第5章 数据库技术应用基础

当今世界，信息已成为极为重要的资源之一，在各行各业、各个专业领域的信息化建设过程中，数据库系统是信息化建设中数据处理的核心系统。数据库技术作为计算机技术的重要分支，已成为对大量数据进行组织与管理的信息系统核心技术和网络信息化管理系统的重要基础，几乎所有的信息管理系统都以数据库为核心进行数据的组织与管理。本章主要内容有：

- 数据库系统的组成与功能；
- 数据库技术的应用与发展；
- 信息实体数据模型和实体联系模型的构建；
- 数据实体关系运算与转换；
- 关系数据库的设计理论；
- 关系模式的规范化；
- 结构化查询语言 SQL。

5.1 数据库技术概述

数据库系统在信息化建设中的广泛使用使数据库技术的应用与发展不断深入人类社会生活的各个领域，从各种信息数据的采集转换到企业管理、银行业务管理、情报检索、档案管理、人口普查等，都离不开数据库管理。在物联网、传感网等信息技术应用迅速普及和发展的今天，数据库系统也需要不断更新、发展和完善，数据库设计、开发与维护领域的高级人才十分稀缺。业界专家普遍认为，我国数据库技术应用与国外相比，差距主要在于应用技术的经验和积累。

5.1.1 数据库技术特点

数据库技术的产生使数据管理进入了一个全新的阶段。数据与程序相互独立，可以最大限度地减少数据的重复性，最大限度地被多个用户共享。数据库技术应用主要有以下几个特点。

1. 数据库中的数据是结构化的

在文件系统中，从整体来看，数据是无结构的，不同文件中的记录型之间没有联系，存取数据时只能按顺序访问。数据库系统中的数据管理与组织不仅反映数据项之间的联系，还能表示记录型之间的联系，这种联系可以通过存储路径实现。例如在学生选课管理中，一个学生可以选修多门课，一门课可被多个学生选修，可用 3 种记录型（学生的基本情况、课程的基本情况以及选课的基本情况）进行管理，如图 5.1 所示。

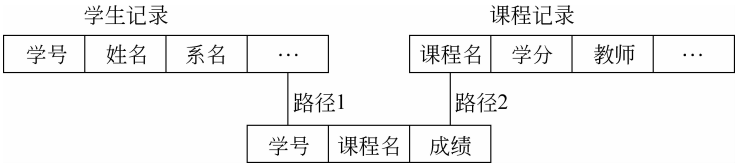

图 5.1　学生选课管理中的数据联系

在查询“张三的学习成绩及学分”时，如果用文件系统实现学生选课的管理，程序员需要编程，并从 3 个文件中查找出所需的信息；如果用数据库系统管理学生选课，可通过关键字关联的数据存取路径实现。利用数据关联存取路径，数据检索可以从一个记录型走到另一个记录型。事实上，学生记录、课程记录与选课记录有着密切的联系，数据存取路径表示了这种联系，这是数据库系统与文件系统的根本区别。

2. 数据库中的数据是面向系统的

数据库中的数据不是面向某个具体应用的，而是面向系统的，这样可减少数据的冗余，实现数据的最大共享，如图 5.2 所示。

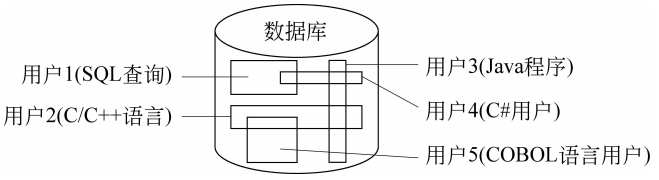

图 5.2　数据库数据的共享

3. 数据库系统比文件系统有更高的数据独立性

数据库系统的数据访问结构分为用户级（应用程序或终端用户）数据逻辑结构、整体数据级的逻辑结构（用户数据逻辑结构的最小并集）和数据存储级的物理结构三级。

数据库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结构的逻辑结构、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。当整体数据的逻辑结构或数据的物理结构发生变化时，应用程序访问可以不变。

