序有了一定的独立性。用户的程序与数据可分别存放在外存储器上，各个应用程序可以共享一组数据，实现了以文件为单位的数据共享。

但由于数据的组织仍然是面向程序，所以存在大量的数据冗余。而且，数据的逻辑结构不能方便地修改和扩充，数据逻辑结构的每一点微小改变，都会影响到应用程序。由于文件之间相互独立，因而它们不能反映现实世界中事物之间的联系，操作系统不负责维护文件之间的联系信息。如果文件之间有内容上的联系，那也只能由应用程序去处理。

在文件系统阶段，数据处理的模式主要为分时和批处理，系统中程序与数据的关系如图 1-3 所示。

3. 数据库系统阶段

20 世纪 60 年代以后，随着计算机在数据管理领域的普遍应用，人们对数据管理技

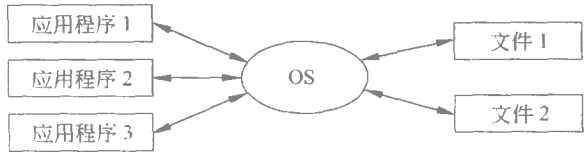


图 1-3　文件系统阶段的数据处理模式

术提出了更高的要求：希望面向企业或部门，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，以降低应用程序研制与维护的费用。数据库技术正是在这样一个应用需求的基础上发展起来的。

数据库技术有如下特点：

- 面向企业或部门，以数据为中心组织数据，形成综合性的数据库，为各应用共享。
- 采用一定的数据模型。数据模型不仅要描述数据本身的特点，而且要描述数据之间的联系。
- 数据冗余小，易修改、易扩充。不同的应用程序根据处理要求，从数据库中获取需要的数据，这样就减少了数据的重复存储，也便于增加新的数据结构，便于维护数据的一致性。
- 程序和数据有较高的独立性。
- 具有良好的用户接口，用户可方便地开发和使用数据库。

。对数据进行统一管理和控制，提供了数据的安全性、完整性以及并发控制。

从文件系统发展到数据库系统，这在信息领域中具有里程碑的意义。在文件系统阶段，人们在信息处理中关注的中心问题是系统功能的设计，因此程序设计占主导地位；而在数据库方式下，数据开始占据了中心位置，数据的结构设计成为信息系统首先关心的问题，而应用程序则以既定的数据结构为基础进行设计。

数据库系统阶段数据的处理模式如图 1-4 所示。

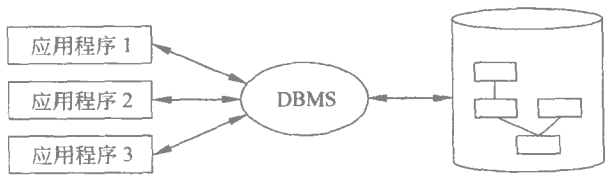


图 1-4　数据库系统阶段的数据处理模式

计算机用于信息处理至今，其应用系统大体可分为 4个层次：

. 电子数据处理

该层次的应用是用计算机代替各种手工工具，实现数据处理的计算机化。其意义在于减轻了用户的劳动强度，节省人力、时间，并且提高了数据处理质量和效率。例如工资的计算与报表打印等。

2. 事务处理系统

事务处理系统实现了部分业务管理，如财务系统、订票系统等。该层次的应用不仅提高事务处理效率，而且提高了管理水平，实现原来手工无法实现的功能。但系统一般不涉及全局，对管理控制与战略计划很少涉及。

3. 管理信息系统(MIS)

该类型的系统是对数据进行深加工，提取有价值的信息用于管理。通常 MIS 系统以一个部门的管理为背景，以基层事务子系统为基础，以完成该部门总体任务为目标。它强调企业中各局部系统之间的信息联系，提供满足各级领导从事管理的信息需求，并作用于管理的全过程。

目前各企事业建立的信息系统大都属于该种类型。

4. 决策支持系统

可以形象地说，决策支持系统是“进入董事会的系统”。该系统依据 MIS 系统提供的信息，支持作出高层管理的有效决策。系统强调对半结构化和非结构化问题的决策，借助模型库、知识库,具有试探、推理、演绎等功能,通过高度人机交互,建立人脑和计算机的协同工作。决策支持系统有助于将人们的高级脑力劳动科学化。

目前，决策支持系统还主要应用于部分专门的领域，如医疗领域中的专家系统等。

1.2.2 数据库体系结构

数据库体系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，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结构将有助于理解数据库设计及应用中所涉及到的一些基本概念。

