

第 1 章 数据库系统概述

数据库技术诞生于 20 世纪 60 年代末, 历经 30 多年(特别是近 20 年), 数据库技术及其应用得到了迅猛的发展, 从早期的层次数据库和网状数据库, 发展为目前占主流地位的关系数据库, 已形成较为完整的理论体系。数据库技术已成为计算机领域中最重要技术之一, 它是软件学科的一个重要分支。数据库技术的出现使得计算机应用迅速渗透到工农业生产、商业、行政、科学研究、工程技术和国防军事的各个部门, 并伴随着因特网的出现遍布社会的每一个角落, 基于数据库技术的计算机应用已成为计算机应用的主要领域。当前, 数据库的理论研究以及新的数据库系统的研究仍在继续, 分布式数据库、面向对象数据库、数据仓库、对象-关系数据库、多媒体数据库等一批新型数据库已逐渐露出头角, 已正在或即将应用到各自的应用领域。

未来社会是信息化社会, 信息已经变为经济发展的战略资源, 而数据库是信息化社会中资源管理与开发利用的基础, 因此, 数据库的建设规模和应用水平已成为衡量一个国家信息化程度的重要标志。

本章主要介绍数据库管理系统和数据库系统的基本概念、基本组成和基本功能, 并对数据库技术的产生、发展历程进行了详细的论述。

1.1 基本概念

用计算机存储和管理现实世界的信息, 必须研究信息与数据之间的关系, 研究数据在计算机中的存储方式和处理方式。

1.1.1 数据与信息

1. 信息

信息是现实世界在人脑中的抽象反映, 是通过人的感官感知出来并经过人脑的加工而形成的反映现实世界中事物的概念。

信息有许多重要的特征:

- (1) 信息来源于物质和能量;
- (2) 信息是可以感知的;
- (3) 信息是可以存储的;

(4) 信息是可以加工、传递和再生的。

2. 数据

数据是一切文字、符号、声音、图像等有意义的组合。它是描述现实世界中各种具体事物或抽象概念的可存储并具有明确意义的信息。它是用符号记录下来的、可以识别的信息。

数据和信息是分不开的。信息是以数据的形式表示的，即它是以数据为载体而表现的。信息是数据的内涵，数据则是表示信息的一种手段。

事物、信息和数据自始至终贯穿于现实世界、信息世界和计算机世界。

数据库技术所研究的问题就是如何科学地组织、存储和管理数据，如何高效地获取和处理数据。这涉及到数据管理和数据管理两个概念。

3. 数据处理

数据处理（又称为信息处理）是指从某些已知的表示某些信息的数据出发，推导整理出一些新的数据，从而又表示出一些新的信息的过程。它涉及到数据的收集、管理、加工直至产生新信息输出的全过程。

4. 数据管理

数据管理是指数据的收集、整理、组织、存储、维护、检索及传送等操作处理过程。

数据处理和数据管理是互相联系的，数据管理中的各种操作都是数据处理业务的必不可少的基本环节，数据管理技术的好坏，直接影响到数据处理的效率。

1.1.2 数据库、数据库管理系统和数据库系统

1. 数据库 (DataBase, 简称 DB)

数据库是指相互关联的数据集合。它是一组长期存储在计算机内，有组织的、可共享的、具有明确意义的数据集合。

数据库可以人工建立、维护和使用，也可以通过计算机建立、维护和使用。

数据库具有以下几个特点：

(1) 它是具有逻辑关系和确定意义的数据集合。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，可为各种用户共享。

(2) 它是针对明确的应用目标而设计、建立和加载的。

(3) 它表现了现实世界的某些方面。

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

能够对数据库进行有效管理的一组计算机程序称为数据库管理系统。它是位于用户与操作系统之间的一层数据管理软件，是一个通用的软件系统。数据库管理系统通常由语言处理、系统运行控制和系统维护三大部分组成，给用户提供了一个软件环境，允许用户快速方便地建立、维护、检索、存取和处理数据库中的信息。我们熟悉的数据库管理系统有 Visual ForPro、SQL Server、Oracle 等。

3. 数据库系统 (DataBase System, 简称 DBS)

数据库系统是指具有管理和控制数据库功能的计算机系统。它通常由 5 部分组成：硬件系统、数据库、软件支持系统、数据库管理员和用户。它们之间的关系如图 1.1 所示。

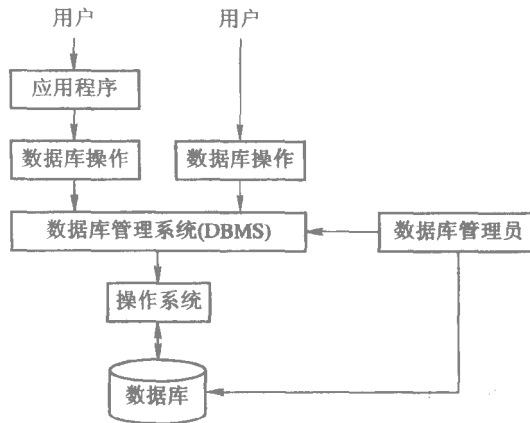


图 1.1 一个简化的数据库系统环境

1.2 数据库技术的产生和发展

数据库技术是数据管理的最新技术，是研究数据库的结构、存储、设计、管理和使用的一门软件学科。数据库技术是在操作系统的文件系统的基础上发展起来的，而且数据库管理系统本身要在操作系统支持下才能工作。

