

第 1 章 数据库概述

数据库技术是数据管理的有效技术，其作为信息系统的核心和基础，有着越来越广泛的应用。数据库管理系统是一个系统软件，用于科学地组织和存储数据、高效地获取和维护数据。数据库系统是在计算机系统中引入数据库之后组成的系统，它用来组织和存取大量数据。在本章中，介绍数据库、数据库管理系统、数据库系统、数据库系统结构、数据模型等内容，它是学习以后各章的基础。

1.1 数据库系统概述

下面介绍数据库的一些基本概念，并强调数据库管理系统是数据库系统的核心组成部分。

1.1.1 数据库

1. 数据

数据（Data）是事物的符号表示，数据可以是数字、文字、图像、声音等。一个学生记录数据如下所示：

161001	周浩然	男	1995-09-14	电子信息工程	52
--------	-----	---	------------	--------	----

2. 数据库

数据库（DataBase）是以特定的组织结构存放在计算机的存储介质中的相互关联的数据集合。

数据库具有以下特征：

- （1）是相互关联的数据集合，不是杂乱无章的数据集合。
- （2）数据存储于计算机的存储介质中。
- （3）数据结构比较复杂，有专门理论支持。

数据库包含了以下含义：

- （4）提高了数据和程序的独立性，有专门的语言支持。
- （5）建立数据库的目的是为应用服务。

1.1.2 数据库管理系统

数据库管理系统（DataBase Management System, DBMS）是位于用户和操作系统之

间的一个数据管理软件，它是在操作系统支持下的系统软件，是数据库应用系统的核心组成部分，它的主要功能如下：

- (1) 数据定义功能：提供数据定义语言，定义数据库和数据库对象。
- (2) 数据操纵功能：提供数据操纵语言，对数据库中的数据进行查询、插入、修改、删除等操作。
- (3) 数据控制功能：提供数据控制语言，进行数据控制，即提供数据的安全性、完整性、并发控制等功能。
- (4) 数据库建立维护功能：包括数据库初始数据的装入、转储、恢复和系统性能监视、分析等功能。

1.1.3 数据库系统

数据库系统 (DataBase System, DBS) 是由数据库、操作系统、数据库管理系统、应用程序、用户、数据库管理员组成的用于存储、管理、处理和维持数据的系统，数据库系统是数据库应用系统的简称，如图 1.1 所示。

数据库系统分为客户-服务器模式 (C/S) 和三层客户-服务器 (B/S) 模式。

1. C/S 模式

应用程序直接与用户打交道，数据库管理系统不直接与用户打交道，因此，应用程序称为前台，数据库管理系统称为后台。因为应用程序向数据库管理系统提出服务请求，所以称为客户程序 (Client)，而数据库管理系统向应用程序提供服务，所以称为服务器程序 (Server)，上述操作数据库的模式称为客户-服务器模式 (C/S)，如图 1.2 所示。

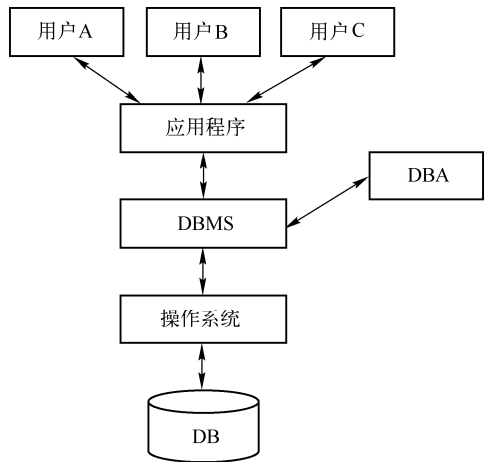


图 1.1 数据库系统的组成

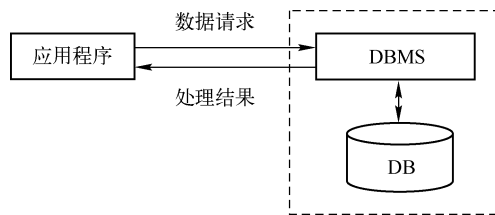


图 1.2 C/S 模式

2. B/S 模式

基于 Web 的数据库应用采用三层客户-服务器模式 (B/S)，第一层为浏览器，第二层为 Web 服务器，第三层为数据库服务器，如图 1.3 所示。

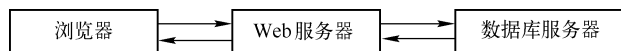


图 1.3 B/S 模式

1.2 数据管理技术的发展

数据管理是指对数据进行分类、组织、编码、存储、检索和维护等工作，数据管理技术的发展经历了人工管理阶段、文件系统阶段、数据库系统阶段，现在正在向更高级的数据库系统发展。