在数据库系统中，对于同一意义下的数据（如学生的各项记录），从计算机中存储的二进制表示到用户处理的数据（诸如学生姓名、学号等概念）之间，存在着许多层次的抽

需要全本请访问 www.ertongbook.com

象和转换。数据库管理系统（DBMS）支持把数据库从逻辑上分为 3 个层次，即三级模式。数据库的三级模式为外模式、概念模式和内模式。

概念模式又称数据库模式（或简称模式），它是数据库中全部数据的逻辑结构和特征的描述。通常以某种数据模型为基础，并用数据库系统提供的模式描述语言（模式 DDL）来定义这些内容。DDL 是 data definition language（数据定义语言）的缩写。用模式 DDL 写出的一个数据库逻辑定义的全部语句称为概念模式。

概念模式可以被看作是现实世界中一个组织或部门中的实体及其联系的抽象模型在具体数据库系统中的实现。

概念模式不同于外模式，它一般与具体的应用程序无关。概念模式也不同于内模式，它不涉及数据的存储细节和硬件环境。

外模式又称用户模式或子模式，通常由概念模式导出，是概念模式的子集。外模式定义了允许用户操作的数据库数据，通常由 DDL 语言描述。外模式是用户看到的数据视图，即与某一应用有关的数据的逻辑表示。由于不同用户需求不同，看待数据的方式不同,使用的数据内容不同,对数据的保密要求不同,因而不同的用户的外模式是不同的。

模式与子模式不但要定义数据的逻辑结构，而且要定义与数据有关的安全性和完整性；不仅要定义数据记录的内部结构，还要定义这些数据之间的联系。

内模式又称存储模式，是对数据的物理结构和存储方式的描述。例如，记录如何存储，索引以什么方式组织等。内模式也是由 DDL 语言描述。

数据库的三级模式定义了数据库的逻辑结构和物理结构，它是数据库设计的主要内容。三级模式框架如图 1-5 所示。

数据库系统的三级模式结构在数据的 3 个抽象级上提供了 2 个层次的映像：内模式到模式的映像以及模式到外模式的映像。

内模式到模式之间的映像提供了数据的物理独立性：当数据的物理结构发生变化时(如改变存储设备,改变数据的存储位置,改变数据存储的组织方式、增加索引等),不影响数据的逻辑结构。

例如，为了提高应用程序的存取效率，数据库管理员和数据库的设计者根据各应用程序对数据的存取要求，对数据的物理组织会进行某些优化。而这些改变不需要对模式和外模式重新定义，也不需修改应用程序。

模式到外模式之间的映像提供了数据的逻辑独立性：即当数据的整体逻辑结构发生变化时（如在概念模式中增加新的数据类型，在原有记录类型间增加新的联系，以及为原有记录增加新的数据项等），可以不影响外模式。因为总可以通过修改外模式/模式映像来实现这一点，除非全局逻辑结构的变化使得外模式中的某些数据项无法再从数据库中导出。

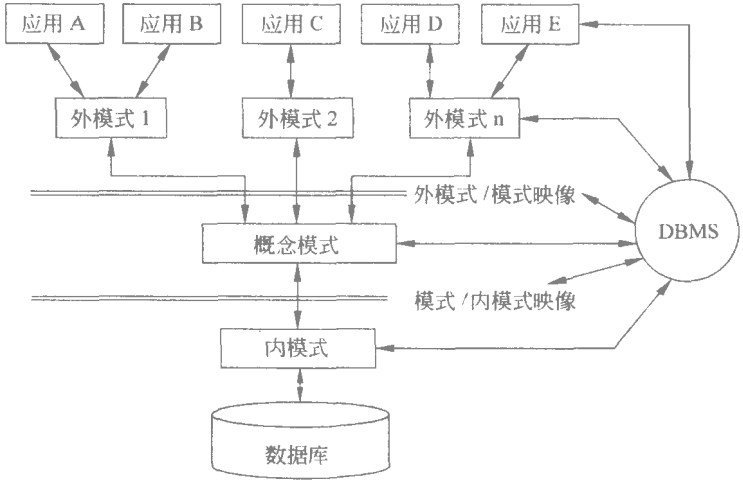


图 1-5　数据库系统结构的三级模式

例如，在数据库建立后，随着功能需求的变化，要求教务管理模式中增加因材施教信息、增设双学位课程的选修及授予学位的信息、增设学生毕业分配去向的信息等。当依据新的功能要求对原模式进行修改或扩充新的结构时，这种修改不影响外部级，不需要重新生成外模式，也不必重新编写应用程序，而仅对模式做部分修改或扩充，对外模式的定义做某些调整。当然，为了完成新的处理功能，需编制新的应用程序，或对原来的应用程序做必要的修改。

