

第 1 章 绪论

【学习目的与要求】

数据库技术是计算机领域中的重要技术之一，是数据管理的最新技术，目前已形成相当规模的理论体系和实用技术。本章主要讲述数据库的有关概念，要理解并熟练掌握数据库的定义，掌握数据库管理系统，了解数据库的发展和每个发展阶段的特点，理解数据库的模式。

数据库技术是 20 世纪 60 年代开始兴起的一门信息管理自动化的新兴学科，是计算机科学中的一个重要分支。随着计算机应用的不断发展，在计算机应用领域中，数据处理越来越占主导地位，数据库技术的应用也越来越广泛。数据库是数据管理的产物，数据管理是数据库的核心任务，内容包括对数据的分类、组织、编码、储存、检索和维护。随着计算机硬件和软件的发展，数据库技术也不断地发展。经过 50 年左右的发展，数据库技术已成为信息系统的核心和基础。目前，它已形成较为完整的理论体系和实用技术。

1.1 数据库的基本概念

数据库技术产生于 20 世纪 60 年代末 70 年代初，聚集了数据处理精华的思想，是管理信息最先进的工具。使用数据库方法管理数据，可以保证数据的安全性、完整性和共享性。在学习数据库知识之前，首先介绍一些数据库最常用的术语和基本概念。

1. 数据、信息及数据处理

(1) 数据

数据 (data) 是描述事物的符号记录，也是数据库中存储的基本对象。数据不仅可以是数值，还可以是文字、图形、符号、声音和图像等。例如，描述一个学生的信息，可用一组数据表示出来。例如，王明，男，1992.12.19，江苏，信息管理与信息系统系。这些符号被赋予了特定的语义，具体描述了学生的特征，因此也就具有了传递信息的功能。

(2) 信息

信息 (information) 是人们用来反映客观世界而记录下来可以鉴别的物理符号。ISO 对信息的定义是信息是对人有用的，影响人们行为的数据。因此信息是具有一定含义的数据，是加工处理后的数据，是对决策有价值的依据。

信息与数据的关系可看做原材料和产品的关系。信息是向人们提供关于现实世界新的事实的知识，数据则是载荷信息的物理符号，两者缺一不可，但又有一定的区别。信息能更直接反映现实的概念，而数据则是信息的具体表现；信息不随载荷它的物理载体而改变，数据

则不然，它在计算机化的信息系统中往往和计算机系统有关。两者也可不断转换。

(3) 数据处理

数据处理 (data processing) 是利用相应的技术和设备对各种数据进行加工的过程。数据处理是对数据的采集、存储、检索、加工、变换和传输的过程。数据处理的基本目的是从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的人们来说是有价值、有意义的数据。数据处理是系统工程和自动控制的基本环节，贯穿于社会生产和社会生活的各个领域。数据处理技术的发展及其应用的广度和深度，极大地影响着人类社会发展的进程。

2. 数据库

数据库 (data base, DB) 是一个长期存储在计算机内的、有组织的、可共享的、统一管理的数据集合。它是一个按数据结构来存储和管理数据的计算机软件系统。数据库的概念实际包括两层意思：数据库是一个实体，它是能够合理保管数据的仓库，用户在该仓库中存放要管理的事务数据，“数据”和“库”两个概念结合成为数据库；数据库是数据管理的新方法和技术，它更适合组织数据、更方便维护数据、更严密控制数据和更有效利用数据。

3. 数据库管理系统

数据库管理系统 (database management system, DBMS) 是一种操纵和管理数据库的系统软件，位于用户与操作系统之间 (图 1.1)，它是数据库系统的核心组成部分，用于建立、使用和维护数据库。它对数据库进行统一的管理和控制，以保证数据库的安全性和完整性。用户通过数据库管理系统访问数据库中的数据，数据库管理员也通过数据库管理系统进行数据库的维护工作。它可使多个应用程序和用户使用不同的方法，在同一时刻或不同时刻去建立、修改和查询数据库。数据库管理系统提供数据定义语言 (data definition language, DDL) 与数据操作语言 (data manipulation language, DML)，供用户定义数据库的模式结构与权限约束，实现对数据的追加、删除等操作。

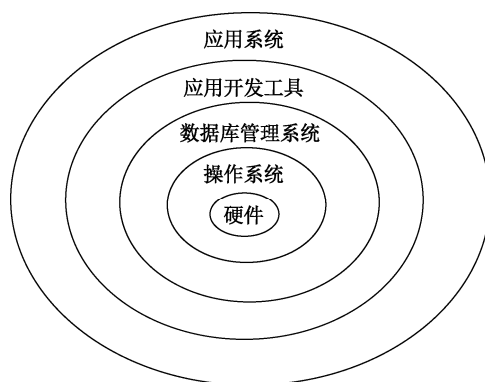


图 1.1 数据库管理系统在数据库系统中的地位

4. 数据库系统

数据库系统 (data base system, DBS) 通常由硬件、软件、数据库和数据管理员组成。其软件主要包括操作系统、各种宿主语言、实用程序及数据库管理系统。数据库由数据库管理

系统统一管理，数据的插入、修改和检索均要通过数据库管理系统进行。数据管理员负责创建、监控和维护整个数据库，使数据能被任何有权使用的人有效使用。数据库管理员一般由业务水平较高、资历较深的人员担任。数据库系统如图 1.2 所示。