例如，根据需要把课程记录中的字段“学分”移出，加到选课记录中，即课程记录中减少一个字段，选课记录中增加一个字段，原来的应用程序可以不变。

4. 数据库系统为用户提供了方便的接口

用户可以用数据库系统提供的查询语言和交互式命令操纵数据库，也可以用高级语言（如 SQL、C/C++、Java、COBOL 等）编写程序以操纵数据库，从而拓宽了数据库的应用范围。

数据库系统中数据的最小存取单位是数据项，若干数据项组成记录，文件系统的最小存取单位是记录。应用程序与数据库的联系如图 5.3 所示。

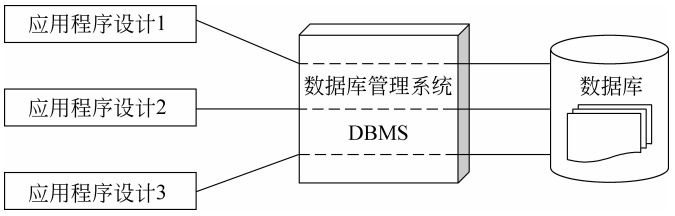

图 5.3　应用程序与数据库的联系

数据库管理系统（Database Management System，DBMS）是一个软件系统，它能够操纵数据库中的数据，对数据库进行统一的控制与管理。

5.1.2　数据库系统的组成

数据库系统（Database System，DBS）是一个实际可运行的系统，它能按照数据库的方式存储和维护数据，并能够向应用程序提供数据。数据库系统通常由数据库、硬件系统、软件系统和数据库管理员（Database Administrator，DBA）4 个部分组成。

1. 数据库

数据库的定义在前面的章节中已经讲述过。数据库的体系结构可划分为两个部分，一部分是存储应用所需的数据，称为物理数据库部分；另一部分是描述部分，可描述数据库的各级结构，这部分由数据字典管理。例如 Oracle 数据库系统可查询其数据字典，了解 Oracle 各级结构的描述。

2. 硬件系统

数据库的运行需要硬件支持系统，中央处理机、主存储器、外存储器等设备是不可缺少的。数据库系统需要足够大的内存以存放支持数据库运行的操作系统与数据库管理系统的核心模块、数据库的数据缓冲区和应用程序以及用户的工作区。例如 Oracle 5.1 版本在微机上运行需要 1.5MB 的内存；Sybase 的微机版本最低需要 12MB 的内存，最好为 16MB 的内存。由于数据库中存储了大量的数据，故需要足够大的磁盘等直接存取设备存取数据或进行数据库的备份，此外还要求硬件系统有较高的信道能力，以提高数据的传输速度。

3. 软件系统

数据库系统的软件主要包括支持 DBMS 运行的操作系统、DBMS 本身及开发工具。为了开发应用系统，还需各种高级语言及其编译系统，例如 Oracle 数据库系统与高级语言 C、Fortran、COBOL 等之间都有接口。Access 关系数据库管理系统内置了 Visual Basic for Application，并允许 Visual Basic 直接访问。不同用户开发的应用可能不同，需用不同的高级语言访问数据库，相应地要把这些高级语言的编译系统装入系统，以供用户使用。

开发工具是为应用开发人员和最终用户提供的高效率的开发应用软件。例如 Oracle 数据库系统提供第四代开发工具。SQL * FORMS 提供一种基于表格的应用开发工具，应用设计人员用它设计格式化画面，应用操作人员通过格式化画面向 Oracle 数据库录入数据，从 Oracle 数据库检索数据。SQL * FORMS 还提供了数据完备性和安全性的检查功能。SQL * GRAPH 是一个交互式的图形生成软件包，它利用从 Oracle 数据库中提取出来的数据生成彩色的拼图、直方图和折线图。大多数数据库系统都提供了开发工具软件，为数据库系统的开发和应用建立了良好的环境。这些开发工具软件都以 DBMS 为核心。