随着计算机硬件和软件的发展，数据管理经历了人工管理、文件系统和数据库系统三个发展阶段。

1.2.1 人工管理阶段

人工管理阶段是指 20 世纪 50 年代中期以前。该时期的计算机应用范围狭窄，主要用于科学计算。在这个时期，计算机硬件发展水平低下，外部存储器只有纸带、卡片、磁带，没有能直接存取的存储设备（如磁盘等）；软件只有汇编语言，还没有开发出操作系统软件，更没有专门的管理数据的软件。数据处理的方式基本上是批处理，数据管理者是人。

这个时期的数据管理的特点是：

- (1) 数据不保存，其主要原因是当时的计算机主要用于科学计算。
- (2) 数据均由应用程序自己管理，没有统一的负责管理数据的专门软件系统。
- (3) 数据不具有独立性，无法进行数据共享。

由于数据是面向应用程序的，不仅数据的逻辑结构和物理结构的设计是由应用程序承担的，而且它们的修改也会引起应用程序的修改，这样必然导致程序员的负担很重，且程序与程序之间将产生大量的冗余数据。

该阶段应用程序和数据之间的关系如图 1.2 所示。

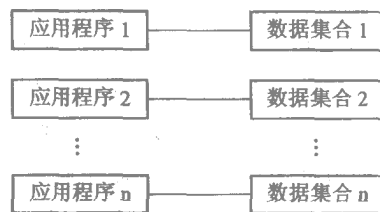


图 1.2 人工阶段应用程序与数据之间的对应关系

1.2.2 文件系统阶段

文件系统阶段是指从 20 世纪 50 年代后期到 60 年代中期。该时期的计算机应用范围逐渐扩大，计算机不仅用于科学计算，而且还大量用于信息管理。计算机硬件有了进一步的发展，出现了磁盘、磁鼓等能直接存取的外存储设备；在软件方面，高级语言和操作系统已经有了完善的产品，并且操作系统中有专门负责管理数据的文件系统功能。数据处理的方式有批处理，也有联机实时处理。数据管理者是计算机。

文件系统管理数据具有如下特点：

(1) 数据可以长期保存在外存储器上，并可以多次进行存取操作。

(2) 程序和数据有了一定的独立性，并分开存放。数据由专门的软件（即文件系统）进行管理，程序和数据之间由软件提供的存取方法进行转换。应用程序与数据之间的联系如图 1.3 所示。

(3) 数据共享性差，数据有较大的冗余。文件是面向应用的，当不同的应用程序中具有部分相同的数据时，也必须建立各自的文件而不能共享相同的数据，这样不仅带来数据的冗余度大，存储空间浪费，而且在进行数据修改时容易造成数据的不一致性。

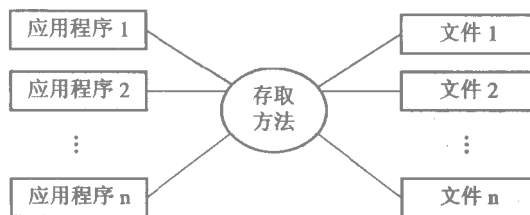


图 1.3 文件系统阶段应用程序与数据之间的对应关系

1.2.3 数据库系统阶段

数据库系统阶段是指从 20 世纪 60 年代后期以来至今。数据管理技术进入数据库系统阶段的标志是 20 世纪 60 年代末期的三件大事：① 1968 年美国 IBM 公司推出层次模型的 IMS(Information Management System) 系统；② 1969 年美国 CODASYL(Conference On Data System Language) 组织发布了 DBTG(Data Base Task Group) 报告，提出网状模型；③ 1970 年美国 IBM 公司的 E. F. Codd 连续发表论文，提出关系模型，奠定了关系数据库的理论基础。

特别是 20 世纪 70 年代以来，计算机应用范围越来越广泛，要求用计算机管理的数据量急剧增长，同时，多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。一方面，计算机硬件发展迅速，硬件价格大幅度下降，出现了大容量磁盘等廉价的内、外存储设备；另一方面，软件价格上升，软件成本增加。为了解决多用户、多应用共享数据的需求，使数据为尽可能多的应用服务，致使数据库技术得到了迅速的发展。

数据库系统管理数据具有如下特点：

(1) 采用特定的数据模型，使数据结构化。数据结构化是数据库和文件系统的本质区别。

在文件系统中，相互独立的文件的记录内部是有结构的，通常采取的是等长或变长的记录格式，但记录之间没有联系，这种数据的独立性只对一个应用而言，仍有局限性，不适应多用户、多应用共享数据的需求。而数据库系统为用户提供一个数据的抽象视图，它能隐藏数据的存储结构和存取方法等细节，并通过数据模型作为实现数据抽象的主要工具，实现了整体数据的结构化，它要求在描述数据时不仅要描述数据本身，还要描述数据之间的联系。

(2) 数据的独立性高。数据库系统提供了三级数据抽象（视图级抽象、概念级抽象和物理级抽象）能力和三种数据库模式（外模式、模式和内模式），实现了数据的物理独立性和逻辑独立性。数据与程序的相互独立，使得可以把数据的定义和描述从应用程序中分离出去，而把数据的存取由数据库管理系统统一进行管理，用户在应用程序中不用考虑存取路径等细节，大大简化了应用程序的编制及应用程序对数据的维护和修改。应用程序和数据之间的联系如图 1.4 所示。