数据库系统的个体含义是指一个具体的数据库管理系统软件和用它建立起来的数据库；它的学科含义是指研究、开发、建立、维护和应用数据库系统所涉及的理论、方法、技术所构成的学科。在这一含义下，数据库系统是软件研究领域的一个重要分支，常称为数据库领域。

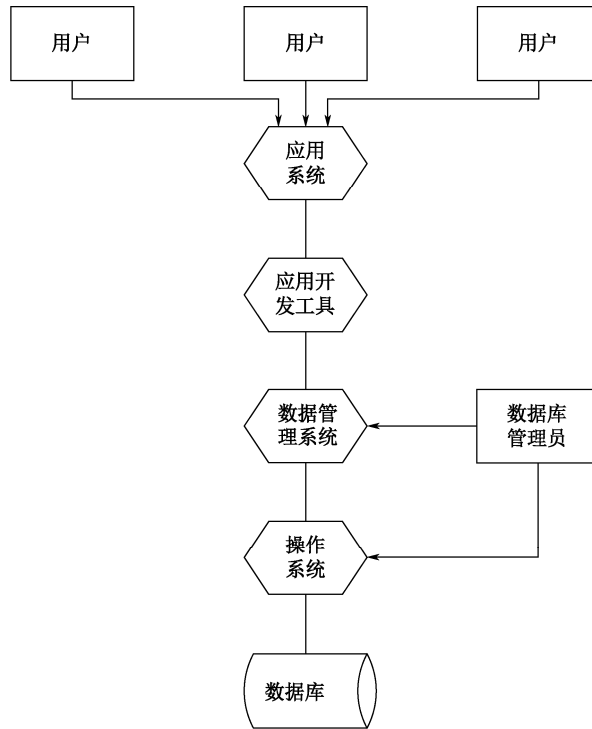


图 1.2 数据库系统

数据库系统是为适应数据处理的需要而发展起来的一种较为理想的数据处理的核心机构。计算机的高速处理能力和大容量存储器提供了实现数据管理自动化的条件。

数据库研究跨越于计算机应用、系统软件和理论研究 3 个领域，其中计算机应用促进新系统的研制开发，为新系统带来新的理论研究，而理论研究又对前两个领域起着指导作用。数据库系统的出现是计算机应用的一个里程碑，它使得计算机应用从以科学计算为主转向以数据处理为主，从而使计算机得以在各行各业乃至家庭中普遍使用。在它之前的文件系统虽然也能处理持久数据，但是文件系统不提供对任意部分数据的快速访问，而这对数据量不断增大的应用来说是至关重要的。为了实现任意部分数据的快速访问，就要研究许多优化技术。这些优化技术往往很复杂，是普通用户难以实现的，所以就由系统软件（数据库管理系统）来完成，而提供给用户的是简单易用的数据库语言。由于对数据库的操作都由数据库管理系统完成，所以数据库就可以独立于具体的应用程序而存在，从而数据库又可以为多个用户所共享。因此，数据的独立性和共享性是数据库系统的重要特征。数据共享节省了大量人力物力，为数据库系统的广泛应用奠定了基础。数据库系统的出现使得普通用户能够方便地

将日常数据存入计算机，并在需要的时候快速访问它们，从而使计算机走出科研机构进入各行各业、进入家庭。

数据库系统有大小之分，大型数据库系统有 SQL Server、Oracle、DB2 等，中小型数据库系统有 FoxPro、Access 等。

数据库系统一般由 4 个部分组成。

1) 数据库：指长期存储在计算机内的、有组织、可共享的数据的集合。数据库中的数据按一定的数学模型组织、描述和存储，具有较小的冗余、较高的数据独立性和易扩展性，并可为各种用户共享。

2) 硬件：构成计算机系统的各种物理设备，包括存储所需的外部设备。硬件的配置应满足整个数据库系统的需要。

3) 软件：包括操作系统、数据库管理系统及应用程序。数据库管理系统是数据库系统的核心软件，是在操作系统的支持下工作，解决如何科学地组织和存储数据，如何高效获取和维护数据的系统软件。其主要功能包括数据定义功能、数据操纵功能、数据库的运行管理和数据库的建立与维护。

4) 人员：主要有 4 类。第一类为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，他们和用户及数据库管理员一起确定系统的硬件配置，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、数据库各级模式的设计。第二类为应用程序员，其负责编写使用数据库的应用程序。这些应用程序可对数据进行检索、建立、删除或修改。第三类为最终用户，他们利用系统的接口或查询语言访问数据库。第四类用户是数据库管理员 (data base administrator, DBA)，其负责数据库的总体信息控制。数据库管理员的具体职责包括：具体定义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，定义数据库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，负责数据库的性能改进、数据库的重组和重构，以提高系统的性能。

1.2 数据库的发展与应用前景

1.2.1 数据库的产生与发展

数据管理的水平是和计算机硬件、软件的发展相适应的。随着计算机技术的发展，人们的数据管理技术经历了 4 个阶段的发展：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