1. 人工管理阶段

20 世纪 50 年代中期以前，人工管理阶段的数据是面向应用程序的，一个数据集只能对应一个程序，应用程序与数据之间的关系如图 1.4 所示。

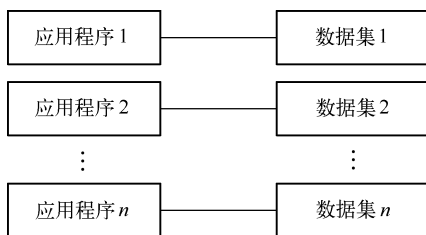


图 1.4 人工管理阶段应用程序与数据之间的关系

人工管理阶段的特点如下：

- (1) 数据不保存。只是在计算某一课题时将数据输入，用完即撤走。
- (2) 数据不共享。数据面向应用程序，一个数据集只能对应一个程序，即使多个不同程序用到相同的数据，也得各自定义。
- (3) 数据和程序不具有独立性。数据的逻辑结构和物理结构发生改变，必须修改相应的应用程序，即要修改数据必须修改程序。
- (4) 没有软件系统对数据进行统一管理。

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期，计算机不仅用于科学计算，也开始用于数据管理。数据处理的方式不仅有批处理，还有联机实时处理。应用程序和数据之间的关系如图 1.5 所示。

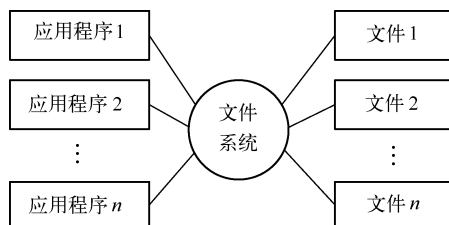


图 1.5 文件系统阶段应用程序与数据之间的关系

文件系统阶段数据管理的特点如下：

- (1) 数据可长期保存。数据以文件的形式长期保存。
- (2) 数据共享性差，冗余度大。在文件系统中，一个文件基本对应一个应用程序，

当不同应用程序具有相同的数据时，也必须各自建立文件，而不能共享数据，数据冗余度大。

(3) 数据独立性差。当数据的逻辑结构改变时，必须修改相应的应用程序，数据依赖于应用程序，独立性差。

(4) 由文件系统对数据进行管理。由专门的软件——文件系统进行数据管理。文件系统把数据组织成相互独立的数据文件，可按文件名访问，按记录存取。应用程序与数据之间有一定的独立性。

3. 数据库系统阶段

20 世纪 60 年代后期开始，数据管理对象的规模越来越大，应用越来越广泛，数据量快速增加。为了实现数据的统一管理，解决多用户、多应用共享数据的需求，数据库技术应运而生，出现了统一管理数据的专门软件——数据库管理系统。