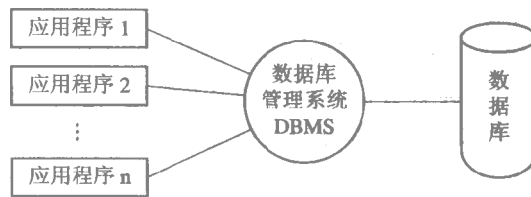


图 1.4 数据库系统阶段应用程序和数据之间的联系

(3) 数据的共享性好，数据冗余度低。数据库系统允许多个用户或多个应用程序同时访问数据库中的相同数据，数据不再面向某个应用，而是面向整个系统，从而支持了数据的共享，节省了存储空间，大大减少了数据冗余，避免了数据之间的不相容性与不一致性。

(4) 为用户提供了方便的用户接口。用户可以使用查询语言（如 SQL）或终端命令对数据库进行访问，也可以借助高级语言（如 C 语言、COBOL 语言等）采用程序方式对数据库进行操作。

(5) 有统一的数据控制功能。数据库系统中的数据由数据库管理系统统一管理，而且管理的是有结构的数据，因此在使用数据时可以有灵活的方式。一个数据库一般都要支持很多应用程序和用户。不同的应用程序和不同的用户对同一个数据库可能有不同的理解，对同一数据库的每一种理解称为这个数据库的一个视图。一个视图可以是一个数据库子集合，也可以是多个数据库的子集按照某种方式构成的虚拟数据库。数据库管理系统提供了定义、维护和操纵视图的机制，使得多个用户可以为他们的应用定义、维护和使用自己的视图。

除此之外，为了适应数据共享的环境，数据库管理系统还必须提供数据控制功能：

数据的安全性：指保护数据，防止不合法使用数据造成数据的泄密和破坏，使每个用户只能按规定对某些数据以某些方式进行访问和处理。

数据的完整性：指数据的正确性、有效性和相容性，即将数据控制在有效的范围内，或要求数据之间满足一定的关系。

并发控制：该机制是用来控制多个事务的并发运行，避免它们之间的相互干扰，保

证每个事务都产生正确的结果。

数据库恢复：在发生某种故障而使数据库当前的状态已经不再正确时，能把数据库恢复到已知为正确的某种状态的功能。

数据库技术是计算机科学技术中发展最快的重要分支之一，它已经成为计算机信息系统和应用系统的重要技术支柱。在短短的 30 多年里，它已从第一代的网状、层次数据库系统，第二代的关系数据库系统，发展到第三代以面向对象模型为主要特征的数据库系统。数据库技术和网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透，互相结合，成为当前数据库技术发展的主要特征。

数据模型是数据库系统的基础。因此，数据库发展阶段的划分应该以数据模型的进展作为主要依据和标志。

• 第一代数据库系统——层次、网状数据库系统。其代表是：① 1969 年 IBM 公司研制的层次模型和数据库管理系统 IMS(Information Management System)；美国数据库系统语言协会 CODASYL 下属的数据库任务组 DBTG 于 20 世纪 60 年代末 70 年代初提出的基于网状结构的方法，它是数据库网状模型的典型代表。

• 第二代数据库系统——关系数据库系统。1970 年，IBM 公司 San Jose 研究室的研究员 E. F. Codd 发表了“大型共享数据库数据的关系模型”论文，提出了数据库的关系模型，开创了数据库关系方法和关系数据理论的研究，为关系数据库技术奠定了理论基础。其典型代表：① IBM San Jose 实验室开发的 System R；② Berkeley 加利福尼亚大学研制的 INGRES。

• 第三代数据库系统。从 20 世纪 80 年代以来，数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长，但传统数据库系统的局限性难以满足新应用的需求。传统数据库系统的局限性主要表现在以下几方面：

(1) 面向机器的语法数据模型。传统数据库中采用的数据模型强调数据的高度结构化，只能存储离散的数据和有限的数据与数据之间的关系，语义表示能力差。它无法表示客观世界中结构复杂、相互联系且语义十分复杂的对象，限制了数据库处理、超文本、图形、图像、CAD 图件、声音等多种复杂对象的能力，更难于处理工程、地理、测绘等领域中的非格式化、非经典数据。

(2) 数据类型简单、固定。

(3) 结构与行为完全分离。传统数据库主要关心数据的独立性及存取数据的效率，是语法数据库，语义表达能力差，难以抽象化地去模拟行为。对象的结构表示可映射到数据库模式，对象的行为特征最多只能由应用程序来表示。

(4) 阻抗失配。它主要是指关系系统中，数据操纵语言和通用程序设计语言之间的失配。

(5) 被动响应。仅能响应和重做用户要求它们做的事情。

(6) 存储、管理的对象有限。仅能存储和管理数据，缺乏知识管理和对象管理的能力。

(7) 事务处理能力较差。仅能支持非嵌套事务，对长事务的响应较慢，且在事务发生故障时恢复比较困难。

有鉴于此，数据库研究人员积极投入对第三代数据库系统的研究。经过大量的研究和讨论，达成共识，第三代数据库系统的基本特征为：支持数据管理、对象管理和知识管理；

必须保持或继承第二代数据库系统的技术；必须对其他系统开放（其开放性表现在：支持数据库语言标准，支持标准网络协议；系统具有良好的可移植性、可连接性、可扩展性和可互操作性）；数据库技术与其他技术相结合（如分布式数据库与分布处理技术相结合；并行数据库与并行处理相结合；多媒体数据库与多媒体技术相结合）等。