20 世纪 50 年代中期以前，人们运用常规的手段从事记录、存储和对数据加工，也就是利用纸张来记录和利用计算工具（算盘、计算机尺）来进行计算，并主要使用人的大脑来管理和利用这些数据。而早期的计算机主要用于数值计算，也无管理数据的软件，因此从计算机内记录的数据上来看，数据量小，数据无结构。用户直接管理，并且数据之间缺乏逻辑组织，数据仅依赖特定的应用，缺乏独立性。在硬件方面，计算机的外存储器（以下简称“外存”）只有磁带、卡片、纸带，没有磁盘等直接存取的存储设备，存储量非常小；在软件方面，

没有操作系统，没有高级语言，数据处理的方式是批处理，即机器一次处理一批数据，直到运算完成为止，然后才能进行另外一批数据的处理，中间不能被打断，原因是此时的外存（如磁带、卡片等）只能顺序输入。人工管理阶段的数据具有以下几个特点。

1) 数据不保存。由于当时计算机主要用于科学计算，数据保存上并不做特别要求，只是在计算某一个课题时将数据输入，用完就退出，对数据不做保存，有时对系统软件也是这样。

2) 数据不具有独立性。数据是作为输入程序的组成部分，即程序和数据是一个不可分割的整体，数据和程序同时提供给计算机运算使用。对数据进行管理，就像现在的操作系统可以以目录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构，也要规定数据的物理结构，程序员对存储结构、存取方法及输入输出的格式有绝对的控制权，要修改数据必须修改程序。例如，要对 100 组数据进行同样的运算，就要给计算机输入 100 个独立的程序，因为数据无法独立存在。

3) 数据不共享。数据是面向应用的，一组数据对应一个程序。不同应用的数据之间是相互独立、彼此无关的，即使两个不同应用涉及相同的数据，也必须各自定义，无法相互利用，互相参照。数据不但高度冗余，而且不能共享。

4) 由应用程序管理数据。数据没有专门的软件进行管理，需要应用程序自己进行管理，应用程序中要规定数据的逻辑结构和设计物理结构（包括存储结构、存取方法、输入和输出方式等），因此程序员负担很重。

综上所述，这一数据管理阶段为无管理阶段。以高校信息管理为例，人工管理阶段程序与数据文件的对应关系如图 1.3 所示。



图 1.3 人工管理阶段程序与数据文件的对应关系

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期，数据管理发展到文件系统阶段。此时的计算机不仅用于科学计算，还大量用于管理。外存有了磁盘等直接存取的存储设备。在软件方面，操作系统中已有了专门的数据管理软件，称为文件系统。从处理方式上讲，不仅有了文件批处理，而且能够联机实时处理，联机实时处理是指在需要的时候随时从存储设备中查询、修改或更新，因为操作系统的文件管理功能提供了这种可能。

这一阶段的数据管理技术得益于计算机的处理速度和存储能力的惊人提高，这一时期的数据处理系统是把计算机中的数据组织成相互独立的被命名的数据文件，并可按文件的名称来进行访问，对文件中的记录进行存取的数据管理技术。数据可以长期保存在计算机外存上，可以对数据进行反复处理，并支持文件的查询、修改、插入和删除等操作，这就是文件系统。以高校信息管理为例，文件系统阶段程序与数据文件的对应关系如图 1.4 所示。

文件系统阶段的特点如下。

1) 数据长期保留。数据可以长期保留在外存上反复处理，即可以经常有查询、修改和删除等操作，所以计算机大量用于数据处理。

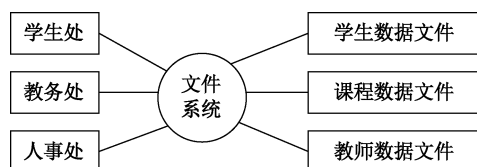


图 1.4 文件系统阶段程序与数据文件的对应关系

2) 数据的独立性。由于有了操作系统，利用文件系统进行专门的数据管理，使得程序员可以集中精力在算法设计上，而不必过多地考虑细节。在保存数据时，只需给出保存指令，而不必所有的程序员都精心设计一套程序，控制计算机物理地实现保存数据。在读取数据时，只要给出文件名，而不必知道文件的具体存放地址。文件的逻辑结构和物理存储结构由系统进行转换，程序与数据有了一定的独立性。数据的改变不一定要引起程序的改变。例如，保存的文件中有 100 条记录，使用某一个查询程序，当文件中有 1000 条记录时，仍然使用保留的这一个查询程序。

3) 可以实时处理。由于有了直接存取设备，也有了索引文件、链接存取文件、直接存取文件等，所以既可以采用顺序批处理，又可以采用实时处理方式。数据的存取以记录为基本单位。

虽然文件系统阶段比人工管理阶段有了很大的改进，但这种方法仍有很多缺点，主要如下所示。

1) 数据共享性差，冗余度大。当不同的应用程序所需的数据有部分相同时，仍需建立各自的独立数据文件，而不能共享相同的数据。因此，数据冗余大，空间浪费严重。并且相同的数据重复存放，各自管理，当相同部分的数据需要修改时比较麻烦，稍有不慎，就造成数据的不一致。例如，学籍管理需要建立包括学生的姓名、班级、学号等数据的文件。这种逻辑结构和学生成绩管理所需的数据结构是不同的。在学生成绩管理系统中，进行学生成绩排列和统计，程序需要建立自己的文件，除了特有的语文成绩、数学成绩、平均成绩等数据外，还要有姓名、班级等与学籍管理系统的文件相同的数据。数据冗余是显而易见的，此外当有本校学生转学离开或外校学生转来时，两个文件都要修改，否则，就会出现有某个学生的成绩，却没有该学生学籍的情况，反之亦然。如果系统庞大，则会牵一发而动全身，一个微小的变动引起一连串的变动，利用计算机管理的规模越大，问题就越多，常常发生实际情况是这样，而从计算机中得到的信息却是另一回事的事件。