数据库系统阶段应用程序和数据之间的关系如图 1.6 所示。

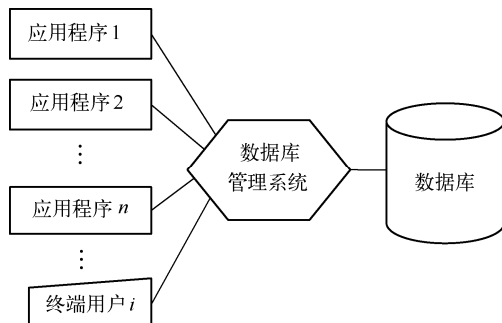


图 1.6 数据库系统阶段应用程序与数据之间的关系

数据库系统与文件系统相比较，具有以下主要特点：

- (1) 数据结构化。
- (2) 数据的共享度高，冗余度小。
- (3) 有较高的数据独立性。
- (4) 由数据库管理系统对数据进行管理。

在数据库系统中，数据库管理系统作为用户与数据库的接口，提供了数据库定义、数据库运行、数据库维护和数据安全性、完整性等控制功能。

1.3 数据库系统结构

从数据库管理系统内部系统结构看，数据库系统通常采用三级模式结构。

数据模式是数据库中全体数据的结构和特征的描述，它仅仅涉及型的描述，不涉及具体的值。模式的一个具体值称为模式的一个实例，同一个模式可以有很多实例。模式是相对稳定的，而实例是相对变动的。模式反映的是数据的结构及其关系，而实例反映的是数据库某一时刻的状态。

模型与模式的区别在于：模型用图形来表示，直观清晰，但计算机无法识别，需要使用一种语言（如由 DBMS 提供的 DDL）来进行描述。模式则是对模式的描述。

1.3.1 数据库系统的三级模式结构

模式（Schema）是指对数据的逻辑结构或物理结构、数据特征、数据约束的定义和描述，它是对数据的一种抽象，模式反映数据的本质、核心或类型等方面。

数据库系统的标准结构是三级模式结构，它包括外模式、模式和内模式，如图 1.7 所示。

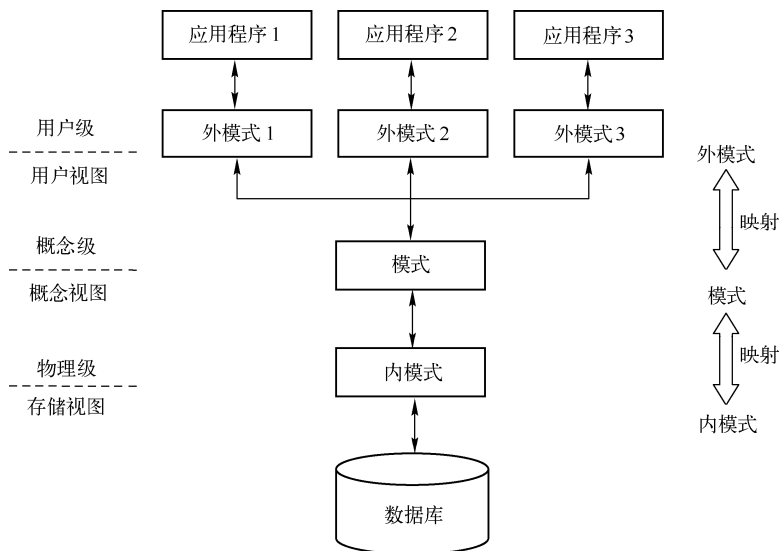


图 1.7 数据库系统的三级模式结构

1. 外模式

外模式（External Schema）又称子模式或用户模式，位于三级模式的最外层，对应于用户级。它是某个或某几个用户所看到的数据视图，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，同一外模式也可以为某一用户的多个应用系统所用，但一个应用程序只能使用一个外模式，它是由外模式描述语言（外模式 DDL）来描述和定义的。

2. 模式

模式（Schema）又称概念模式，也称逻辑模式，位于三级模式的中间层，对应于概念级。它是由数据库设计者综合所有用户的数据，按照统一观点构造的全局逻辑结构，是所有用户的公共数据视图（全局视图）。一个数据库只有一个模式，它是由模式描述语言（模式 DDL）来描述和定义的。

3. 内模式

内模式（Internal Schema）又称存储模式，位于三级模式的底层，对应于物理级。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。一个数据库只有一个内模式，它是由内模式描述语言（内模式 DDL）来描述和定义的。

1.3.2 数据库的二级映射功能和数据独立性

为了能够在内部实现这三个抽象层次的联系和转换，数据库管理系统在这三级模式

之间提供了两级映射：外模式/模式映射，模式/内模式映射。

1. 外模式/模式映射

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。数据库系统都有一个外模式/模式映射，它定义了该外模式与模式之间的对应关系。