1.3 数据库系统结构

从不同的角度考查，可得出不同的数据库系统结构。从数据库管理系统角度来看，数据库系统通常采用三级模式结构；从数据库最终用户的角度来看，数据库系统的结构分为单用户结构、主从式结构、分布式结构和客户/服务器结构。下面我们就一一进行介绍。

1.3.1 数据库系统的三级模式结构

虽然实际的数据库系统软件产品种类繁多，它们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，数据的存储结构各不相同，但从数据库管理系统的角度看，它们在体系结构上通常都具有三级模式的特征（微机上的个别小型数据库系统除外），并提供两级映象功能。

在数据库系统中，用户看到的数据和计算机中存放的数据是两回事，它们之间通过两次映象变换相互联系起来。

1. 三级模式

数据库的三级模式结构是指数据库系统是由外模式、模式和内模式三级构成。

1) 外模式（又称子模式或用户模式）

外模式是数据库用户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，是与某一个应用有关的数据的逻辑表示。它由若干个外部记录类型组成。

用户使用数据操纵语言的语句对数据库进行操作，实际上就是对外模式的外部记录进行操作。用户对数据库的操作，只能与外模式发生联系，按照外模式的结构存储和操纵数据，不必关心模式。

2) 模式（又称逻辑模式或概念模式）

模式是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图。它由若干个概念记录类型组成。

3) 内模式（又称存储模式）

内模式是数据物理结构和存储结构的描述，即是数据在数据库内部的表示方式。它定义所有的内部记录类型、索引和文件的组织方式，以及数据控制方面的细节。

一个数据库中只有一个内模式和一个模式，但可有多个外模式。

数据库的内模式依赖于它的全局逻辑结构，但独立于外模式，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有效的组织，以实现较好的时间和空间效率。

数据按外模式的描述提供给用户，按内模式的描述存储在磁盘中。模式提供了一种约束其他两级的相对稳定的中间观点，它使得两级中的任何一级改变都不受另一级的牵制。

模式位于数据库系统模式结构的中间层，不涉及数据的物理存储细节和硬件环境，也与具体的应用无关。数据库模式以某一种数据模型为基础，综合考虑所有用户的需求，并将这些需求结合成一个逻辑整体。定义模式时不仅要定义数据的逻辑结构（如数据记录由哪些数据项构成，数据项的名字、类型、取值范围等），而且要定义与数据有关的安全性、完整性要求，定义这些数据之间的联系。

数据库的外模式是面向应用程序的，它定义在模式之上，独立于内模式和存储设备。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，通常它是模式的子集。一方面，对模式中的同一数据，在外模式中的结构、类型、长度、保密级别等都可不同。另一方面，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。每个用户只能看见和访问所对应的外模式中的数据，数据库中的其余数据对他们来说是不可见的。

由此可见，数据库系统的三级模式是对数据的三个抽象级别，它把数据的具体组织留给了数据库管理系统去管理，使用户能逻辑地、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与存储方式。

图 1.5 描述了数据库系统的三级模式结构及其映象关系。

为了能够在内部实现这三个抽象层次的联系和转换，数据库系统在这三级模式之间提供了外模式 / 模式映象和模式 / 内模式映象两级映象。

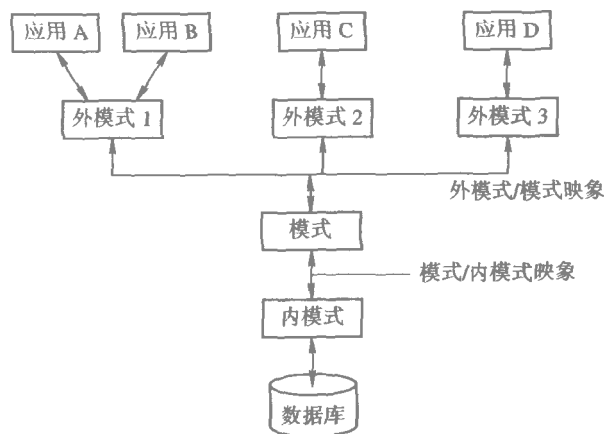


图 1.5 数据库系统的三级模式结构

2. 二级映象

1) 外模式 / 模式映象

外模式 / 模式映象定义了各外模式和模式之间的对应关系。这些映象定义通常包含在各自外模式的描述中。当模式改变时，由数据库管理员对各个外模式 / 模式映象作相应的改变，而外模式仍然保持不变，从而应用程序不必修改，保证了数据的逻辑独立性（指当总体逻辑结构改变时，通过对映象的相应改变而保持局部逻辑结构不变，从而应用程序也可以不必改变）。

2) 模式 / 内模式映象

模式 / 内模式映象定义了数据全局逻辑结构与存储结构之间的对应关系。模式 / 内模式映象是惟一的，该映象定义通常包含在模式描述中。当数据库的存储结构发生改变时，由数据库管理员对模式 / 内模式映象作相应的改变，而使模式保持不变，从而保证了数据的物理独立性（指当数据的存储结构改变时，数据的逻辑结构可以不变，从而应用程序也不必改变）。

1.3.2 数据库系统的体系结构

前面介绍的数据库系统的三级模式结构是从数据库管理系统的角度看的数据库系统结构。若从用户的角度看，数据库系统的体系结构又可分为单用户、主从式结构、分布式结构和客户 / 服务器结构。