2) 数据和程序缺乏足够的独立性。文件中的数据是面向特定应用的，文件之间是孤立的，不能反映现实世界事物之间的内在联系。例如，在上面的学籍文件与成绩文件之间没有任何联系，计算机无法知道两个文件中的哪两条记录是针对同一个人的。要对系统进行功能的改变是很困难的。在上面的例子中，要将学籍管理和成绩管理从两个应用合并成一个应用中，则需要修改原来的某一个数据文件的结构，增加新的字段，还需要修改程序，后果就是浪费时间和重复工作。此外，应用程序所用的高级语言的改变，也将影响到文件的数据结构。例如，BASIC 语言生成的文件，COBOL 语言就无法如同是自己的语言生成的文件一样顺利地使用。总之数据和程序之间缺乏足够的独立性是文件系统的一个大问题。

文件管理系统在数据量非常庞大的情况下，已经不能满足需要。美国在 20 世纪 60 年代进行阿波罗计划的研究。阿波罗飞船由约 200 万个零部件组成，分散在世界各地制造。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洛克威尔 (Rockwell) 公司曾研制

了一个计算机零件管理系统。系统共用了 18 盘磁带，虽然可以工作，但效率极低，维护困难。18 盘磁带中的 60% 是冗余数据。这个系统一度成为实现阿波罗计划的最大障碍。应用的需要推动了技术的发展。文件管理系统面对大量数据时的困境促使人们去研究新的数据管理技术，数据库技术便应运而生。例如，最早的数据库管理系统之一 IMS 就是上述的洛克威尔公司在实现阿波罗计划中与 IBM 公司合作开发的，从而保证了阿波罗飞船 1969 年顺利登月。

3. 数据库系统阶段

20 世纪 60 年代后期，计算机性能得到进一步提高，更重要的是出现了大容量磁盘，存储容量大大增加且价格下降。在此基础上，才有可能克服文件系统管理数据时的不足，从而满足和解决实际应用中多个用户、多个应用程序共享数据的要求，从而使数据能为尽可能多的应用程序服务，这就出现了像数据库这样的数据管理技术。数据库的特点是数据不再只针对某一个特定的应用，而是面向全组织，具有整体的结构性，共享性高，冗余度减小，具有一定的程序与数据之间的独立性，并且对数据进行统一的控制。以高校信息管理为例，数据库系统阶段程序与数据文件的对应关系如图 1.5 所示。

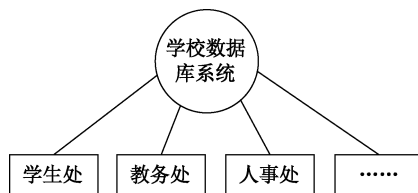


图 1.5 数据库系统阶段程序与数据文件的对应关系

数据库系统的目标是解决数据冗余问题，实现数据独立性，实现数据共享并解决由于数据共享而带来的数据完整性、安全性及并发控制等一系列问题。为实现这一目标，数据库的运行必须有一个软件系统来控制，这个系统软件称为数据库管理系统。数据库管理系统将程序员进一步解脱出来，就像当初操作系统将程序员从直接控制物理读写中解脱出来一样。程序员此时不需要再考虑数据中的数据是不是因为改动而造成不一致，也不用担心由于应用功能的扩充，而导致程序重写和数据结构重新变动。在这一阶段，数据管理具有下面的优点。

1) 数据结构化。数据结构化是数据库系统与文件系统的根本区别。在文件系统中，相互独立的文件的记录内部是有结构的，传统文件的最简单形式是等长同格式的记录集合。这样就可以节省许多储存空间。数据的结构化是数据库的主要特征之一。这是数据库与文件系统的根本区别。至于这种结构化是如何实现的，则与数据库系统采用的数据模型有关，后面会有较详细的描述。

2) 数据共享性高，冗余度小，易扩充。数据库从整体的观点来看待和描述数据，数据不再是面向某一应用，而是面向整个系统。这样就减小了数据的冗余，节约了存储空间，缩短了存取时间，避免了数据之间的不相容和不一致。对数据库的应用可以很灵活，面向不同的应用，且存取相应的数据库的子集。当应用需求改变或增加时，只要重新选择数据子集或者加上一部分数据，便可以满足更多更新的要求，也就是保证了系统的易扩充性。

3) 数据独立性高。数据库提供数据的存储结构与逻辑结构之间的映像或转换功能，使得当数据的物理存储结构改变时，数据的逻辑结构可以不变，从而程序也不用改变。这就是数据与程序的物理独立性。也就是说，程序面向逻辑数据结构，不去考虑物理的数据存放形式。数据库可以保证数据的物理改变不引起逻辑结构的改变。

