

# 单元 5 数据库技术基础



数据库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。随着计算机技术和相应技术领域的发展，数据库技术得到了极大的发展。

软件是新一代信息技术（Information Technology, IT）的灵魂，是数字经济发展的基础，而数据库作为十分复杂、跨技术领域很多的关键基础软件，一般应用软件的数据处理都要与其进行数据交互，所以数据库对于计算产业的重要性不言而喻。

# 目录

5.1 数据库技术概述

5.2 数据库系统

5.3 数据库管理系统

5.4 关系数据库

5.5 结构查询语言

5.6 非关系数据库

5.7 数据库设计基础



# 5. 1 数据库技术概述

数据库技术就是存储、处理、管理数据的一门计算机技术，是计算机信息系统的重要技术基础和支柱。数据库技术涉及许多基本概念，主要包括信息、数据、数据处理、数据库、数据库管理系统以及数据库系统等。

## 5.1.1 数据库技术的相关概念

### 1. 数据

描述事物的符号可称为数据，数据是数据库中存储的基本对象。例如，新生入学时，一般要填写入学登记，把自己的基本情况写下来，例如：姓名为张三，性别是男，身高为170 cm。那么，张三、男、170 cm就可以称为数据。数据表示了登记者的一种特征或特性。数据种类可以是数字、文字、图形、图像、声音、语言等。

### 2. 数据库

数据库是长期存储在计算机内、有组织的、可共享的大量相关数据的集合。数据库中的数据按一定的数据模型组织、描述和存储。数据库具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

数据库有以下两个比较典型的特点。

① 把在特定的环境中与某应用程序相关的数据及其关系集中在一起，并按照一定的结构形式进行存储，即集成性。

② 数据库中的数据能被多个应用程序所使用，即共享性。

### 3. 数据库管理系统

数据库管理系统是数据库系统的核心组成部分，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。数据库管理系统的功能主要包括数据库的定义和操作、运行和管理、建立和维护等。

数据库管理系统有以下特点。

- ① 采用复杂的数据模型表示数据结构，数据冗余小，易扩充，实现了数据共享。
- ② 具有较高的数据和程序独立性，数据库的独立性有物理独立性和逻辑独立性。
- ③ 提供了方便的用户接口。
- ④ 提供了4个方面的数据控制功能，分别是并发控制、恢复、完整性和安全性。数据库中各个应用程序所使用的数据由数据库管理系统统一规定，按照一定的数据模型组织和建立，由系统统一管理和集中控制。
- ⑤ 增加了系统的灵活性。

### 4. 数据库应用系统

数据库应用系统（Database Application System, DBAS）是在数据库管理系统支持下建立的计算机应用系统。数据库应用系统包括数据库、数据库管理系统、数据库应用程序、数据库管理员、普通用户、硬件，例如，以数据库为基础的财务管理系统、人事管理系统、图书管理系统等。无论是面向内部业务和管理的管理信息系统，还是面向外部提供信息服务的开放式信息系统，从实现技术角度而言，都是以数据库为基础和核心的计算机应用系统。

数据库应用系统有以下特点。

① 结构特性：与数据库状态有关，即与数据模型所反映的实体及实体间的联系的静态特性有关。结构设计就是设计各级数据库模式，决定数据库系统的信息内容，由数据库设计来实现。

② 行为特性：与数据库状态转换有关，即改变实体及其特性的操作。它决定数据库系统的功能，由事务处理等应用程序设计来实现。

根据数据库应用系统的结构和行为两方面的特性，系统设计开发分为两个部分：一部分是作为数据库应用系统核心和基石的数据库设计开发；另一部分是相应的数据库应用软件的设计开发。这两部分是紧密相关、相辅相成的，组成统一的数据库工程。

## 5. 数据库系统

数据库系统（Database System, DBS）是实现有组织地、动态地存储大量关联数据，方便多用户访问计算机的硬件、软件和数据库资源组成的系统，即采用数据库技术的计算机系统。

数据库系统有以下特点。

① 数据结构化，数据的共享性、独立性好，数据存储粒度小，数据库管理系统为用户提供了友好的接口。

② 数据库系统的基础是数据模型，现有的数据库系统均是基于某种数据模型的。

③ 数据库系统的核心是数据库管理系统。

④ 数据库系统一般由数据库、数据库管理系统、数据库应用系统、数据库管理员、普通用户和硬件构成。

## 6. 数据库系统、数据库管理系统、数据库应用系统三者的区别与联系

### (1) 三者的区别

#### ① 本质不同。

数据库系统是一种软件系统，是为满足数据处理的需求而发展起来的一种较为理想的数据处理系统，也是一个为实际可运行的存储、维护和应用系统提供数据的软件系统，是存储介质、处理对象和管理系统的集合体。