当模式改变时，由数据库管理员对各个外模式/模式映射做相应改变，可以使外模式保持不变。

应用程序是依据数据的外模式编写的，保证了数据与程序的逻辑独立性，简称数据逻辑独立性。

2. 模式/内模式映射

数据库中只有一个模式，也只有一个内模式，所以模式/内模式映射是唯一的，它定义了数据库全局逻辑结构与存储结构之间的对应关系。当数据库的存储结构改变时，由数据库管理员对模式/内模式映射做相应改变，可以使模式保持不变，从而应用程序也不必改变。保证了数据与程序的物理独立性，简称数据物理独立性。

在数据库的三级模式结构中，数据库模式即全局逻辑结构是数据库的中心与关键，它独立于数据库的其他层次。

数据库的内模式依赖于它的全局逻辑结构，但独立于数据库的用户视图即外模式，也独立于具体的存储设备。

数据库的外模式面向具体的应用程序，它定义在逻辑模式之上，但独立于内模式和存储设备。

数据库的二级映射保证了数据库外模式的稳定性，从而根本上保证了应用程序的稳定性，使得数据库系统具有较高的数据与程序的独立性。数据库的三级模式与二级映射使得数据的定义和描述可以从应用程序中分离出去。

1.3.3 数据库管理系统的工作过程

数据库管理系统控制的数据操作过程基于数据库系统的三级模式结构与二级映射功能，下面通过读取一个用户记录的过程反映数据库管理系统的工作过程，如图 1.8 所示。

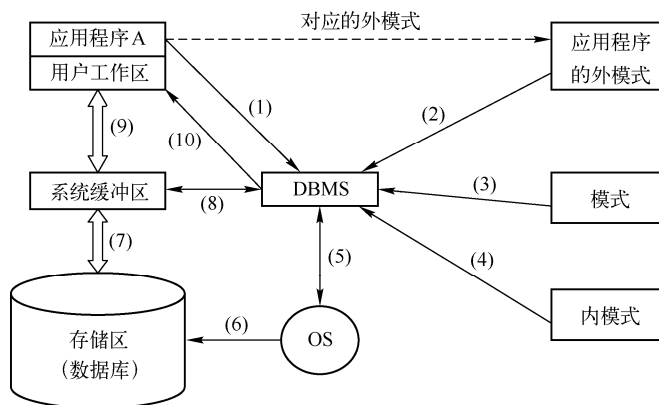


图 1.8 应用程序从数据库中读取一个用户记录的过程

- (1) 应用程序 A 向 DBMS 发出从数据库中读用户记录的命令。
- (2) DBMS 对该命令进行语法检查、语义检查，并调用应用程序 A 对应的外模式，检查 A 的存取权限，决定是否执行该命令。如果拒绝执行，则转 (10) 向用户返回错误信息。
- (3) 在决定执行该命令后，DBMS 调用模式，依据外模式/模式映射的定义，确定应读入模式中的哪些记录。
- (4) DBMS 调用内模式，依据模式/内模式映射的定义，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 (OS) 发出执行读取所需物理记录的命令。
- (6) 操作系统执行从物理文件中读数据的有关操作。
- (7) 操作系统将数据从数据库的存储区送至系统缓冲区。
- (8) DBMS 依据内模式/模式、模式/外模式映射的定义 (仅为模式/内模式、外模式/模式映射的反方向，并不是另一种新映射)，导出应用程序 A 所要读取的用户记录格式。
- (9) DBMS 将用户记录从系统缓冲区传送到应用程序 A 的用户工作区。
- (10) DBMS 向应用程序 A 返回命令执行情况的状态信息。

以上为 DBMS 读一个用户记录的过程，DBMS 向数据库写一个用户记录的过程与此类似，只是过程基本相反而已。由 DBMS 控制的用户记录的存取操作，就是由很多读或写的基本过程组合完成的。

1.4 数据模型

模型是对现实世界中某个对象特征的模拟和抽象，数据模型 (Data Model) 是对现实世界数据特征的抽象，它用来描述数据、组织数据和对数据进行操作。数据模型是数据库系统的核心和基础。

1.4.1 两类数据模型

数据模型需要满足三方面的要求：能比较真实地模拟现实世界，容易为人所理解，便于在计算机上实现。

在开发设计数据库应用系统时需要使用不同的数据模型，它们是概念模型、逻辑模型、物理模型。根据模型应用的不同目的，按不同的层次可将它们分为两类，第一类是概念模型，第二类是逻辑模型、物理模型。