1. 单用户数据库系统

该系统是最简单的数据库系统。在单用户数据库系统中，整个数据库系统，包括应用程序、数据库管理系统、数据库等都装在一台计算机上，由一个用户独占，不同的计算机之间不能共享数据。单用户数据库系统和计算机之间的关系如图 1.6 所示。

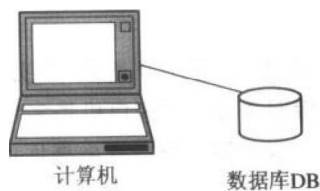


图 1.6 单用户数据库系统和计算机之间的关系

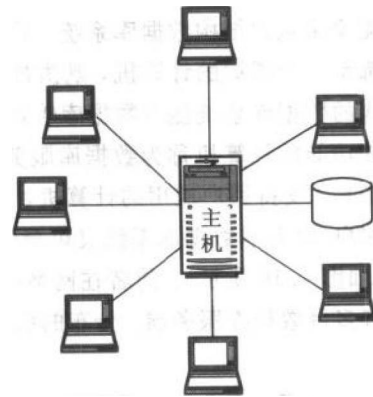


图 1.7 主从式结构的数据库系统

2. 主从式结构的数据库系统

该系统是指一个主机带多个终端的多用户结构的数据库系统。在这种结构下，整个数据库系统，包括应用程序、数据库管理系统、数据库等都集中存放在主机上，所有处理任务都由主机来完成，各个用户通过主机的终端并发地存取数据库中的数据，达到共享数据资源的目的。主从式结构的数据库系统如图 1.7 所示。

3. 分布式结构的数据库系统

该系统是指数据库中的数据在逻辑上是一个整体，但物理地分布在计算机网络的不同结点上。网络中的每个结点（通常是一个通用计算机）都可以独立处理本地数据库中的数据，执行局部应用；也可以同时存取和处理多个异地数据库中的数据，执行全局应用。分布式结构的数据库系统如图 1.8 所示。

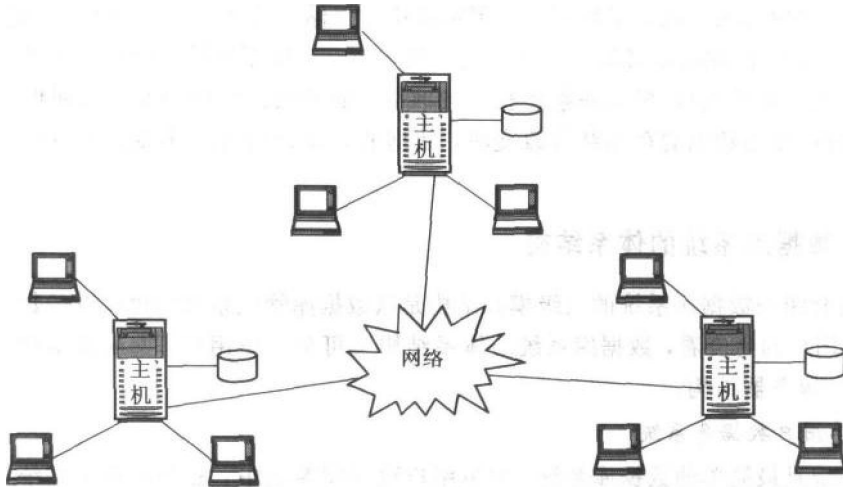


图 1.8 分布式结构的数据库系统

4. 客户/服务器结构的数据库系统

无论主从式结构数据库系统中的主机，还是分布式结构数据库系统中的每个结点机，它们都是一个通用的计算机，既执行数据库管理系统的功能，又执行应用程序。客户/服务器结构的数据库系统能将数据库管理系统功能和应用分开。网络中专门用于执行数据库管理系统功能的计算机称为数据库服务器，简称服务器；将安装数据库管理系统的外围应用开发工具、支持用户应用的计算机，称为客户机。

客户/服务器数据库系统又可分为集中的服务器结构（如图 1.9 所示）和分布的服务器结构（如图 1.10 所示）。前者在网络中仅有一台数据库服务器，而客户机是多台。后者在网络中有多台数据库服务器。分布的服务器结构是客户/服务器与分布式数据库的结合。

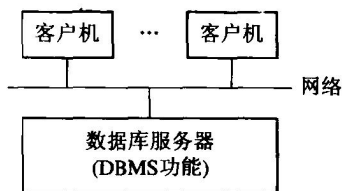


图 1.9 集中的服务器结构

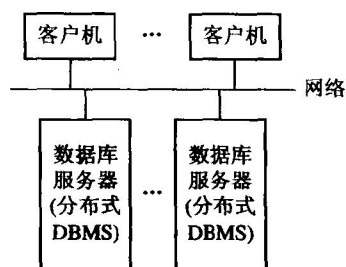


图 1.10 分布的服务器结构

采用客户/服务器结构的数据库系统的特点：

(1) 客户端的用户请求被传送到数据库服务器，数据库服务器进行处理后，只将结果返回给用户（而不是所有数据），从而极大地减少了网络负载，提高了系统的处理速度和性能；