数据库管理系统是实现把用户抽象的逻辑数据处理后，转换成计算机中具体的物理数据处理的软件。

数据库应用系统是在数据库管理系统的支持下建立的一种计算机应用系统。

#### ② 组成不同。

数据库系统通常由软件、数据库和数据管理员等组成。其软件主要包括操作系统、各种宿主语言、实用程序以及数据库管理系统。

数据库管理系统由数据库语言和数据库管理例行程序组成。

数据库应用系统由数据库系统、应用程序、用户等组成。

### (2) 三者的联系

三者都用于管理数据库，其中数据库系统和数据库应用系统的组成成分中都包含数据库管理系统，这两者都是通过数据库管理系统来实现对数据库的管理和操控的。

## 5.1.2 数据管理技术的发展

数据库技术是现代信息科学与技术的重要组成部分，是计算机数据处理与信息管理的核心。数据库技术研究和解决了计算机信息处理过程中大量数据有效组织和存储的问题，在数据库系统中实现减少数据存储冗余、数据共享、保障数据安全以及高效地检索数据和处理数据。数据库技术的根本目标是解决数据的共享问题。

数据库技术是因数据管理任务的需要而产生的，数据管理是指对数据进行分类、组织、编码、存储、检索和维护，是数据处理的核心问题。数据管理技术的发展经历了人工管理、文件系统、数据库系统、高级数据库、大数据等多个阶段。

## 5.1.3 数据模型

数据模型（Data Model）是数据特征的抽象，是数据库系统中用以提供信息表示和操作手段的形式构架。数据模型包括数据库数据的结构部分、操作部分和约束条件。

### 1. 数据模型按应用层次分类

数据模型按应用层次分成3种类型，分别是概念数据模型、逻辑数据模型、物理数据模型。

## （1）概念数据模型

概念数据模型（Conceptual Data Model）是指按照用户的观点来对数据和信息进行建模，是一种面向用户、面向客观世界的模型，主要用于数据库设计，用来描述世界的概念化结构。它帮助数据库设计人员在设计的初始阶段，摆脱计算机系统及数据库管理系统的具体技术问题，集中精力分析数据与数据之间的联系等，与具体的数据库管理系统无关。概念数据模型必须换成逻辑数据模型，才能在数据库管理系统中实现。

在概念数据模型中常用的有E-R（Entity-Relationship Approach）模型（主要用实体-联系方法表示，也称为实体-联系模型）、扩充的E-R模型、面向对象模型及谓词模型。

## （2）逻辑数据模型

逻辑数据模型（Logical Data Model）是一种面向数据库系统的模型，是具体的数据库管理系统所支持的数据模型，如层次数据模型（Hierarchical Data Model）、网状数据模型（Network Data Model）等。此模型既面向用户，又面向系统，主要用于数据库管理系统的实现。

## （3）物理数据模型

物理数据模型（Physical Data Model）是一种面向计算机物理表示的模型，描述了数据在存储介质上的组织结构，它不但与具体的数据库管理系统有关，而且还与操作系统和硬件有关。每一种逻辑数据模型在实现时都有其对应的物理数据模型。为了保证其独立性与可移植性，大部分物理数据模型的实现工作由数据库管理系统自动完成，而设计者只设计索引、聚集等特殊结构。



## 2. 基本数据模型

基本数据模型（结构数据模型）是指按照计算机系统的观点来对数据和信息进行建模，主要用于数据库管理系统的实现。基本数据模型是数据库系统的核心和基础，通常由数据结构、数据操作和完整性约束3部分组成。其中数据结构是对系统静态特性的描述，数据操作是对系统动态特性的描述，完整性约束是一组完整性规则的集合。

数据库管理系统总是基于某种基本数据模型，常用的基本数据模型有层次模型、网状模型、关系模型和面向对象模型。

### （1）层次模型

层次模型是数据库技术中发展较早、技术上比较成熟的一种数据模型。它把数据按自然的层次关系组织起来，以反映数据之间的隶属关系。将数据组织成有向、有序的树结构，也叫树形结构。用树形结构表示实体类型及实体间的联系，采用关键字来访问每一层次的每一部分。结构中的节点代表数据记录，连线描述位于不同节点的数据间的隶属关系（一对多的关系）。层次模型的树形结构如图5-8所示。

层次模型的优点是记录之间的联系通过指针来实现，存取方便且速度快，查询效率较高，数据修改和数据库扩展容易实现，检索关键属性十分方便。层次模型的缺点是只能表示一对多联系，虽然有多种辅助手段实现多对多联系，但比较复杂，用户不易掌握。层次顺序的严格和复杂导致数据的查询和更新操作很复杂，应用程序的编写也比较复杂。

## (2) 网状模型

网状模型通过网状结构表示数据间的联系，将数据组织成有向图结构，用有向图表示实体类型及实体间的联系，用连接指令或指针来确定数据间的显式连接关系，是具有多对多类型的数据组织方式。网状模型中的节点代表数据记录，连线描述位于不同节点的数据间的联系，数据之间没有明确的从属关系，一个节点可与其他多个节点建立联系，即节点之间的联系是任意的，任何两个节点之间都能发生联系，可表示多对多的联系。网状模型结构如图5-9所示，该图描述了教研室、教师、课程、学生、任课与选课之间的关系。

网状模型的优点是记录之间的联系通过指针实现，多对多联系也容易实现，查询效率高，能明确而方便地表示数据间的复杂关系。其缺点是编写应用程序的过程比较复杂，程序员必须熟悉数据库的逻辑结构。