第一类中的概念模型，按用户的观点对数据和信息建模，是对现实世界的第一层抽象，又称信息模型。它通过各种概念来描述现实世界的事物以及事物之间的联系，主要用于数据库设计。

第二类中的逻辑模型，按计算机的观点对数据建模，是概念模型的数据化，是事物及事物之间联系的数据描述，提供了表示和组织数据的方法。主要的逻辑模型有层次模型、网状模型、关系模型、面向对象数据模型、对象关系数据模型和半结构化数据模型等。

第二类中的物理模型，是对数据底层的抽象，它描述数据在系统内部的表示方式和存取方法，如数据在磁盘上的存储方式和存取方法。它是面向计算机系统的，由数据库

管理系统具体实现。

为了把现实世界具体的事物抽象、组织为某一个数据库管理系统支持的数据模型，需要经历一个逐级抽象的过程，将现实世界抽象为信息世界，然后将信息世界转换为机器世界。即首先将现实世界的客观对象抽象为某一种信息结构，这种信息结构不依赖于具体计算机系统，不是某一个数据库管理系统支持的数据模型，而是概念级的模型，然后，将概念模型转换为计算机上某一个数据库管理系统支持的数据模型，如图 1.9 所示。

从概念模型到逻辑模型的转换由数据库设计人员完成，从逻辑模型到物理模型的转换主要由数据库管理系统完成。

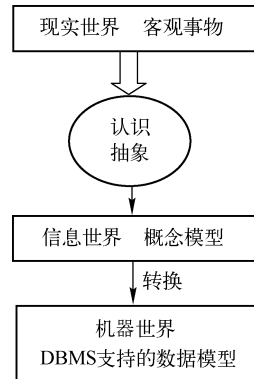


图 1.9 现实世界客观事物的抽象过程

1.4.2 概念模型

概念模型（Conceptual Model）是现实世界到机器世界的一个中间层次。

概念模型又称信息模型，它按用户的观点对数据和信息进行建模，是描述现实世界的概念化结构，它独立于数据库逻辑结构和具体的 DBMS。

1. 概念模型的基本概念

概念模型有以下基本概念：

（1）实体（Entity）：客观存在并可相互区别的事物称为实体。实体用矩形框表示，框内为实体名。实体可以是具体的人、事、物或抽象的概念，例如，在学生成绩管理系统中，“学生”就是一个实体。

（2）属性（Attribute）：实体所具有的某一特性称为属性。属性采用椭圆框表示，框内为属性名，并用无向边与其相应实体连接。例如，在学生成绩管理系统中，学生的特性有学号、姓名、性别、出生日期、专业、总学分，它们就是学生实体的 6 个属性。

（3）码（Key）：能唯一标识实体的最小属性集，又称为键或关键字。例如，学号是学生实体的码。

（4）实体型（Entity Type）：用实体名及其属性名集合来抽象和刻画同类实体，称为实体型。例如，学生（学号，姓名，性别，出生日期，专业，总学分）就是一个实体型。

（5）实体集（Entity Set）：同型实体的集合称为实体集，例如，全体学生记录就是一个实体集。

（6）联系（Relationship）：实体之间的联系，可分为两个实体集之间的联系、多个实体集之间的联系、单个实体集内的联系。

两个实体集之间的联系包括一对一的联系、一对多的联系、多对多的联系。

● 一对一的联系（1：1）

如果实体 A 中的每个实例在实体 B 中至多有一个（也可以没有）实例与之关联，反

之亦然，则称实体 A 与实体 B 具有一对一联系，记作：1 : 1。

例如，一个班只有一个正班长，而一个正班长只属于一个班，班级与正班长两个实体间具有一对一的联系。

● 一对多的联系 (1 : n)

如果实体 A 与实体 B 之间存在联系，并且对于实体 A 中的一个实例，实体 B 中有多个实例与之对应；而对实体 B 中的任意一个实例，在实体 A 中都只有一个实例与之对应，则称实体 A 到实体 B 的联系是一对多的，记为 1 : n。