4. 数据库管理员

数据库管理员(DBA)、系统分析员、应用程序员和用户是管理、开发和使用数据库的主要人员。这些人员的职责和作用是不同的，因此涉及不同的数据抽象级别，有不同的数据视图。

数据库管理员可以是一个人，也可以是由几个人组成的小组，他们全面负责管理、维护和控制数据库系统，一般来说，应由业务水平较高和资历较深的人员担任。DBA 的具体职责如下。

(1) 决定数据库的信息内容。数据库中存放什么信息是由 DBA 决定的，他们确定应用的实体，包括属性及实体间的联系，完成数据库模式的设计，并同应用程序员一起完成用户界面的设计工作。

(2) 决定数据库的存储结构和存取策略。DBA 负责确定数据的物理组织、存放方式及数据存取方法。

(3) 定义存取权限和有效性检验。用户对数据库的存取权限、数据的保密级别和数据的约束条件都是由 DBA 确定的。

(4) 建立数据库。DBA 负责原始数据的装入及建立用户数据库。

(5) 监督数据库的运行。DBA 负责监视数据库的正常运行。当出现软硬件故障时，应能及时排除，使数据库恢复到正常状态，并负责数据库的定期转储和日志文件的维护等工作。

(6) 重组和改进数据库。DBA 通过各种日志和统计数字分析系统性能。当系统性能下降(如存取效率和空间利用率降低)时，应对数据库进行重新组织，同时根据用户的使用情况不断改进数据库的设计，以提高系统性能，满足用户的需要。

5. 用户

用户分为应用程序和最终用户(End User)两类,他们通过数据库系统提供的接口和开发工具软件使用数据库。目前常用的接口方式有菜单驱动、表格操作、生成报表以及利用数据库与高级语言的接口编程等。

6. 数据库系统结构

数据库系统结构是一个多极结构,一方面能方便地存储数据,同时又能高效安全地组织数据。现有的数据库系统都采用三级模式和二级映射结构,其体系结构如图 5.4 所示。

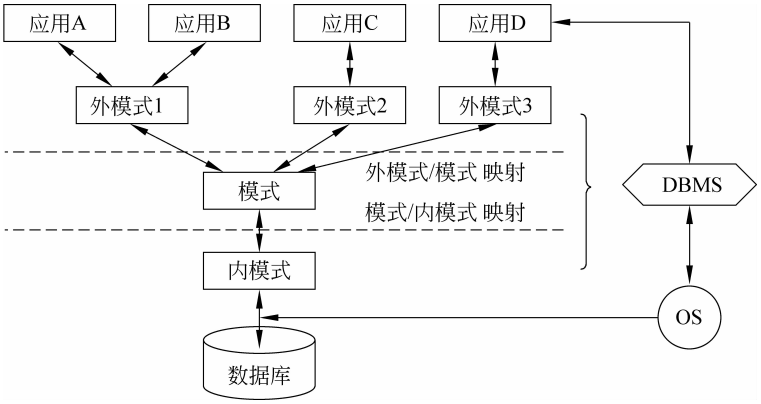

图 5.4　数据库系统体系结构

1) 模式

模式又称概念模式,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。

2) 外模式

外模式又称子模式或用户模式,是数据库用户所看到和使用的局部数据的逻辑结构和特征的描述,也就是用户看到和使用的数据。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,数据库中其余的数据对他们来说是不可见的。

3) 内模式

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,是数据在数据库内部的表示方式。一个数据库只有一个内模式。

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数据库管理系统,使用户能逻辑、抽象地处理数据,而不必关心数据在计算机中具体的表示方式和存储方式。

为了实现这三个抽象层次的联系和转换,数据库系统在三级模式中提供了两种映射:

- 外模式和模式之间的映射;
- 模式和内模式之间的映射。

正是由于这二级映射功能，才使得数据库系统中的数据具有较高的逻辑独立性和物理独立性。

5.1.3 数据库系统功能

数据库是存储在外存储器上的逻辑相关的数据的集合，并按一定的方式进行组织和管理。数据库数据相互关联又彼此独立，并以一定的组织方式存储，具有较少的数据冗余，能被多个应用程序或用户共享。