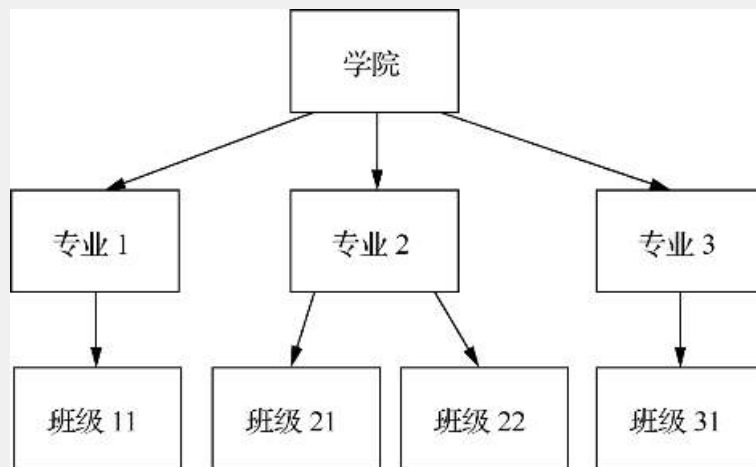


图5-8 层次模型的树形结构

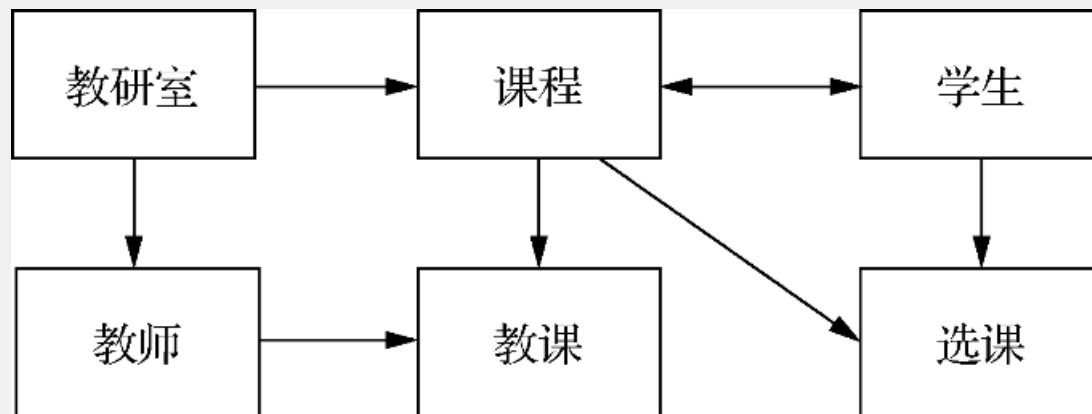


图5-9 网状模型结构

### (3) 关系模型

关系模型用表格结构表达实体集，用外键表示实体间的联系，可以简单、灵活地表示各种实体及其关系。以记录组或数据表的形式组织数据，以便利用各种实体与属性之间的关系进行存储和变换，不分层也无指针，是建立空间数据和属性数据之间的关系的一种非常有效的数据组织方法。关系模型实体之间的联系如图5-10所示。

关系模型的优点是结构简单、清晰、灵活，概念（关系）单一，用户易懂、易用；满足所有由布尔逻辑运算和数学运算规则形成的查询要求；能搜索、组合和比较不同类型的数据；增加和删除数据非常方便；存取路径对用户透明，数据独立性、安全性好，简化了数据库开发工作。

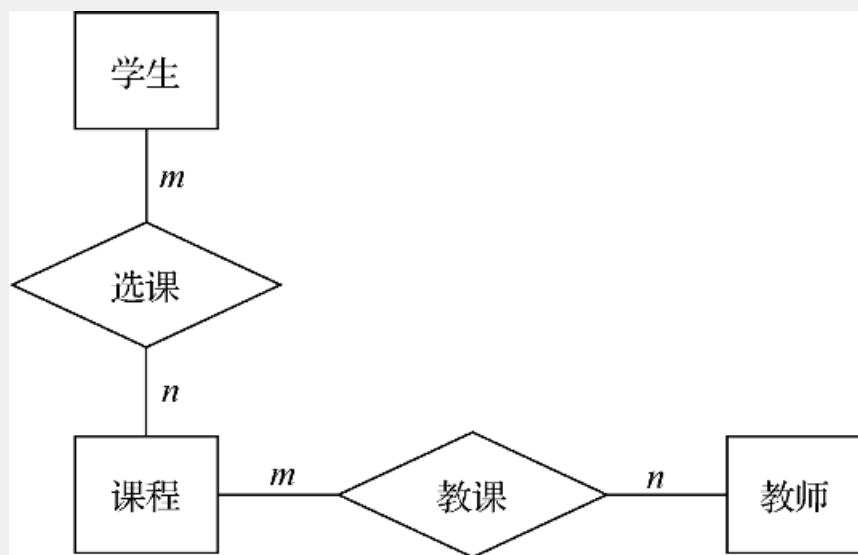


图5-10 关系模型实体之间的联系

#### (4) 面向对象模型

面向对象模型是用面向对象的观点来描述现实世界实体的逻辑组织、对象间限制、联系等的模型。面向对象数据库模式是类的集合，面向对象模型提供了一种类层次结构。在面向对象数据库模式中，一组类可以形成一个类层次，一个面向对象的数据库可能有多个类层次。在一个类层次中，一个类继承其所有超类的全部属性、方法和消息。