(2) 由于客户与服务器都能在多种不同的硬件和软件平台上运行，并可以采用不同的

数据库开发工具，因此，应用程序的编写更灵活，移植性更好，且软件维护更容易。

1.4 数据库管理系统

数据库管理系统是数据库系统的核心，是用于建立、使用和维护数据库的一组软件。一般情况下，数据库管理系统建立在操作系统的基础上，是位于操作系统与用户之间的一层数据管理软件，负责对数据库进行统一的管理和控制。数据库管理系统分为多个程序模块，每个模块实现数据库系统的一种功能。数据库管理系统不仅承担执行各种应用程序对数据库中的数据的操作命令，还要承担数据库的维护工作，以保证数据库的安全性和完整性。

1.4.1 数据库管理系统的主要功能

基于不同的数据模型的数据库管理系统的功能一般是不同的，即使是同种数据模型的数据库管理系统，它们在用户接口和系统功能等方面经常也是不相同的。虽然不同的数据库管理系统要求的硬件资源、软件环境有所不同，其功能与性能也存在不同的差别，但数据库管理系统的主要功能均包括以下几个方面：

1. 数据库定义功能

DBMS 一般提供数据定义语言来定义构成数据库结构的外模式、模式和内模式，定义两级映射，定义保证数据的完整性约束、保密限制等的约束条件。

2. 数据库操纵功能

DBMS 一般提供数据操纵语言实现对数据库中数据的基本操作，即查找（或检索）、插入、修改和删除操作。

3. 数据库运行控制功能

DBMS 包括对数据库进行并发控制、存取控制（安全性检查）、完整性约束的检查和执行、数据库内部的维护（如索引、数据字典的自动维护）等。这是 DBMS 运行时的核心部分，所有数据库的操作都要在这些控制程序的统一管理下进行，以保证数据库的正确有效。

4. 数据库的建立和维护功能

DBMS 包括数据库初始数据的载入、转换；数据库的转储、恢复；数据库的重组织与重构造等功能。

5. 数据组织、存储和管理

需要在数据库中存放多种数据，如数据字典、用户数据和存取路径等，数据库管理系统负责对这些数据分类进行组织、存储和管理，以确定用何种文件结构和存取方式物理地组织这些数据，如何实现数据之间的联系，以便提高存储空间利用率，提高查找、增加、删除、修改等数据操纵的时间效率。

6. 数据通信接口

它提供数据库管理系统与其他软件系统进行通信的功能

现代数据库管理系统应该具有：友好的用户界面；高级的用户接口；数据查询处理和优化；数据目录和管理；数据的并发控制；数据的恢复功能；数据的安全性和完整性约束检查；数据的访问控制等。

1.4.2 数据库管理系统的组成

根据数据库管理系统所需完成的功能，数据库管理系统通常由以下几部分组成：

1. 数据库语言

该语言是给用户提供的。数据库语言包括两个子语言：数据定义子语言和数据操纵子语言。

如众所周知的 SQL 语言，它是一个集数据定义和数据操纵子语言为一体的典型数据库语言。几乎目前出现的关系数据库系统产品都提供 SQL 语言作为标准数据库语言。数据库语言与数据模型密切相关。基于不同数据模型的数据库系统的语言也不完全相同。

1) 数据定义子语言

数据定义语言 Data Definition Language ,简称 DDL 包括数据库模式定义和数据库存储结构与存取方法定义两方面。数据库模式定义处理程序接收用 DDL 表示的数据库的外模式、模式、存储模式及它们之间的映射的定义，通过各种模式翻译程序负责将它们翻译成相应的内部表示形式，存储到数据库系统中称为数据字典的特殊文件中，作为数据库管理系统存取和管理数据的基本依据。而数据库存储结构和存取方法定义处理程序接收用 DDL 表示的数据库的存储结构和存取方法定义，在存储设备上创建相关的数据库文件，建立起相应物理数据库。

2) 数据操纵子语言

数据操纵语言 Data Manipulation Language ,简称 DML) 用来表示用户对数据库的操作请求，是用户与 DBMS 之间的接口。一般对数据库的主要操作包括：查询数据库中的信息；向数据库插入新的信息；从数据库删除信息；修改数据库中的某些信息等。

DML 通常又分为两类。一类是嵌入主语言（如 C 语言），由于这种 DML 语言本身不能独立使用，故称为宿主型的语言。另一类是交互式命令语言。由于这种 DML 语言本身能独立使用，故又称为自主型或自含型的语言。

2. 数据库管理的例程序

数据库管理例程序随系统不同而各异。一般包括以下几部分：

(1) 语言翻译处理程序。该程序包括 DDL 翻译程序、DML 处理程序、终端查询语言解释程序、数据库控制语言的翻译程序等。

(2) 系统运行控制程序。该程序包括系统初启程序、文件读写与维护程序、存取路径管理程序、缓冲区管理程序、安全性控制程序、完整性检查程序、并发控制程序、事务管理程序、运行日志管理程序和通信控制程序等。

(3) 公用程序。该程序包括定义公用程序和维护公用程序。定义公用程序包括信息格式定义、概念模式定义、外模式定义和保密定义公用程序等。维护公用程序包括数据装入、数据库更新、重组、重构、恢复、统计分析、工作日记、转储和打印公用程序等。

1.4.3 数据库管理系统的工作流程

在数据库系统中，当一个应用程序或用户需要存取数据库中的数据时，应用程序、数据库管理系统、操作系统和计算机硬件等几方面必须协同工作，共同完成用户的请求。在这个较复杂的运行过程中，数据库管理系统起着关键的桥梁作用。