1. 数据的完整性

数据的完整性可保证数据库存储数据的正确性。例如预订同一班飞机的旅客不能超过飞机的定员数；订购货物时，订货日期不能大于发货日期。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，可以保证数据库中数据的正确性。

2. 数据的安全性

并非每个应用都应该存取数据库中的全部数据。例如建立一个人事档案的数据库，只有那些需要了解工资情况并且有一定权限的工作人员才能存取这些数据。数据的安全性可保护数据库不被非法使用，防止数据丢失和被盗。

3. 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，使数据库中的数据完整性受到破坏，从而导致数据的不一致性。数据库的并发控制可防止这种现象的发生，提高了数据库的利用率。

4. 数据库的恢复

任何系统都不可能永远正确无误地工作，数据库系统也是如此。数据库系统在运行过程中会出现硬件或软件的故障。数据库系统具有恢复能力，能把数据库恢复到最近某个时刻的正确状态。

5.1.4 数据库技术应用发展

随着计算机技术的不断发展，出现了分布式数据库、面向对象数据库和智能型知识数据库等，它们通常被称为高级数据库技术。

1. 分布式数据库

由于计算机网络通信的迅速发展，使得分散在不同地理位置的计算机能够实现数据的通信和资源的共享，已建立并使用中的许多数据库也需要互联，因此产生了分布式数据库系统。

分布式数据库是分布在计算机网络不同节点上的数据的集合，它有以下两个主要特点。

(1) 网络每个节点上的数据库都具有独立处理的能力。大多数数据处理都是就地完成的，不能处理的才交给其他处理机处理。

(2) 计算机之间用通信网络连接。每个节点上的应用可访问本节点上数据库中的数据，这种应用称为局部应用。也可以通过网络访问其他节点上的数据库数据，这种应用称为全局应用。

分布式数据库在物理上是分布的，在逻辑上是统一的。在分布式数据库系统中适当地增加了数据冗余，个别节点的失效不会引起系统的瘫痪，而且多台处理机可并行工作，提高了数据处理的效率。

2. 面向对象数据库

随着计算机的发展，数据库的应用领域不断扩大，从商务领域(如存款取款、财务管理、人事管理等)拓宽到计算机集成制造系统(CIMS)、计算机辅助设计(CAD)、计算机辅助生产管理等应用领域。这些新的应用领域对数据库技术提出了新的要求。

在面向对象的数据库系统(Object-Oriented Database Systems，OODBS)中，一切概念上的存在，小至单个整数或数字串，大至由许多部件构成的系统均称为对象。一个对象有数据部分和程序部分，例如职工张三是一个对象，25 岁，每月工资为 1500 元。这个对象的数据部分是：姓名是张三，年龄为 25，工资是 1500 元；修改对象张三的年龄或工资，或检索对象属性(例如姓名、年龄、工资)的值时，所使用的程序构成了对象的程序部分。面向对象的数据库系统比一般数据库系统具有更多的特点和应用领域。面向对象的方法将成为标准方法，未来的软件系统将建立在面向对象的概念上。

3. 知识库系统的特点

人工智能的发展要求计算机不仅能够管理数据，还能管理知识，允许用户说明处理数据的规则，这种功能可用知识库系统实现。

知识库是把有关知识的数据信息从应用程序中分离出来加到数据库中，它把人工智能的知识获取技术和机器学习的理论引入数据库系统，通过抽取隐含在数据库实体间的逻辑蕴含关系和隐含在应用中的数据操纵之间的因果联系等，形式化地描述数据库中实体联系的语义网络，并把语义知识自动提供给推理机，从已有的事实知识推理出新的事实知识。知识库是一门新的学科，研究知识表示、结构、存储、获取等技术。知识库是专家系统、知识处理系统的重要组成部分。