数据库的这种多层次的体系结构提供了高度的数据独立性。其中，数据库的全局逻辑描述是独立于其他所有结构描述的，在定义数据库结构时，概念模式应首先定义。内模式则是将概念模式中所定义的数据进行适当的组织并加以存储，以实现较好的时空运行效率。

总之，三级模式都是数据库管理的结构框架，而按照这些数据框架装入的数据才是数据库的内容。在设计数据库时，主要是定义数据库的各级模式，即设计数据库的逻辑结构和物理结构；而当用户使用数据库时，所关心的才是数据库的内容。数据库的模式通常是相对稳定的，而数据库中的数据是经常变化的。

三级模式是认识数据库的一种依据，但在具体的数据库产品中，三级结构的划分常常是不清晰的，故不可生搬硬套。

1.2.3　数据库的系统组成与工作流程

一个完整的数据库系统由数据库（即物理数据库）、软件、硬件和数据库管理员等几

部分组成。下面介绍一下数据库系统中的几个主要成分。

1. 数据库管理系统(database management system,DBMS)

数据库管理系统是对数据进行管理的软件系统，它是数据库系统的核心软件。数据库系统的一切操作，包括创建各种数据库对象，如表、视图、存储过程等，以及应用程序对这些对象的操作(如插入数据到表中,对表中数据的检索、修改、删除等),都是通过数据库管理系统进行的。

数据库管理系统主要提供以下功能：

①数据库定义功能

DBMS 通常都提供数据定义语言（DDL）。数据库设计人员通过这些语句来描述和定义数据库的结构，如数据库模式、内模式、外模式及安全保密等信息。这些定义中包含了数据库对象属性特征的描述、对象属性所满足的完整性约束条件、对象上允许的操作以及对象允许哪些用户程序存取等。

②数据存取功能

DBMS 提供数据操纵语言（data manipulation language,DML）。用户使用 DML 语言实现对数据库中数据的基本操作，如查询、插入、修改和删除等。一般大型数据库管理系统支持两种方式的数据库访问：交互方式访问和嵌入方式访问。交互方式允许用户在联机终端或客户机上以命令方式执行 DML 语句；而嵌入方式则采用将 DML 嵌入某一主语言（又称宿主语言，如 C、PASCAL 等),程序中的 DML 语言完成对数据库的访问，而主语言本身的语句完成对数据的加工与处理。

③数据库管理功能

DBMS 提供对数据进行管理和控制的机制，其中包括并发控制功能（当多个用户程序并发地存取数据库中的数据时，负责维护数据库中数据的完整性、一致性等）、存取控制功能（负责执行对数据的安全、保密及合法性检查等）、以及数据库内部的维护（如数据字典维护）。

2. 物理数据库

物理数据库是指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，在教务管理信息系统中，各应用程序获取的数据来自教务管理数据库。这个数据库中可能存放着学生、课程、选课、教师、教材等信息。

用户最为关心的是物理数据库中的数据，用户常说的数据库也就是物理数据库。

3. 数据字典

除了用户直接使用的数据外，还有另一类“数据”，它们是有关数据库的定义信息，如用户名及权限、数据库表的定义等。这些信息通常存放在称为“数据字典（data dictionary）”的地方，人们形象地把数据字典看作是有关物理数据库的数据库。

数据字典是数据库管理系统的组成部分之一，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，数据库管理系统通过它对物理数据库中的数据进行管理与维护。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，这些对象占用的空间信息，这些对象允许哪些用户使用等。但是，用户一般不能改动数据字典的内容。

4. 软件支持系统

数据库系统的软件除 DBMS 外，还需要一个软件支撑环境，其中包括操作系统、应用系统开发工具、各种宿主语言及实用程序等。 DBMS 是在操作系统的文件系统基础上发展起来的，而且一般是在操作系统的支持下工作的，这一点要特别注意。

应用系统开发工具是数据库系统为应用开发人员提供的一组开发工具集，其中常见的有：面向表格应用（业务人员的操作界面）的自动生成工具，面向报表的自动生成工具，绘制统计图表的自动生成工具等。这些开发工具大部分是可视化的第四代语言开发工具，采用图形界面，支持客户机／服务器运行模式，开发应用系统效率较高。

5. 硬件支持系统

硬件系统主要是指计算机。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的外存要足够大，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。

6. 数据库管理员及相关人员

数据库有关人员主要包括数据库管理员（DBA）、系统设计与开发人员以及普通用户。数据库系统中不同人员涉及不同级别的数据。

用户是指数据库的最终使用者，他们通过应用程序的操作界面使用数据库，完成日常业务处理。

系统设计与开发人员负责应用系统的设计与开发，其中包括需求分析、数据库设计、系统功能设计与实现等。

数据库管理员（DBA）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专人负责数据库系统的日常管理和维护工作。