应用程序 A 通过数据库管理系统从数据库中访问一个数据需要经过以下几个步骤：

(1) 用户通过应用程序 A 中的某条 DML 语句向 DBMS 发出从数据库中访问数据记录的命令。

(2) DBMS 接收到该命令后，对该命令进行语法、语义检查，并从数据字典中调用应用程序 A 对应的子模式，检查此应用程序的存取权限，决定是否执行该命令。如果拒绝执行，则向应用程序 A 返回访问失败信息，结束；否则，接着执行。

(3) 在执行该命令后，DBMS 调用相应的模式描述，依据子模式 / 模式映象的定义，确定应读入模式中的哪些记录。

(4) DBMS 调用相应的物理模式描述，依据模式 / 物理模式映象的定义，决定应从哪个文件、用什么存取方式、读入哪个或哪些物理记录。

(5) DBMS 向操作系统发出执行读取所需物理记录的命令。

(6) 操作系统执行读数据的有关操作。

(7) 操作系统将数据从数据库的存储区送到系统缓冲区。

(8) DBMS 依据子模式 / 模式映象的定义，导出应用程序 A 所要读取的记录格式。

(9) DBMS 将数据记录从系统缓冲区传送到应用程序 A 的变量区。

(10) DBMS 向应用程序 A 返回命令执行情况的状态信息。

(11) 应用程序 A 根据返回的状态信息决定是否使用程序变量区中的数据。

图 1.11 表示了应用程序 A 通过数据库管理系统从数据库中访问一个数据的全过程。

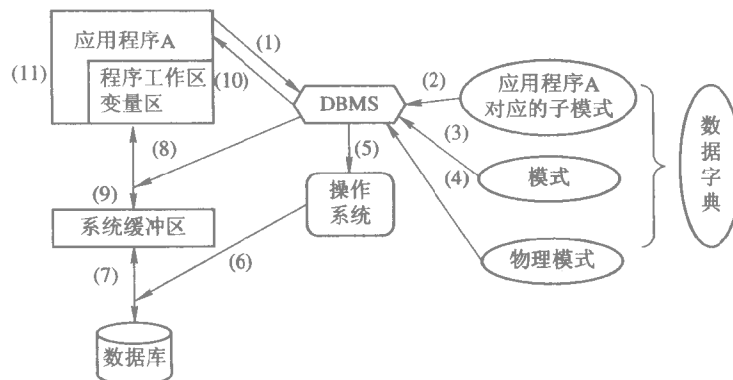


图 1.11 应用程序 A 访问数据库中一个数据的全过程

通过上面对应用程序从数据库中访问一个数据步骤的介绍，说明了数据库管理系统的大致工作过程和在数据库系统中起到的作用。

1.5 数据库系统

数据库系统是指具有管理和控制数据库功能的计算机系统。它通常由 5 部分组成：硬件系统、数据库、软件支持系统、数据库管理员和用户（参见图 1.1）。

1.5.1 数据库

前面已经简单地介绍了这个概念。下面从另一个角度解释它。

数据库是与一个特定组织各项应用有关的全部数据的集合。

数据库通常由两部分组成：物理数据库和描述数据库。

(1) 应用所需要的数据的集合，称为物理数据库（或数据库），它是数据库的主体。

(2) 关于各级数据结构的描述，称为描述数据库，由数据字典管理。

数据字典 (Data Dictionary, 简称 DD) 是数据库系统中各种描述信息和控制信息的集合，是数据库设计和管理的有力工具，又称为“数据库的数据库”。

数据字典的基本内容包括：数据项、组项（若干个数据项的组合）、记录、文件、外模式、模式、内模式、外模式 / 模式映射、模式 / 内模式映射、用户管理信息和数据库控制信息等。它的主要作用是：供数据库管理系统快速查找有关对象的信息；供数据库管理员查询，以掌握整个系统的运行情况；支持数据库设计和系统分析。现在有的大型系统中，把数据字典单独抽出来成一个系统，成为一个软件工具，使之成为一个比数据库管理系统更高级的用户与数据库之间的接口。

1.5.2 硬件系统

硬件系统包括中央处理机、内存、外存、输入 / 输出设备、数据通道等硬件设备。对数据库系统来说，特别要关注内存和外存的容量、I/O 存取速度、通道能力、可支持终端数和性能稳定性等指标，有些应用中还需要考虑系统支持联网能力等。

1.5.3 软件支持系统

软件支持系统包括数据库管理系统、操作系统、各种宿主语言和实用程序等。数据库管理系统是管理数据库的软件，它必须在操作系统的支持下才能工作。为了开发应用程序，还需要各种宿主语言，并且与数据库管理系统要有良好的接口。

1.5.4 数据库管理员

数据库管理员 (DataBase Administrator, 简称为 DBA) 是负责全面管理和控制数据库系统正常运行的人员，他承担着创建、监控和维护整个数据库结构的责任。DBA 的素质在一定程度上决定了数据库应用的水平，所以他们是数据库系统最重要的人员。

DBA 的主要职责包括：

(1) 确定数据库中的内容和结构。DBA 根据用户的需求，创建数据库的结构，确定数据库中存放的内容。DBA 必须参与数据库设计的全过程，与用户、应用程序员和系统分析