数据库还提供了数据的总体逻辑结构与某类应用所涉及的局部逻辑结构之间的映像或转

换功能。当总体的逻辑结构改变时，局部逻辑结构可以通过这种映像的转换保持不变，从而程序也不用改变，这就是数据与程序的逻辑独立性。举例来讲，在进行学生成绩管理时，姓名等数据来自于数据的学籍部分，成绩来自于数据的成绩部分，经过映像组成局部的学生成绩，由数据库维持这种映像。当总体逻辑结构改变时，如学籍和成绩数据的结构发生了变化，数据库为这种改变建立一种新的映像，就可以保证局部数据——学生数据的逻辑结构不变，程序是面向这个局部数据的，所以程序就无须改变。

4)统一的数据管理和控制功能，包括数据的安全性控制、数据的完整性控制及并发控制、数据库恢复。数据库是多用户共享的数据资源，对数据库的使用经常是并发的。为保证数据的安全可靠和正确有效，数据库管理系统必须提供一定的功能来保证。

数据库的安全性是指防止非法用户非法使用数据库而提供的保护。例如，不是学校的成员不允许使用学生管理系统，学生允许读取成绩但不允许修改成绩等。

数据的完整性是指数据的正确性和兼容性。数据库管理系统必须保证数据库的数据满足规定的约束条件，常见的有对数据值的约束条件。例如，在建立上面的例子中的数据库时，数据库管理系统必须保证输入的成绩值大于0，否则，系统将发出警告。

数据的并发控制是多用户共享数据库必须解决的问题。要说明并发操作对数据的影响，必须首先明确，数据库是保存在外存中的数据资源，而用户对数据库的操作是先读入内存操作，修改数据时，是在内存中修改读入的数据副本，然后再将这个副本写回到储存的数据库中，以实现物理的改变。

从文件系统到数据库系统，标志着数据管理技术质的飞跃。20世纪80年代后不仅在大、中型计算机上实现并应用了数据管理的数据库技术，如 Oracle、Sybase、Informix 等，在微型计算机上也可使用数据库管理软件，如常见的 Access、FoxPro 等软件，使数据库技术得到广泛应用和普及。上述3个阶段数据管理技术比较如表 1.1 所示。

表 1.1 3个阶段数据管理技术比较

	人工管理	文件管理	数据库管理
应用背景	科学计算	科学计算、管理	大规模管理
硬件背景	无直接存储存取设备	磁盘、磁鼓	大容量磁盘
软件背景	没有操作系统	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、分布处理、批处理
数据的管理者	用户（程序员）	文件系统	数据库系统
数据的针对者	特定应用程序	面向某一应用	面向整性应用
数据的共享性	无共享	共享差，冗余大	共享好，冗余小
数据的独立性	无独立性	独立性差	独立性好
数据的结构化	无结构	记录有结构，整体无结构	整体结构化
数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

4. 高级数据库阶段

20世纪80年代以来关系数据库理论日趋完善，逐步取代网状和层次数据库占领了市场，并向更高阶段发展。目前数据库技术已成为计算机领域中最重要技术之一，它是软件科学中的一个独立分支，正在朝分布式数据库、数据库机、知识库系统、多媒体数据库方向发展。

特别是现在的数据仓库和数据挖掘技术的发展，大大推动了数据库向智能化和大容量化的发展趋势，充分发挥了数据库的作用。

1.2.2 数据库的应用前景

数据、计算机硬件和数据库应用，这三者推动着数据库技术与系统的发展。数据库要管理的数据的复杂度和数据量都在迅速增长；计算机硬件平台的发展仍在继续；数据库应用迅速向深度、广度扩展。尤其是互联网的出现，极大地改变了数据库的应用环境，向数据库领域提出了前所未有的技术挑战。这些因素的变化推动着数据库技术的进步，出现了一批新的数据库技术，如 Web 数据库技术、并行数据库技术、数据仓库与联机分析技术、数据挖掘与商务智能技术、内容管理技术、海量数据管理技术等。

1.3 数据库管理系统

1.3.1 数据库管理系统的主要功能

目前有许多数据库产品，如 Oracle、Sybase、Informix、Microsoft SQL Server、Microsoft Access、Visual FoxPro 等，其各有自己独特的功能，在数据库市场上占有一席之地。随着新型数据模型及数据管理实现技术的推进，可以预期数据库管理系统软件的性能还将更新和完善，应用领域也将进一步拓宽。但是，不管是功能强大的大型数据库管理系统还是功能相对弱小的数据库管理系统，每个数据库管理系统都应该具有如下主要功能。

(1) 数据定义

数据库管理系统提供数据定义语言，供用户定义数据库的三级模式结构、两级映像及完整性约束和保密限制等约束。数据定义语言主要用于建立、修改数据库的库结构。数据定义语言所描述的库结构仅仅给出了数据库的框架，数据库的框架信息被存放在数据字典（data dictionary）中。

(2) 数据操作

数据库管理系统提供数据操作语言，供用户实现对数据的追加、删除、更新、查询等操作。

(3) 数据库的运行管理

数据库的运行管理功能是数据库管理系统的运行控制、管理功能，包括多用户环境下的并发控制、安全性检查和存取限制控制、完整性检查和执行、运行日志的组织管理、事务的管理和自动恢复，即保证事务的原子性。这些功能保证了数据库系统的正常运行。