5.2 数 据 模 型

数据之间的联系称为数据模型，通常有 3 种，即层次模型、网状模型和关系模型。与之对应的数据库称为层次数据库、网状数据库和关系数据库。

5.2.1 数据模型

数据模型是对现实世界进行抽象的工具。现实世界是复杂多变的，目前任何一种科学技术手段都不可能将现实世界按原样进行复制和管理，只能抽取某个局部的特征，构造反映这个局部的模型，帮助人们理解和表达数据处理的静态特征及动态特征。

1. 数据模型的类型

数据模型是数据库技术的核心，数据库管理系统都是基于某种数据模型的。目前使用的数据模型基本上可分为两种类型，一种类型是概念模型，也称信息模型。这种模型不涉及信息在计算机中的表示和实现，是按用户的观点进行数据信息建模，强调语义表达能力。这种模型比较清晰、直观，容易被理解。另一种类型是数据模型，这种模型是面向数据库中数据逻辑结构的，如关系模型、层次模型、网状模型和面向对象的数据模型等，用户可以使用这种数据模型定义和操纵数据模型中的数据。

2. 数据模型的构成

数据模型包括 3 个部分：数据结构、数据操纵和数据的完整性约束。

数据结构是存储在数据库中的对象的类型的集合。例如建立一个科技开发公司的人事管理的数据库，每个人的基本情况（姓名、单位、出生年月、工资、工作年限等）说明对象人的特征，构成在数据库中存储的框架，即对象的类型。公司中每个技术人员可以参加多个项目，每个项目可有多人参加，这类对象之间存在数据关联，这种数据关联也要存储在数据库中。数据库系统是按数据结构的类型组织数据的。由于采用的数据结构类型不同，因此通常把数据库分为层次数据库、网状数据库、关系数据库和面向对象数据库等。

数据操纵是指对数据库中各种对象实例的操作。例如根据用户的要求检索、插入、删除、修改对象实例的值。

数据的完整性约束是指在给定的数据模型中，数据及数据关联所遵守的一组通用的完整性规则，它能保证数据库中数据的正确性、一致性。例如数据库主键的值不能取空值（没有定义的值）；关系数据库中，每个非空外键值（foreign key）必须与某一主键值相匹配。这类完整性约束是数据模型所必须遵守的通用的完整性规则。另一类完整性约束是用户根据数据模型提供的完整性约束机制自己定义的。例如在销售管理中，发货日期要在订货日期之后。