例如，一个班可有若干学生，一个学生只能属于一个班，班级与学生两个实体间具有一对多的联系。

● 多对多的联系 (m : n)

如果实体 A 与实体 B 之间存在联系，并且对于实体 A 中的一个实例，实体 B 中有多个实例与之对应；而对实体 B 中的一个实例，在实体 A 中也有多个实例与之对应，则称实体 A 到实体 B 的联系是多对多的，记为 m : n。

例如，一个学生可选多门课程，一门课程可被多个学生选修，学生与课程两个实体间具有多对多的联系。

2. 概念模型表示方法

概念模型较常用的表示方法是实体-联系模型 (Entity-Relationship Model, E-R 模型)。

在 E-R 模型中：

- (1) 实体采用矩形框表示，把实体名写在矩形框内。
- (2) 属性采用椭圆框表示，把属性名写在椭圆框内，并用无向边将其与相应的实体框相连。
- (3) 联系采用菱形框表示，联系名写在菱形框中，用无向边将参加联系的实体矩形框分别与菱形框相连，并在连线上标明联系的类型，例如 1 : 1、1 : n 或 m : n。如果联系也具有属性，则将属性框与菱形框也用无向边连上。

【例 1.1】 画出学生成绩管理系统中的学生实体图和课程实体图。

学生实体有学号、姓名、性别、出生日期、专业、总学分等 6 个属性，课程实体有课程号、课程名、学分、教师号等 4 个属性，它们的实体图如图 1.10 所示。

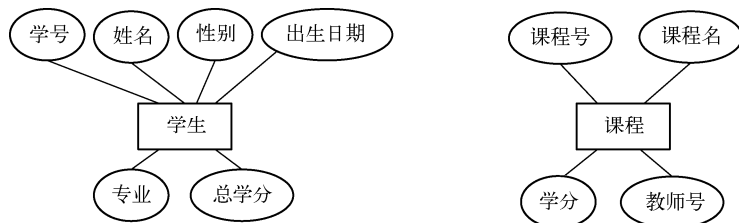


图 1.10 学生成绩管理系统中的学生实体图和课程实体图

【例 1.2】 画出学生成绩管理系统的 E-R 模型。

学生成绩管理系统有学生、课程两个实体，它们之间的联系是选课，学生选修一门课程后都有一个成绩，一个学生可选多门课程，一门课程可被多个学生选修，学生成绩管理系统的 E-R 模型如图 1.11 所示。

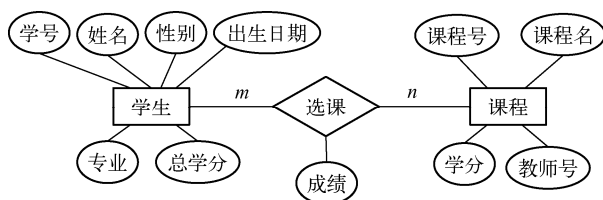


图 1.11 学生成绩管理系统的 E-R 模型

【例 1.3】 画出商店销售管理系统中的员工、订单、商品、部门实体图。

员工、订单、商品、部门实体如下：

员工：员工号、姓名、性别、出生日期、地址、工资。

订单：订单号、客户号、销售日期、总金额。

商品：商品号、商品名称、商品类型代码、单价、库存量、未到货商品数量。

部门：部门号、部门名称。

它们的实体图如图 1.12 和图 1.13 所示。



图 1.12 商店销售管理系统中的员工实体图和订单实体图

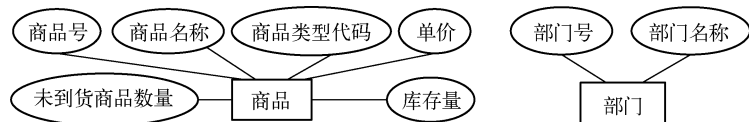


图 1.13 商店销售管理系统中的商品实体图和部门实体图

【例 1.4】 画出商店销售管理系统的 E-R 模型。

商店销售管理系统中的实体存在如下联系：

- (1) 一个部门拥有多个员工，一个员工只属于一个部门。
- (2) 一个员工可开出多个订单，一个订单只能由一个员工开出。
- (3) 一个订单可订购多类商品，一类商品可有多个订单。