# 5. 2 数据库系统

数据库系统是为满足数据处理的需求而发展起来的一种较为理想的数据处理系统，也是一个为实际可运行的存储、维护和应用系统提供数据的软件系统，是存储介质、处理对象和管理系统的集合体。

## 5.2.1 数据库系统的发展阶段

数据库系统的发展主要经历了以下3个阶段。

- ① 第一代数据库系统是指层次模型数据库系统（基于树形结构）和网状模型数据库系统（基于有向图结构）。
- ② 第二代数据库系统是指支持关系模型的关系数据库系统[最先由埃德加·弗兰克·科德(E. F. Codd) 提出关系模型]。
- ③ 面向对象的技术与数据库技术相结合便产生了第三代数据库系统。

## 5.2.2 算法的描述方法

通常，一个完整的数据库系统由数据库、数据库管理系统、数据库应用程序、数据库管理员、普通用户和硬件组成。普通用户与数据库应用程序交互，数据库应用程序与数据库管理系统交互，数据库管理系统访问数据库中的数据。一个完整的数据库系统还应包括硬件，数据库存放在计算机的外存中，数据库应用程序、数据库管理系统等软件都需要在计算机上运行，因此，数据库系统中必然会包含硬件，但本单元不涉及硬件方面的内容。

数据库系统的组成结构如图5-11所示。

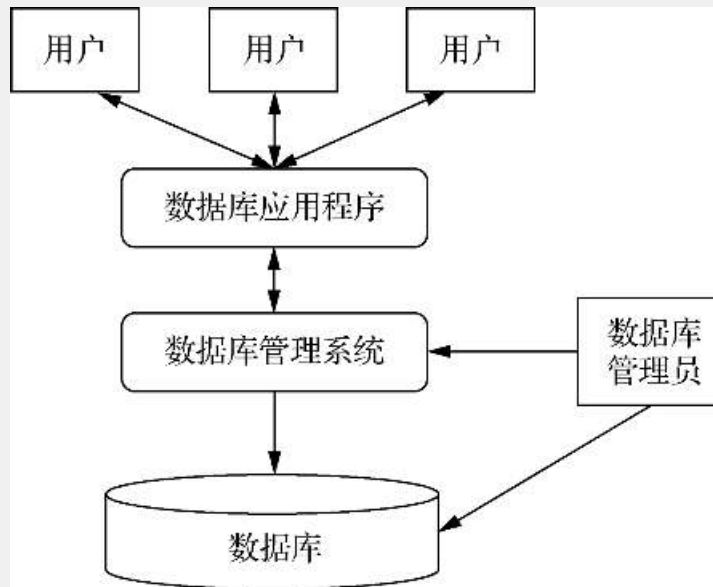


图5-11 数据库系统的组成结构

## 1. 数据库

数据库就是一个长期存储在计算机内、有组织的、集成的、可共享的、统一管理的相关数据集。数据库是一个有结构的数据集，也就是说，数据是按一定的数据模型来组成的，数据模型可用数据结构来描述。数据模型不同，数据的组织结构以及操纵数据的方法也就不同。现在的数据库大多数是以关系模型来组织数据的，可以简单地把关系模型的数据结构（即关系）理解为一张二维表。以关系模型组织起来的数据库称为关系数据库。在关系数据库中，不仅存放着各种用户数据，如与图书有关的数据、与借阅者有关的数据、与借阅图书有关的数据等，而且还存放着与各个表结构定义有关的数据，这些数据通常称为元数据。

数据库具有以下特点。

- ① 数据库是一个集成的数据集，也就是说，数据库中集中存放着各种各样的数据。
- ② 数据库是一个为各种用户共享的数据集，也就是说，数据库中的数据可以被不同的用户使用，每个用户可以按自己的需求访问相同的数据库。
- ③ 数据库是一个统一管理的数据集，也就是说，数据库由数据库管理系统统一管理，任何数据访问都是通过数据库管理系统来完成的。
- ④ 数据库具有冗余度较小、数据间联系紧密、数据独立性较高等特点。

## 2. 数据库管理系统

数据库管理系统是一种用来管理数据库的商品化软件。访问数据库的请求都是通过数据库管理系统来完成的。数据库管理系统提供了对数据库操作的许多命令，这些命令所组成的语言中常用的就是SQL。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件，它为用户或应用程序提供访问数据库的方法，包括数据库的建立、查询、更新以及各种数据控制。

常见的数据库管理系统有Microsoft公司开发的SQL Server、Oracle公司开发的Oracle、Sybase 公司开发的Sybase、IBM公司开发的DB2等。

说明：“数据库管理系统”这一术语通常指的是某个特定厂商的特定数据库产品，如MySQL、Microsoft SQL Server、Microsoft Access、Oracle等，但有时人们使用“数据库”这个术语来代指数据库管理系统，这种用法是不恰当的。甚至还有人用“数据库”这一术语来代指数据库系统，这种用法就更不恰当了。所以对于数据库、数据库管理系统、数据库应用程序、数据库系统等术语要明确其含义，合理使用这些术语。



### 3. 数据库应用程序

虽然已经有了数据库管理系统，但是在很多情况下，数据库管理系统无法满足对数据管理的要求。数据库应用程序（Database Application）的使用可以满足对数据管理的更高要求，还可以使数据管理过程更加直观和友好。数据库应用程序负责与数据库管理系统进行通信、访问和管理数据库管理系统中存储的数据，允许用户插入、修改、删除数据库中的数据。