(4) 数据组织、存储与管理

数据库管理系统要分类组织、存储和管理各种数据，包括数据字典、用户数据、存取路径等，需要确定以何种文件结构和存取方式在存储级上组织这些数据，如何实现数据之间的联系。数据组织和存储的基本目标是提高存储空间利用率，选择合适的存取方法提高存取效率。

(5) 数据库的保护

数据库中的数据是信息社会的战略资源，所以数据的保护至关重要。数据库管理系统对

数据库的保护通过 4 个方面来实现：数据库的恢复、数据库的并发控制、数据库的完整性控制、数据库安全性控制。数据库管理系统的其他保护功能还有系统缓冲区的管理及数据存储的某些自适应调节机制等。

(6) 数据库的维护

这一部分包括数据库的数据载入、转换、转储、数据库的重组重构及性能监控等功能，这些功能分别由各个使用程序来完成。

(7) 通信

数据库管理系统具有与操作系统的联机处理、分时系统及远程作业输入的相关接口，负责处理数据的传送。对网络环境下的数据库系统，还应该包括数据库管理系统与网络中其他软件系统的通信功能及数据库之间的互操作功能。

1.3.2 数据库管理系统的组成

按照功能划分，数据库管理系统通常由以下 6 个部分组成。

(1) 模式翻译

模式翻译提供数据定义语言，用它书写的数据库模式被翻译为内部表示，数据库的逻辑结构、完整性约束和物理储存结构保存在内部的数据字典中。数据库的各种数据操作（如查找、修改、插入和删除等）和数据库的维护管理都是以数据库模式为依据的。

(2) 应用程序的编译

应用程序的编译把包含着访问数据库语句的应用程序，编译成在数据库管理系统支持下可运行的目标程序。

(3) 交互式查询

交互式查询提供易使用的交互式查询语言，如 SQL。数据库管理系统负责执行查询命令，并将查询结果显示在屏幕上。

(4) 数据的组织与存取

数据的组织与存取提供数据在外部储存设备上的物理组织与存取方法。

(5) 事务运行管理

事务运行管理提供事务运行管理及运行日志，事务运行的安全性监控和数据完整性检查，事务的并发控制及系统恢复等功能。

(6) 数据库的维护

数据库的维护为数据库管理员提供软件支持，包括数据安全控制、完整性保障、数据库备份、数据库重组及性能监控等维护工具。

基于关系模型的数据库管理系统已日臻完善，并已作为商品化软件广泛应用于各行各业。

1.4 数据库系统

考察数据库系统的结构可以有多种不同的层次或不同的角度。

从数据库管理系统的角度看，数据库系统通常采用三级模式结构，这是数据库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/服务器结构和浏览器/应用服务器/数据库服务器等结构，这是数据库系统外部的体系结构。

1.4.1 数据库系统的三级模式

人们为数据库设计了一个严谨的体系结构，数据库领域公认的标准结构是三级模式结构，它包括外模式、概念模式、内模式，有效地组织、管理数据，提高了数据库的逻辑独立性和物理独立性。用户级对应外模式、概念级对应概念模式、物理级对应内模式，使不同级别的用户对数据库形成不同的视图。

在数据模型中有记录型（type）和记录值（value）的概念。记录型指对某一类数据的结构和属性的描述，如学生（学号、姓名、性别、系名、年龄）表示了一个记录型。而记录值是记录型的一个具体赋值，如（99001、王明、男、计算机、19）表示了一个记录值。

模式（schema）是数据库中全体数据的逻辑结构和特征的描述，它仅仅涉及类型的描述，而不涉及具体的值。模式的一个具体值称为模式的一个实例（instance）。同一个模式可以有很多实例。模式是相对稳定的，实例是相对变动的，因为数据库中的数据总在不断地更新。模式反映的是数据的结构及其联系，而实例反映的是数据库某一时刻的状态。

实际的数据库管理系统产品种类很多，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，数据的存储结构也各不相同。但是无论什么模型的数据库系统，它们在体系结构上具有相同的性质，即采用三维模式结构并提供两级映像功能。

数据库系统的三维模式结构是指数据库系统由外模式、概念模式和内模式三级构成，如图 1.6 所示。

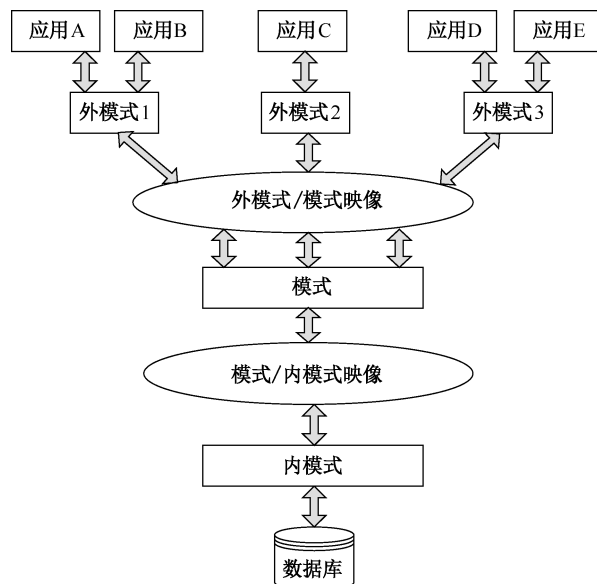


图 1.6 数据库系统的三维模式结构

1. 外模式

外模式也称用户模式，它是数据库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。一个数据库可以有多个外模式。应用程序都是和外模式打交道的。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用户来说是不可见的。