商店销售管理系统的 E-R 模型如图 1.14 所示。

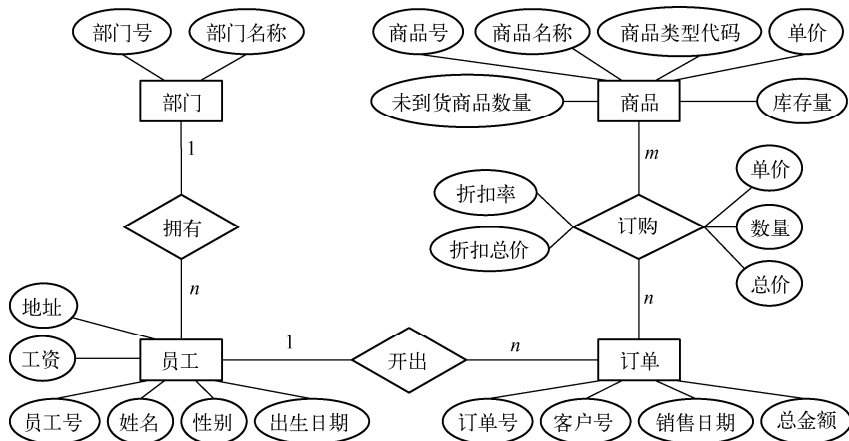


图 1.14 商店销售管理系统的 E-R 模型

1.4.3 数据模型组成要素

数据模型是严格定义的一组概念的集合，一般由数据结构、数据操作、数据完整性约束三部分组成。

1. 数据结构

数据结构用于描述系统的静态特性，是所研究的对象类型的集合。数据模型按其数据结构分为层次模型、网状模型和关系模型等。数据结构研究的对象是数据库的组成部分，包括两类：一类是与数据类型、内容、性质有关的对象，例如关系模型中的域、属性等，另一类是与数据之间联系有关的对象，例如关系模型中反映联系的关系等。

2. 数据操作

数据操作用于描述系统的动态特性，是指对数据库中各种对象及对象的实例允许执行的操作的集合，包括对象的创建、修改和删除，对对象实例的检索、插入、删除、修改及其他有关操作等。

3. 数据完整性约束

数据完整性约束是一组完整性约束规则的集合，完整性约束规则是给定数据模型中数据及其联系所具有的制约和依存的规则。

数据模型三要素在数据库中都是严格定义的一组概念的集合。在关系数据库中，数据结构是表结构定义及其他数据库对象定义的命令集，数据操作是数据库管理系统提供的数据库操作（操作命令、语法规则、参数说明等）命令集，数据完整性约束是各关系表约束的定义及操作约束规则等的集合。

1.4.4 常用的数据模型

常用的数据模型有层次模型、网状模型、关系模型、面向对象数据模型、对象关系数据模型、半结构化数据模型等，下面介绍层次模型、网状模型和关系模型。

1. 层次模型

层次模型采用树状层次结构组织数据，树状结构的每个节点表示一个记录类型，记录类型之间的联系是一对多的联系。层次模型有且仅有一个根节点，位于树状结构顶部，其他节点有且仅有一个父节点。某大学按层次模型组织数据的示例如图 1.15 所示。