数据库应用程序是利用某种程序语言，为实现某些特定功能而编写的程序，如查询程序、报表程序等。这些程序为最终用户提供便于使用的可视化界面，最终用户通过界面输入必要的的数据，应用程序接收最终用户输入的数据，对其进行加工处理，并将其转换成数据库管理系统能够识别的SQL语句，然后传给数据库管理系统，由数据库管理系统执行该语句，并负责从数据库若干个数据表中找到符合查询条件的数据，再将查询结果返回给应用程序，应用程序将得到的结果显示出来。由此可见，应用程序为最终用户访问数据库提供了有效途径和简便方法。

### 4. 用户

用户是使用数据库的人员，数据库系统中的用户一般有以下3类。

- (1) 数据库管理员
- (2) 应用程序员
- (3) 最终用户

## 5.2.3 数据库系统的三级模式结构

1975年，美国国家标准协会的计算机与信息处理委员会下属的标准计划和需求委员会（ANSI-SPARC）的数据库系统研究组为数据库系统建立了三级模式结构，即将数据库系统的结构划分为3个层次：外模式、概念模式和内模式。数据库系统的层次结构如图5-12所示。

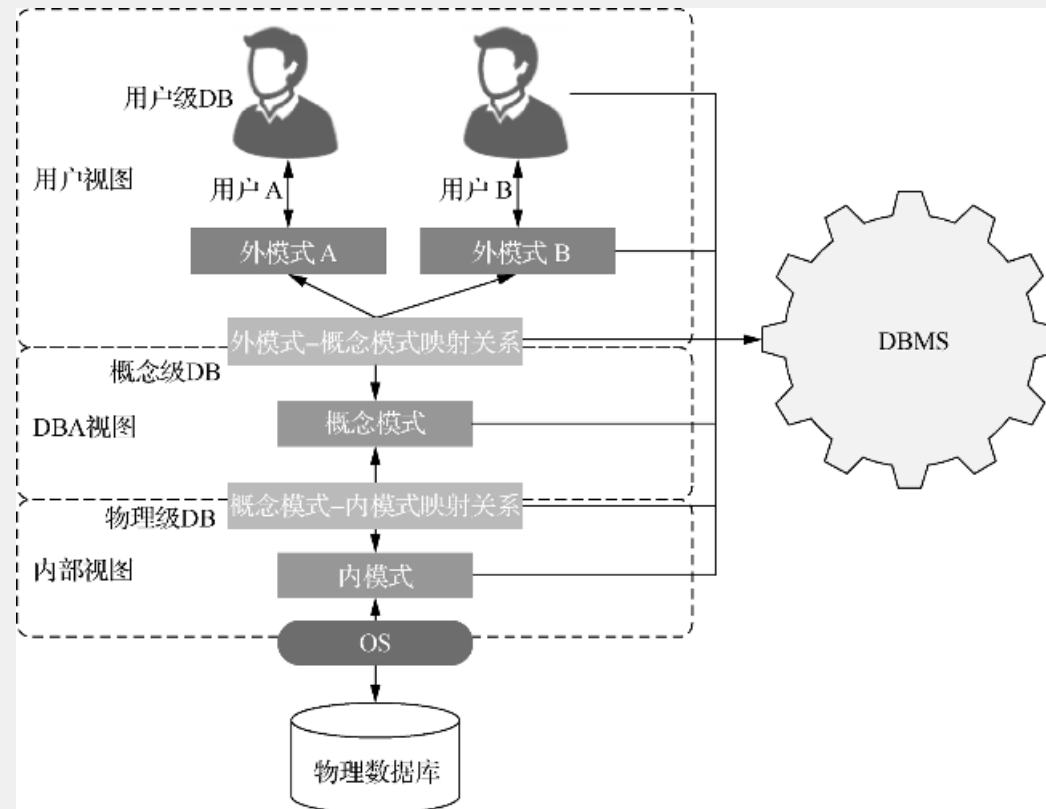


图5-12 数据库系统的层次结构

## 1. 数据库系统结构的3个层次

### (1) 外模式

外模式也称为用户模式或子模式，一个数据库可以有多个外模式，但一个应用程序只能使用一个外模式。外模式主要描述组成用户视图的各个记录的组成、相互关系、数据项的特征、数据的安全性和完整性约束条件等。

外模式是数据库用户（包括应用程序员和最终用户）能够看见和使用的那部分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。

外模式是应用程序与数据库系统之间的接口，是保证数据库安全性的有效措施。用户可使用数据定义语言（Data Definition Language, DDL）和数据操纵语言（Data Manipulation Language, DML）来定义数据库的结构和对数据库进行操纵。用户可以使用数据操作语句或应用程序去操作数据库中的数据，对用户而言，只需要按照所定义的外模式进行操作，而无须了解概念模式和内模式等的内部细节。

### (2) 概念模式

概念模式也称为模式、逻辑模式或关系模式，是所有用户共享的数据库数据视图，它构建了数据项值的框架，每个数据库只有一个概念模式。该模式提供了数据库的整体逻辑结构和特征的完整描述，主要涵盖现实世界实体及其性质与联系，概念模式的定义包括记录、数据项、数据完整性约束条件、安全性约束，以及记录之间的联系等。