2. 模式

模式也称逻辑模式或概念模式，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。模式实际上是数据库数据在逻辑级上的视图。一个数据库只有一个模式。定义模式时不仅要定义数据的逻辑结构，而且要定义数据之间的联系，定义与数据有关的安全性、完整性要求。

3. 内模式

内模式也称存储模式，一个数据库只有一个内模式。它是数据物理结构和存储方式的描述，是数据在数据库内部的表示方式。例如，记录的存储方式是顺序结构存储还是 B 树结构存储；索引按什么方式组织；数据是否压缩，是否加密；数据的存储记录结构有何规定等。

数据库的三级模式是数据库在 3 个层次上的抽象，使用户能够逻辑地、抽象地处理数据而不必关心数据在计算机中的物理表示和存储。实际上，对于一个数据库系统而言，物理级数据库是客观存在的，它是进行数据库操作的基础；概念级数据库中不过是物理数据库的一种逻辑的、抽象的描述（即模式）；用户级数据库则是用户与数据库的接口，它是概念级数据库的一个子集（外模式）。

用户应用程序根据外模式进行数据操作，通过外模式与模式之间的映射，定义和建立某个外模式与模式间的对应关系，将外模式与模式联系起来，当模式发生改变时，只要改变其映射，就可以使外模式保持不变，对应的应用程序也可保持不变；另一方面，通过模式与内模式之间的映射，定义建立数据的逻辑结构（模式）与存储结构（内模式）间的对应关系，当数据的存储结构发生变化时，只需改变模式与内模式之间的映射，就能保持模式不变，因此应用程序也可以保持不变。

1.4.2 数据库系统的二级映像功能和数据独立性

1. 二级映像

数据库系统的三级模式是数据的 3 个抽象级别，它把数据的具体组织留给数据库管理系统管理，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。同时，它通过二级映像建立了模式间的联系与转换，使得概念模式与外模式也能通过映像而获得其实体。此外，二级映像也保证了数据库系统中数据的独立性，即数据的物理组织改变与逻辑概念级改变相互独立，使得只需调整映像方式而不必改变用户模式。

所谓映像（mapping）就是一种对应规则，说明映像双方如何进行转换。

（1）外模式/模式映像

外模式/模式映像定义了外模式与模式之间的映像关系。由于外模式和模式的数据结构可能不一致，即记录类型、字段类型的命名和组成可能不同，因此需要这个映像说明外部记录和概念之间的对应性。

通过外模式与模式之间的映像把描述局部逻辑结构的外模式与描述全局逻辑结构的模式联系起来。由于一个模式与多个外模式对应，因此，对于每个外模式都有一个外模式/模式映

像用于描述该外模式与模式之间的对应关系。外模式/模式映像通常放在外模式中描述。

有了外模式/模式映像，当模式改变时，如增加新的属性、修改属性的类，外模式/模式映像做相应的改变，使外模式保持不变，则以外模式为依据的应用就不会受到影响，从而保证了数据与程序之间的逻辑独立性，也就是数据的逻辑独立性。

(2) 模式/内模式映像

模式/内模式映像定义了内模式与模式之间的映像关系。例如，说明逻辑记录 and 字段在内部是如何表示的。该映像定义通常包含在模式描述中。

通过模式与内模式之间的映像把描述全局逻辑结构的模式与描述物理联系起来。由于数据库只有一个模式，也只有一个内模式，因此，模式/内模式映像也只有一个，通常就放在内模式中描述。

有了模式/内模式映像，当内模式改变时，如存储设备或存储方式有所改变，只要对模式/内模式映像做相应的改变，使模式保持不变，则应用程序就不会受到影响，从而保证了数据与程序之间的物理独立性，也就是数据的物理独立性。

从上面的介绍可以看出，由于有两层映像，在内模式发生变化，甚至模式发生变化时，都可以使外模式在最大限度上保持不变。由于应用程序是在外模式所描述的数据结构的基础上编写的，外模式的稳定性就保证了应用程序的稳定性。而这正是数据库结构采用三层模式、两层映像为系统提供了高度的数据独立性所得到的结果。

2. 数据独立性

数据独立性是指应用程序和数据库的数据结构之间相互独立，不受影响。数据独立性分为物理独立性和逻辑独立性。

(1) 物理独立性

物理独立性是指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的，即数据在磁盘上怎样存储由数据库管理系统管理，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构，这样当数据的物理存储改变时，应用程序不必改变。

(2) 逻辑独立性

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，即当数据的逻辑结构改变时，用户程序也可以不变。

1.4.3 数据库系统的体系结构

从最终用户角度来看，数据库系统分为单用户结构、主从式结构、分布式结构和客户/服务器结构等。

1. 单用户结构

单用户结构的数据库系统运行于单台计算机上。整个数据库系统（包括应用程序、数据库管理系统、数据）都装在一台计算机上，某时间段仅为一个用户所独占，且不同计算机之间不能共享数据，如图 1.7 所示。