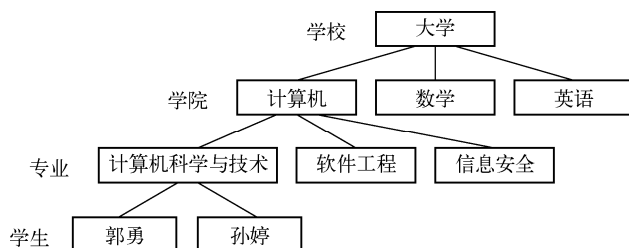


图 1.15 层次模型示例

层次模型简单易用，但现实世界很多联系是非层次性的，如多对多的联系等，表达起来比较笨拙且不直观。

2. 网状模型

网状模型采用网状结构组织数据，网状结构的每个节点表示一个记录类型，记录类型之间可以有多种联系。按网状模型组织数据的示例如图 1.16 所示。

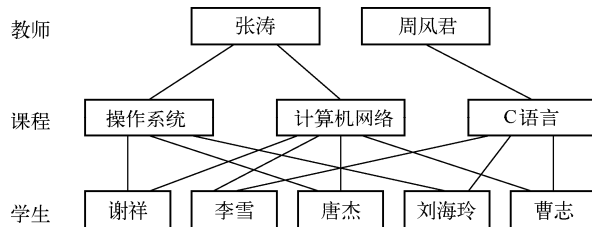


图 1.16 网状模型示例

网状模型可以更直接地描述现实世界，层次模型是网状模型的特例，但网状模型结构复杂，用户不易掌握。

3. 关系模型

关系模型采用关系的形式组织数据，一个关系就是一张二维表，二维表由行和列组成。按关系模型组织数据的示例如图 1.17 所示。

学生关系框架

学号	姓名	性别	出生日期	专业	总学分
----	----	----	------	----	-----

成绩关系框架

学号	课程号	分数
----	-----	----

学生关系

学号	姓名	性别	出生日期	专业	总学分
161001	周浩然	男	1995-09-14	电子信息工程	52
161002	王丽萍	女	1997-02-21	电子信息工程	48
161004	程杰	男	1996-10-08	电子信息工程	50

成绩关系

学号	课程号	分数
161001	1002	93
161002	1002	70
161004	1002	86
161001	1007	92

图 1.17 关系模型示例

关系模型建立在严格的数学概念基础上，数据结构简单清晰，用户易懂易用，关系数据库是目前应用最广泛、最重要的一种数学模型。

1.5 小结

本章主要介绍了以下内容：

- (1) 数据库 (DataBase, DB) 是长期存放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、描述和存储，具有尽可能小的冗余度、较高的数据独立性和易扩张性。

数据库管理系统 (DataBase Management System, DBMS) 是数据库系统的核心组成部分，它是在操作系统支持下的系统软件，是对数据进行管理的大型系统软件。用户在数据库系统中的一些操作都是由数据库管理系统来实现的。

数据库系统 (DataBase System, DBS) 是在计算机系统中引入数据库后的系统构成。数据库系统由数据库、操作系统、数据库管理系统、应用程序、用户、数据库管理员 (DataBase Administrator, DBA) 组成。

(2) 数据管理技术的发展经历了人工管理阶段、文件系统阶段、数据库系统阶段，现在正在向更高一级的数据库系统发展。

(3) 数据库系统的标准结构是三级模式结构，它包括外模式、模式和内模式，数据库管理系统在这三级模式之间提供了两级映射：外模式/模式映射，模式/内模式映射。数据库的三级模式与二级映射使得数据的定义和描述可以从应用程序中分离出去。

(4) 数据模型 (Data Model) 是现实世界数据特征的抽象，根据模型应用的不同目的，按不同的层次可将它们分为两类，第一类是概念模型，第二类是逻辑模型、物理模型。数据模型是数据库系统的核心和基础。

概念模型 (Conceptual Model) 又称信息模型，它按用户的观点对数据和信息进行建模，是描述现实世界的概念化结构。它独立于数据库逻辑结构和具体的 DBMS。概念模型较常用的表示方法是实体-联系模型 (Entity-Relationship Model, E-R 模型)。

数据模型一般由数据结构、数据操作、数据完整性约束三部分组成。常用的数据模型有层次模型、网状模型、关系模型、面向对象数据模型、对象关系数据模型、半结构化数据模型等。

习 题 1

一、选择题

- 1.1 数据库中存储的是_____。
- A. 数据
B. 数据以及数据之间的联系
C. 数据模型
D. 信息
- 1.2 信息的数据表现形式是_____。
- A. 只能是文字
B. 只能是图形
C. 只能是声音
D. 以上皆可
- 1.3 数据库具有_____、最小的冗余度和较高的程序与数据的独立性等特点。
- A. 数据结构化
B. 程序结构化
C. 程序标准化
D. 数据模块化
- 1.4 数据库 (DB)、数据库系统 (DBS) 和数据库管理系统 (DBMS) 的关系是_____。
- A. DBMS 包括 DBS 和 DB
B. DBS 包括 DBMS 和 DB
C. DB 包括 DBS 和 DBMS
D. DBS 就是 DBMS，也就是 DB