数据库管理员通常对他们管理和维护的数据库的总体结构及细节了解得十分透彻。例如，数据库中有多少用户，他们存取和访问哪些数据，这些数据使用的频率如何，这些数据是如何分布的，如何优化整个系统的性能等。数据库管理员的主要任务包括：

- 负责数据库系统软件的安装和维护。
- 参与数据库设计。
- 监控数据库的运行（系统运行是否正常，系统效率如何），并进行系统参数的配置与调整。
- 权限管理（确保数据库的安全性，使多个用户程序能方便地共享数据，又防止非法用户获取系统信息）。
  日常维护（如定期对数据库中的数据进行备份，当系统出现故障时，能够在最短时间内把数据库恢复到某一正确状态）。
- 根据需要对数据库中的数据进行扩充或重组。
- 数据库有关文档管理。

数据库管理员在数据库管理系统的正常运行中起着非常重要的作用。

引入数据库系统后，计算机系统层次如图 1-6 所示。

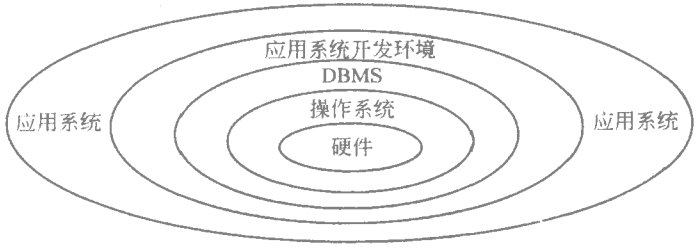


图 1-6 计算机系统层次图

下面通过一个用户对数据库的访问过程，描述一下数据库系统的工作流程及各组成部分的相互关系，如图 1-7 所示。

假如用户提交了一条 SQL 语句来查询数据库中的一组数据，数据库系统是这样工作的：

① DBMS 首先对 SQL 语句进行语法检查，然后从数据字典中找到该用户对应的外模式并进行权限检查，如有错误，会将错误信息返回给用户。

② DBMS 从数据字典读取相应的模式定义，并从外模式映像到模式，从中导出概

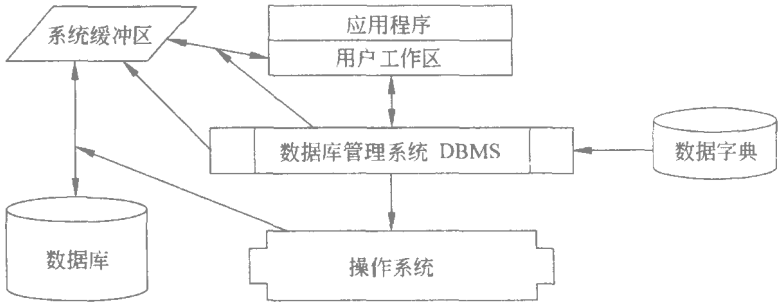


图 1-7 数据库系统工作流程

念模式应该读入哪些记录。

③ DBMS 继续读入内模式描述，并从模式映像到内模式，即将模式下的记录映像到内模式的记录格式，确定应读入哪些物理记录及具体的地址信息（如某个文件中某一数据块里的若干个记录）。

④ DBMS 根据上述地址信息向操作系统发出读取记录的命令。

⑤操作系统执行读取数据的命令，将物理数据库中的数据从硬盘读入系统缓冲区,并将执行结果通知 DBMS。

⑥ DBMS 根据查询命令及数据字典的定义信息，将系统缓冲区中的内容转换成用户所需的记录格式。

⑦ DBMS 将数据记录从系统缓冲区传送到用户工作区。

从上面的流程可以看出，在数据库系统中，DBMS 处于中心地位，它的工作要随时参照数据字典中的定义信息。

另外，DBMS 并不是直接读取数据库中的数据，而是通过操作系统访问数据库中的实际数据。当然，对于操作系统而言，它管理信息的单位不是数据库中的元素（例如表），而是文件。总之，数据库系统是基于操作系统的。对于它们之间的分工，用户不一定需要了解,但对于 DBA 来说，为了做好系统的维护与管理工作，必须清楚这层关系。

1.3 数 据 模 型

1.3.1 数据模型的基本概念

数据模型是模型化数据和信息的工具。数据模型根据其用途不同而具有不同的层次,如图 1-8 所示。

需要全本请访问 www.ertongbook.com