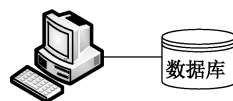


图 1.7 单用户结构的数据库系统

2. 主从式结构

主从式结构的数据库系统是指一个主机带多个终端的

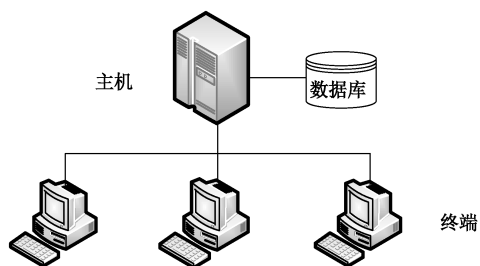


图 1.8 主从式结构的数据库系统

多用户数据库系统,如图 1.8 所示。在这种结构中,数据库系统(包括应用程序、数据库管理系统、数据)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、使用数据库,共享数据资源。主从式数据库系统中的主机是一个通用计算机,既执行数据库管理系统功能又执行应用程序。

主从式结构的数据库系统的特点是数据集中、处理任务集中。其优点是数据易于管理和维护,缺点是主机的任务可能会过于繁重,从而使系统性能大幅度下降;当主机出现故障时,整个系统都不能使用,因此系统的可靠性不高。

3. 分布式结构

分布式结构的数据库系统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,如图 1.9 所示。网络中的每个结点都可以独立处理本地数据库中的数据,执行局部应用;也可以同时存取或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是计算机网络发展的必然产物,它适应了地理上分散的团体或组织对数据库应用、远程共享的需求。

分布式结构数据库系统是以网络为平台的数据库系统,数据分布对用户是透明的。分布式结构的特点是数据分布、处理任务也分布。其优点是充分共享、高效地使用远程资源,各结点又能独立自治;缺点是数据的分布给数据的处理、管理及维护带来了困难。

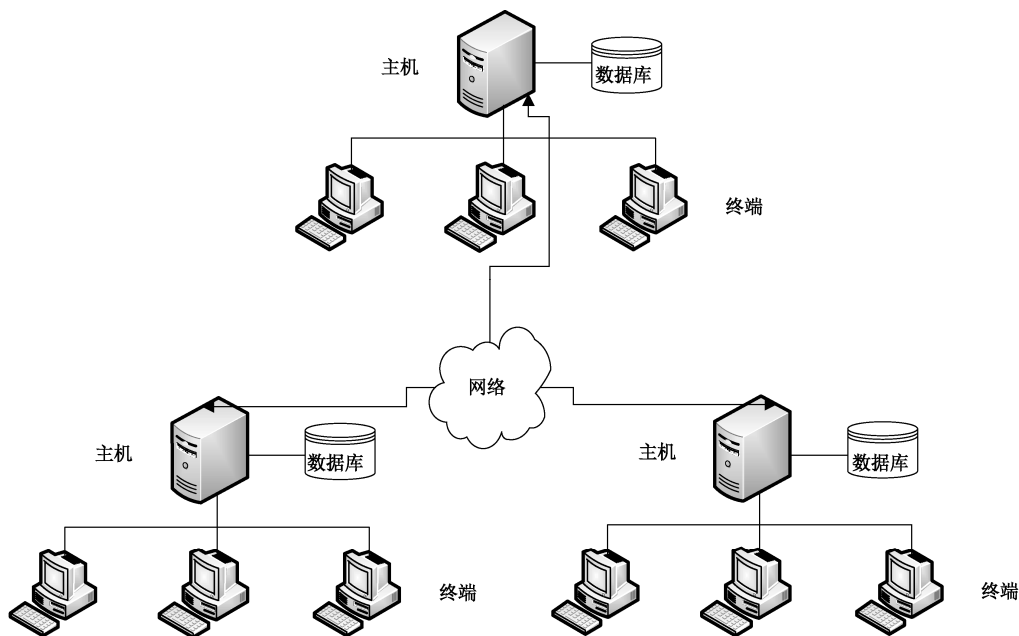


图 1.9 分布式结构的数据库系统

4. 客户/服务器结构

随着工作站功能的增强和广泛使用,人们开始把数据库管理系统功能和应用分开,网络

中某些结点上的计算机专门用于存放数据库和执行数据库管理系统功能，称为数据库服务器，简称服务器。其他结点上的计算机安装数据库管理系统的外围应用开发工具，支持用户的交互与应用，称为客户机，这就是客户/服务器结构的数据库系统，如图 1.10 所示。目前主要有两种模式：客户/服务器模式（client/server, C/S）和浏览器/服务器模式（browser/server, B/S）。

客户/服务器结构数据库系统的优点是显著提高了系统的性能、吞吐量和负载能力。客户/服务器结构数据库往往更加开放，有多种不同的硬件和软件平台及更加灵活的数据库应用开发工具，应用程序具有更强的可移植性，同时也可以减少软件维护开销；缺点是服务器主机的任务重，交互数据频率高。

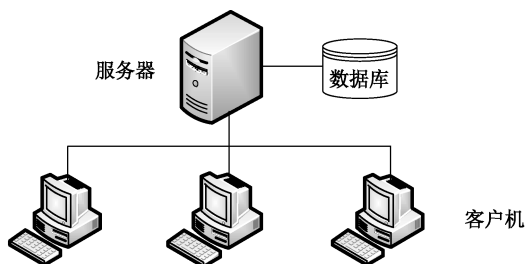


图 1.10 客户/服务器结构的数据库