数据模型是数据库技术的关键，可用数据模型描述数据库的结构和语义。

5.2.2 构建信息实体数据模型

数据模型是对现实世界的抽象描述。在组织数据模型时，人们首先将现实世界中存在的客观事物用某种信息结构表示出来，然后再转换为计算机能表示的数据形式。

1. 信息实体的数据转换

信息是指客观世界中存在的事物在人们头脑中的反映，人们把这种反映用文字、图形等形式记录下来，经过命名、整理、分类就形成了信息，如图 5.5 所示。

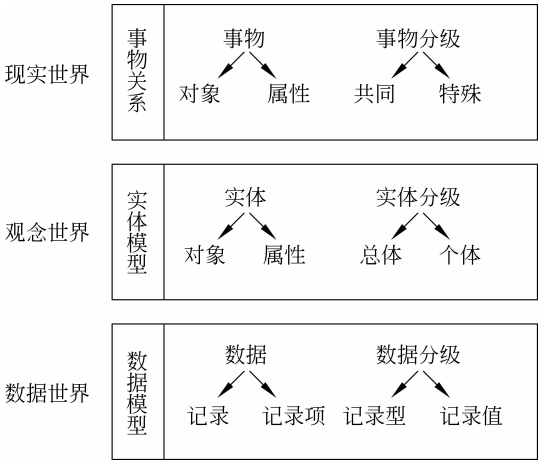

图 5.5　信息数据的形成

在信息领域中，数据库技术用到的术语有实体、属性、实体集、键等。

- 实体(Entity)：实体是客观存在并可相互区分的事物。例如人、部门、雇员等都是实体。实体可以指实际的对象，也可以指抽象的对象。
- 属性(Attribute)：属性是实体所具有的特性，每一特性都称为实体的属性。例如学生的学号、班级、姓名、性别、出生年月等都为学生的属性。属性是描述实体的特征，每一属性都有一个值域。值域的类型可以是整数型、实数型或字符串型等，如学生的年龄是整数型，姓名是字符串型。
- 实体集：具有相同属性(或特性)的实体的集合称为实体集。例如全体教师是一个实体集，全体学生也是一个实体集。
- 键(Key)：也称关键字，能唯一标识一个实体的属性及属性值。例如，学号是学生实体的键，而姓名不能作为键，因为有重名。

2. 信息实体的数据联系

数据模型反映了现实世界中事物间的各种联系，即实体间的联系。联系通常有两种：一种是实体内部的联系，即实体中属性间的联系；一种是实体与实体之间的联系。在数据模型中，不仅要考虑实体属性间的联系，更要考虑实体与实体间的联系，下面主要讨论后一种联系。

实体间的联系是错综复杂的，但就两个实体的联系来说，有以下 3 种情况。

1) 一对一的联系

如果 A 中的任意一个个体至多对应于 B 中的一个个体，反之 B 中的任意一个个体至多对应于 A 中的一个个体，则称 A 与 B 是一对一的联系，如图 5.6 所示。

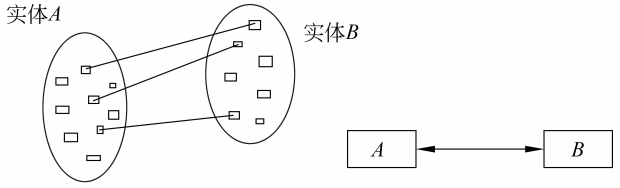

图 5.6　一对一的联系

这是最简单的一种实体间的联系，它表示了两个实体集中的个体间存在一对一的联系。例如，一所学校有一位正校长、每个班级有一个班长等。这种一对一的联系记为 1∶1。

2）一对多的联系

如果 A 中至少有一个个体对应于 B 中一个以上的个体，反之 B 中任一个体对应于 A 中至多一个个体，则称 A 与 B 是一对多的联系，如图 5.7 所示。

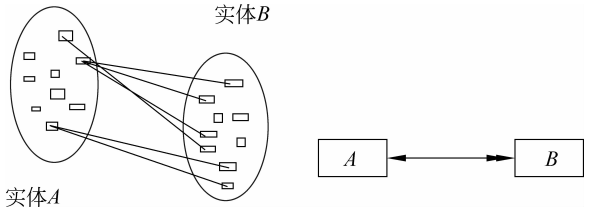

图 5.7　一对多的联系

实体间存在的另一种联系是一对多的联系。例如一个教研室有许多教师、一个班级有许多学生等。这种一对多的联系记为 1∶M。

3）多对多的联系

如果 A 中至少有一个个体对应于 B 中一个以上的个体，反之 B 中至少有一个个体对应于 A 中一个以上的个体，则称 A 与 B 是多对多的联系，如图 5.8 所示。

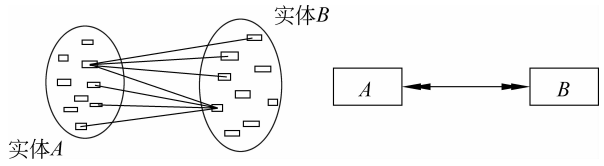

图 5.8　多对多的联系

例如，一个教师教许多学生、一个学生有许多教师教课等。多对多的联系表示了实体集之间存在着交叉联系的相互关系，其中一个实体集中的任一实体与另一实体集中的实体间存在一对多的联系，反之亦然。这种联系记为 N∶M。

5.2.3　构建实体联系模型

实体联系模型（Entity-Relationship Model，ER 模型）是一个面向问题的概念模型，即