概念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，与具体的数据值无关，与具体的应用程序、开发工具及程序设计语言也无关。

### (3) 内模式

内模式也称为存储模式，一个数据库只有一个内模式。内模式是数据库内部数据存储结构和存储方式的描述，是数据在数据库内部的表示方式，它定义了数据库内部记录的类型、存储域的表示、存储记录的物理顺序、文件的组织方式以及数据控制方面的细节等。

内模式是整个数据库的底层表示，不同于物理层，它假设外存是无限的线性地址空间。

概念模式是数据库的中心与关键。内模式、概念模式和外模式之间的关系如下。

- ① 内模式依赖于概念模式，独立于外模式和存储设备。
- ② 外模式面向具体的应用，独立于内模式和存储设备。
- ③ 应用程序依赖于外模式，独立于概念模式和内模式。

### 2. 数据库系统的3级抽象

数据库系统划分为3个抽象级：用户级、概念级、物理级。

#### (1) 用户级数据库

用户级数据库对应于外模式，是非常接近用户的一级数据库，是用户可以看到和使用的数据库，又称为用户视图。用户级数据库主要由外部记录组成，不同的用户级数据库可以互相重叠，用户的操作都是针对用户视图进行的。一个数据库可以有多个不同的用户级数据库，每个用户级数据库由数据库某一部分的抽象表示所组成。

## （2）概念级数据库

概念级数据库对应于概念模式，介于用户级数据库和物理级数据库之间，是所有用户级数据库的最小并集，是数据库管理员可看到和使用的数据库，又称为DBA视图。概念级数据库由概念记录组成，一个数据库应用系统只存在一个DBA视图，它把数据库作为一个整体的抽象表示。概念级数据库把用户级数据库有机地结合成一个整体，综合平衡考虑所有用户要求，实现数据的一致性，最大限度降低数据冗余，准确地反映数据间的联系。

## （3）物理级数据库

物理级数据库对应于内模式，是数据库的底层表示，它描述数据的实际存储组织，是非常接近物理存储的一级数据库，又称为内部视图。物理级数据库由内部记录组成，物理级数据库并不是真正的物理存储，而是非常接近物理存储。

### 3. 数据库系统的两级独立性

数据库系统两级独立性是指物理独立性和逻辑独立性，数据库系统的3个抽象级通过两级映射（外模式-模式映射，模式-内模式映射）进行相互转换，形成统一的整体。

#### （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的数据是相互独立的。当数据的物理存储改变时，应用程序不需要改变。物理独立性存在于概念模式和内模式之间的映射转换，用于说明物理组织发生变化时应用程序的独立程度。

#### （2）逻辑独立性

逻辑独立性是指用户的应用程序与数据库中的逻辑结构是相互独立的。当数据的逻辑结构改变时，应用程序不需要改变。逻辑独立性存在于外模式和概念模式之间的映射转换，用于说明概念模式发生变化时应用程序的独立程度。

逻辑独立性比物理独立性更难实现。

## 5.2.4 几种新型的数据库系统

1. 面向对象数据库系统
2. 并行数据库系统
3. 分布式数据库系统
4. 多媒体数据库系统
5. 演绎数据库系统
6. 主动数据库系统
7. 专家数据库系统
8. 空间数据库系统
9. 工程数据库系统
10. 数据仓库

# 5. 3 数据库管理系统

数据库管理系统是一种操纵和管理数据库的大型软件，是一种能够提供数据录入、修改、查询的数据操作软件。用户通过数据库管理系统访问数据库中的数据，数据库管理员通过数据库管理系统进行数据库的维护工作。

## 5.3.1 数据库管理系统的功能

数据库管理系统用于建立、使用和维护数据库，它可以支持多个应用程序和用户用不同的方法在同一时刻或不同时刻去建立、修改和询问数据库。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。数据库系统中只有数据库管理系统才能直接访问数据库，如MySQL就是一种数据库管理系统，其主要优点是跨平台、开放源代码、速度快、成本低，是目前十分流行的开源小型数据库管理系统。



## 1. 数据定义

数据库管理系统提供数据定义语言（Data Definition Language, DDL），DDL主要用于建立、修改数据库的库结构，DDL所描述的库结构仅仅给出了数据库的框架，通过DDL可以方便地定义数据库中的各种对象。例如，可以使用DDL定义图书借阅数据库中的图书信息数据表、借阅者数据表、图书借阅数据表的表结构。

## 2. 数据操纵

数据库管理系统提供数据操纵语言（Data Manipulation Language, DML），供用户实现对数据的追加、删除、更新、查询等操作。通过DML可以实现数据库中对数据进行的基本操作，例如，向数据表中插入一行数据、修改数据表中的数据、删除数据表中的行、查询数据表中的数据等。

## 3. 安全控制和并发控制

数据库管理系统提供数据控制语言（Data Control Language, DCL）。通过DCL可以控制什么情况下谁可以执行什么样的数据操作。另外，由于数据库是共享的，多个用户可以同时访问数据库（并发操作），这可能会引起访问冲突，从而导致数据的不一致。数据库管理系统还提供并发控制的功能，以避免并发操作时可能带来的数据不一致的问题。

#### 4. 数据库的运行管理

数据库的运行管理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）等功能，这些功能保证了数据库系统的正常运行。