员密切结合，设计概念模式、数据库逻辑模式以及各个用户的外模式。

(2) 确定数据库的存储结构和存取方法。DBA 决定数据库的存储结构和存取方法，设计数据库的内模式。

(3) 重新构造数据库。根据用户的需要，DBA 能对数据库进行较大的修改，包括修改内模式或模式。

(4) 维护数据库，定义数据的安全性要求和完整性约束条件。DBA 负责维护数据库，为了保证数据库中的数据的安全性，数据库中的内容对 DBA 应该是有封锁的。同时根据需要，DBA 应编写完整性规则，以监督数据库的运行。

(5) 监控数据库的使用和运行。DBA 负责监视数据库系统的运行情况，分析数据库系统的性能，及时处理运行过程中出现的问题，控制和授权不同用户访问数据库的权限，定期对数据库进行重新组织。

(6) 转储和恢复数据库。为了减少硬件、软件或人为故障对数据库系统的破坏，DBA 必须提供周期性的转储数据，填写维护日志文件和恢复策略。

通过前面介绍，我们已经知道，在一台能够满足数据库应用开发需求的计算机上先安装一个具体的数据库管理系统，而它必须安装在一个具体的操作系统之上，然后开发人员根据用户需求开发一个具体的应用系统，从而形成一个完整的数据库系统。数据库管理员的任务就是管理和维护这个数据库系统进行正常运行。

在具备了硬件环境、操作系统等其他系统软件和某个具体的数据库管理系统的情况下，对数据库应用开发人员来说，就是如何使用这个环境来表达用户的要求，并转换成有效的数据库结构，构成较优的数据库模式等，这就涉及数据库设计问题。

本书主要介绍数据库系统的基本概念、基本原理、基本方法和应用技术，并对数据库设计中涉及的问题也进行了详细介绍。

小 结

数据库系统是一个复杂的系统，它是采用了数据库技术的计算机系统，又是一个实际可运行的，按照数据库方法存储、维护和向应用系统提供数据支持的系统。它由硬件系统、数据库、软件支持系统、数据库管理员和用户组成。数据库管理系统是位于应用程序（或用户）和操作系统之间的一层管理软件，它是数据库系统的核心。数据库应用系统是包含了数据库管理系统、数据库和用户的应用程序的系统。

本章介绍了数据库、数据库管理系统、数据库系统、数据库的三级模式的基本概念，叙述了数据库技术的产生和发展，从两个不同角度讨论了数据库系统的结构，详细论述了数据库管理系统的主要功能、基本组成和 workflows，并对数据库系统的组成做了全面的介绍。

习 题 1

1. 试述下列概念：
 - (1) 数据；
 - (2) 数据库；
 - (3) 数据库管理系统；
 - (4) 数据库系统；
 - (5) 内模式；
 - (6) 模式；
 - (7) 外模式；
 - (8) 数据的物理独立性；
 - (9) 数据的逻辑独立性；
 - (10) 数据字典。
2. 数据库系统与文件系统的主要区别是什么？
3. 试述数据库系统的特点和结构。
4. 试述数据库管理系统的主要功能和组成。
5. 试述数据库管理员的主要职责是什么？
6. 什么是数据库系统的核心？
7. 数据管理技术经历了哪几个阶段？
8. 指出下列缩写的含义：

DML, DBMS, DDL, DBS, DBA, DD
9. 试述数据库系统的三级模式结构，并解释数据库为什么要采用这种结构。
10. 请描述应用程序访问数据库的全过程，体会数据库管理系统在数据库系统中的核心作用。

第 2 章 数据模型

模型是对现实世界特征的模拟和抽象，而数据模型(Data Model)是模型的一种，它是对现实世界数据特征的抽象。在数据库中，用数据模型这个工具来抽象、表示和处理现实世界中的数据和信息。

现有的数据模型可分为三类：基于对象的逻辑模型、基于记录的逻辑模型和物理模型。

基于对象的逻辑模型用于在逻辑层和视图层上描述数据。其特点是可提供灵活的结构组织能力，允许显式定义数据约束。几种较著名的逻辑模型是：实体-联系模型、面向对象的模型、语义数据模型和功能数据模型。

基于记录的逻辑模型也用于在逻辑层和视图层上描述数据。与基于对象的数据模型不同，基于记录的模型既可用于定义数据库的全局逻辑结构，又可用于提供关于实现的高层描述。

基于记录的逻辑模型的特点是它用一些固定格式的记录来描述数据库结构。每个记录类型定义了固定数目的字段（或属性），通常每个字段的长度也是固定的，使用定长记录可以简化数据库的物理层实现。这种简单性同许多基于对象的模型形成对照，后者丰富的结构常常会导致物理层的记录变长。基于记录的模型中广为接受的是关系模型、网状模型和层次模型。

物理数据模型用于在最底层上描述数据。与逻辑数据模型不同，实际使用中的物理数据模型较少。常用的是一致化模型和框架存储模型。

数据模型是数据库系统的基础，各种机器上实现的 DBMS 软件都是基于某种数据模型的。一般地讲，数据模型是严格定义的一组概念的集合。这些概念精确地描述了系统的静态特性、动态特性和完整性约束条件。因此数据模型通常由数据结构、数据操作和完整性约束三部分组成。

2.1 实体-联系模型

不同的数据模型提供给我们模型化数据和信息的工具不同。根据模型应用的不同目的，可以将这些模型划分为两个层次：一类模型是概念模型，也称信息模型，它是按用户的观点来对数据和信息建模，主要用于数据库设计；另一类模型是数据模型，主要包括网状模型、层次模型、关系模型等，它是按计算机系统的观点对数据建模，主要用于 DBMS