#### 5. 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。

#### 6. 数据库的保护

数据库中的数据是信息社会的战略资源，所以对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、数据库的安全性控制。数据库管理系统的其他保护功能还有系统缓冲区的管理、数据存储的某些自适应调节等。

## 7. 数据库的维护

数据库的维护包括数据库的数据载入、转换、转储，数据库的重组和重构，性能监控等功能，这些功能分别由各个应用程序来完成。

## 8. 数据库事务管理

数据库中的数据是可供多个用户同时使用的共享数据，为保证数据能够安全、可靠地运行，数据库管理系统提供了事务管理功能，该功能保证数据能够并发使用并且不会产生相互干扰的情况，而且在数据库发生故障时能够对数据库进行正确恢复。

## 9. 数据库备份与恢复

数据库管理系统提供了备份数据库和恢复数据库的功能。

## 10. 传送数据与相互通信

数据库管理系统提供负责数据传输的接口，这些接口与操作系统的联机处理、分时系统以及远程作业输入相关。网络环境下的数据库系统还应该具有数据库管理系统与网络中其他软件系统的通信功能以及数据库之间的互操作功能。

## 5.3.2 常用的数据库管理系统产品介绍

目前常用的数据库管理系统产品主要是MySQL、SQL Server、Oracle、Sybase、DB2、Access等产品。不同的数据库管理系统，有不同的特点，也有相对独立的应用领域和用户支持。

### 1. MySQL

MySQL是一种开源小型关系数据库管理系统，MySQL现隶属于Oracle公司。MySQL软件分为社区版和商业版，其体积小、速度快、总体拥有成本低，目前被广泛地应用在中小型网站中。

### 2. SQL Server

SQL Server是由微软公司开发和推广的关系数据库管理系统，具有使用方便、可伸缩性好、与相关软件集成程度高等优点。SQL Server数据库引擎为关系型数据和结构化数据提供了安全、可靠的存储功能，使用户可以构建和管理用于业务的高可用和高性能的数据应用程序。SQL Server的最初版本适用于中小型企业，但是应用范围在不断扩展。

### 3. Oracle

Oracle是Oracle（甲骨文）公司提供的以分布式数据库为核心的一组软件产品，是目前使用最为广泛的数据库管理系统之一。Oracle作为一个通用的数据库管理系统，它具有完整的数据管理

理功能；作为一个关系数据库，它是一款关系完备的产品；作为分布式数据库，它实现了分布式处理功能。

Oracle被认为是业界比较成功的关系数据库管理系统，是一种运行稳定、功能齐全、使用方便、可移植性好、高效率、可靠性好、适应高吞吐量的数据库管理系统。对于数据量大、事务处理繁忙、安全性要求高的企业，Oracle是一个比较理想的选择。随着Internet的普及，Oracle适时地将自己的产品紧密地和网络计算结合起来，成为在Internet应用领域的数据库厂商中的佼佼者。Oracle可以运行在UNIX、Windows等主流操作系统平台上，支持几乎所有的工业标准，并获得了最高级别的ISO标准安全性认证。Oracle采用完全开放策略，可以使客户选择合适的解决方案，同时对开发商提供全力支持。

#### 4. Sybase

Sybase是Sybase公司研制的一种关系数据库管理系统，是一种典型的可运行在UNIX或Windows平台上C/S环境下的大型数据库管理系统。Sybase提供了一套应用程序编程接口，可以与非Sybase数据源及服务器集成，允许在多个数据库之间复制数据，适用于创建多层应用。

## 5. DB2

DB2是IBM公司研制的关系数据库管理系统，主要应用于大型应用系统，具有较好的可伸缩性，可支持从大型计算机到单用户环境。DB2有很多不同的版本，可以运行在从掌上产品到大型计算机的不同的终端上。DB2提供了高层次的数据利用性、完整性、安全性、可恢复性，以及从小规模到大规模应用程序的执行能力，支持无关平台的基本功能和SQL命令。

1968年，IBM公司推出的信息管理系统（Information Management System, IMS）是层次模型数据库系统的典型代表，是第一个大型的商用数据库管理系统。1970年，IBM公司的研究员首次提出了数据库系统的关系模型，开始了对数据库关系方法和关系数据理论的研究，为数据库技术奠定了基础。IBM公司在关系数据库理论方面一直走在业界的前列。DB2于1983年首次发布，2001年，IBM公司兼并了数据库公司Informix，并将其所拥有的先进特性融入DB2中，使DB2的系统性能和功能有了进一步提高。

## 6. Access

Access是微软公司在1992年推出的一个入门级小型桌面数据库管理系统，它具有界面友好、易学易用、开发简单、接口灵活，可以方便地生成各种数据对象，利用存储的数据建立窗体和报表等特点。作为Office套件的一部分，Access可以与Office集成，实现无缝连接，是一种典型的桌面数据库管理系统。但其性能和安全性都一般。Access能够利用Web检索和发布数据，实现与互联网的连接，主要适用于中小型企业应用系统，或作为C/S系统中的客户端数据库。

## 5.3.3 国产数据库管理系统简介

随着数字经济的发展，数据量爆炸式增长，国产化替代加速，国产数据库迎来巨大机遇。从国内市场竞争格局来看，华为、阿里巴巴、达梦数据库、人大金仓等国产数据库供应商占据的市场份额越来越大。

国产数据库管理系统厂商经过多年的技术研发和经验积累，经历引进技术、研究、创新再自主创新，实力不断增强，产品应用越来越广泛，专利申请数量越来越多，国产阵营日益强盛，完全实现国产数据库指日可待。

1. 华为GaussDB: AI-Native分布式数据库
2. 华为openGauss: 开源关系数据库
3. 阿里云PolarDB: 关系型云原生分布式数据库
4. 蚂蚁集团OceanBase: 原生分布式关系数据库
5. 阿里云AnalyticDB: 实时分析型数据库
6. 达梦数据库DM8: 通用关系数据库
7. 人大金仓KingbaseES: 关系数据库
8. 南大通用GBase: 数据库产品
9. PingCAP TiDB: 分布式关系数据库
10. 腾讯云TDSQL: 分布式数据库

# 5. 4 关系数据库

关系数据库是指使用关系模型来组织和管理数据的数据库。

## 5.4.1 关系的基本运算

关系运算是一个数学名词。基本运算有两类：一类是传统的集合运算（并、差、交等），另一类是专门的关系运算（选择、投影、连接、除法等）。有些查询需要进行几个基本运算的组合，经过若干步骤才能完成。

### 1. 传统的集合运算

传统的集合运算包括并（Union）、差（Difference）、交（Intersection）、笛卡儿积等，均为二目运算。这些针对集合的运算是以元组为运算的基本元素进行的，是从行的角度展开的运算。



## 2. 专门的关系运算

关系运算的运算对象是关系，运算结果亦为关系。

### (1) 选择运算

从关系中找出满足给定条件的那些元组（行）称为选择（Selection）。其中的条件是以逻辑表达式给出的，值为真的元组将被选取。这种运算是从水平方向抽取元组。

### (2) 投影运算

从关系模式中挑选若干属性（列）组成新的关系称为投影（Projection）。这是从列的角度进行的运算，相当于对关系进行垂直分解。

投影之后不仅取消了原关系中的某些列，而且还可能取消某些元组，这是因为取消了某些属性（列）后，可能出现重复的行，应该取消这些完全相同的行。

### (3) 连接运算

连接（Join）运算是从两个关系的笛卡儿积中选择属性满足一定条件的元组，将不同的两个关系连接成一个新关系。新关系中的元组是通过连接原有关系的元组而得到的。新关系中属性的名字采用原有关系属性名加上原有关系名作为前缀，这种命名方法保证了新关系中属性名的唯一性，尽管原有不同关系中的属性可能是同名的。

### (4) 除法运算

在关系代数中，除法运算可理解为笛卡儿积的逆运算。

## 5.4.2 关系数据库概述

关系数据库是一种数据库类型，关系数据库是指采用了关系模型来组织数据的数据库，其以行和列的形式存储数据，以便用户理解和快速访问。关系模型可以简单理解为二维表格模型，关系模型由关系数据结构、关系操作集合、关系完整性约束3部分组成。而关系数据库就是由二维表及其之间的关系组成的数据组织。

在关系数据库中，数据存放在包含一系列行和列的二维表中，一个关系数据库包含多个二维表。关系数据库所包含的表之间是有关联的，关联主要由主键和外键所体现的参照关系实现。用户通过查询来检索数据库中的数据，而查询是一系列用于限定数据库中某些区域的执行代码。

关系数据库的**优点**如下。

- ① 易于维护：都使用表结构，格式一致。
- ② 使用方便：通用的SQL可实现复杂查询。
- ③ 复杂操作：支持SQL，可实现一个表及多个表之间非常复杂的查询。

关系数据库的**缺点**如下。

- ① 读写性能比较差，尤其是对海量数据的读写效率。
- ② 表的结构固定，灵活度稍欠缺。
- ③ 难以满足高并发读写需求，对传统关系数据库来说，硬盘I/O是一个很大的瓶颈。

传统关系数据库旨在处理大量结构化数据，这使得关系数据库特别适用于处理结构化的大数据，因为它们依赖于SQL，并且可以使用数据库管理系统控制数据。但是，更大、更复杂的数据集包含的数据种类越来越多，意味着数据的结构化程度越来越低，且来自新的源，这就需要使用支持处理非结构化或半结构化数据的非关系数据库（如NoSQL）。

## 5.4.3 关系数据库的相关概念

### 1. 关系模型

关系模型是一种以二维表的形式表示实体数据和实体之间联系的数据模型，关系模型的数据结构是一个由行和列组成的二维表格，每个二维表称为关系，每个二维表都有一个名字，如“图书信息”“出版社”等。目前大多数数据库管理系统所管理的数据库都是关系数据库，MySQL数据库就是关系数据库。

例如，表5-5所示的“图书信息”数据表中和表5-6所示“出版社信息”数据表就是两张二维表，分别描述“图书”实体对象和“出版社”实体对象。

另外“图书信息”数据表和“出版社信息”数据表有一个共同字段，即出版社ID，在“图书信息”数据表中该字段的命名为“出版社”，在“出版社信息”数据表中该字段的命名为“出版社ID”，虽然命名有所区别，但其数据类型、长度相同，字段值有对应关系，这两个数据表可以通过该字段建立关联。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/928052031064007006>