

理 论 篇

第 1 章 数据库系统基础知识

数据库技术产生于 20 世纪 60 年代，它的出现得益于计算机技术的产生和发展，而它又使计算机得到更广泛的应用。数据库技术是一门研究如何存储、使用和管理数据的技术，是计算机数据管理的最新发展阶段，它能把大量的数据按照一定的结构存储起来，在数据库管理系统的集中管理下实现数据共享。数据库技术是计算机领域的一个重要分支。在计算机应用的三大领域（科学计算、数据处理和过程控制）中，数据处理约占其中的 70%，而数据库技术就是作为一门数据处理技术发展起来的。在信息技术日益普及的今天，现代信息社会各行各业中都涉及对数据的运用和管理，数据库技术的作用变得越来越重要。作为信息系统的核心技术和基础，数据库系统几乎触及到人类社会生活的各个方面，成为信息技术中一种必备的技能。

1.1 数据库系统的基本概念

1.1.1 数据管理技术的发展

对数据进行加工计算，产生出新数据的过程称为数据处理。数据处理的核心是数据管理，即对数据进行收集、分类、组织、存储、编码、检索和维护等操作。随着计算机软硬件技术的发展，数据管理技术也在不断发展，计算机进行数据管理大概经历了如下三个阶段。

1. 人工管理阶段（20 世纪 50 年代）

人工管理阶段处于 20 世纪 50 年代中期以前，计算机主要用于数值计算。在这一阶段，硬件方面，外存储器只有卡片机、纸带机或磁带机，没有像硬盘一样可以快速、随时存储的外存储器；软件方面，没有操作系统和数据管理软件支持，数据处理方式是批处理。在这一管理方式下，应用程序与数据之间不可分割，当数据有所变动时，程序则随之改变，数据的独立性差；另外，各个应用程序之间的数据不能相互传递，数据无法共享。在人工管理阶段，数据作为应用程序的组成部分不能独立存在，由程序员在应用程序中进行管理，无专门的软件对数据进行管理。

人工管理阶段的数据处理具有以下特点：

- （1）数据不能保存。
- （2）数据不具有独立性。
- （3）数据不能共享。
- （4）数据的冗余度极大。
- （5）没有相应的软件系统负责数据的管理工作。

2. 文件管理阶段（20 世纪 60 年代）

20 世纪 50 年代后期至 20 世纪 60 年代中、后期，随着硬件方面磁鼓、磁盘等存储器的研制并投入使用，以及软件方面高级语言和操作系统软件的出现，计算机的应用不仅仅局限于数值计算，同时也开始以“文件”的方式介入数据处理。在文件管理阶段，是把有关的数据组织成数据文件，这种数据文件可脱离应用程序而独立存在，可长期保存在硬盘中并多次存取。由于使用专门的文件管理系统实施数据管理，应用程序与数据文件相互独立，同时数据的逻辑结构与物理结构之间也具有一定的相对性。

在文件管理阶段，数据可以以操作系统文件的形式长期存在计算机中，操作系统的文件管理系统提供了对数据的输入和输出操作接口，进而提供数据存取方法，但没有专门的应用软件对数据进行管理。

文件管理阶段的数据处理具有如下特点：

- （1）数据由专门的软件（文件系统）负责管理。
- （2）数据有了一定的独立性。
- （3）数据能长期保存。
- （4）数据文件不再只从属于一个应用程序。
- （5）数据仍有一定的冗余。
- （6）数据有可能存在不一致性。

3. 数据库管理阶段（20 世纪 60 年代后期）

进入 20 世纪 60 年代后期，随着计算机应用的日益发展，计算机用于数据处理的范围越来越广，数据处理的数据量也越来越大，仅仅基于文件系统的数据库管理技术已很难满足应用的需求。与此同时，计算机硬件技术也在飞速发展，磁盘存储技术取得重大突破，大容量磁盘开始进入市场，数据库管理软件的改善成为许多软件公司的重要目标。在实际需求迫切、硬件与软件竞相拓展的环境中，数据库系统应运而生。在数据库系统阶段，应用程序独立于数据，实现了数据的独立性，并且出现了专门管理数据的软件——数据库管理系统（DataBase Management System，简称 DBMS）。数据由数据库管理系统统一管理和控制，数据库管理系统还提供了数据的安全性、完整性、并发控制和数据库恢复等功能。数据管理中数据的定义、操作及控制统一由数据库管理系统来完成。其特点如下：

（1）数据结构化

文件系统中的文件不存在联系，从总体上看其数据是没有结构的，而在数据库系统中将各种应用的数据按一定的结构形式（即数据模型）组织到一个结构化的数据库中，不仅考虑了某个应用的数据结构，而且考虑了整个组织（即多个应用）的数据结构。也就是说数据库中的数据不再仅仅针对某个应用，而是面向全组织；不仅数据内部是结构化的，整体也是结构化的；不仅描述了数据本身，也描述了数据间的有机联系，从而较好地反映了现实世界事物间的自然联系。

（2）数据的高共享性与低冗余性

数据的结构化使得数据可为多个应用所共享，而数据共享又可极大地减少数据冗余，不仅减少了不必要的存储空间，更为重要的是可以避免数据的不一致性。所谓数据的一致性是指在系统中同一数据的不同出现场合应保持相同的值，减少数据冗余是保证系统一致性的基础。

（3）数据独立性

数据独立性是数据与程序间的互不依赖性，即数据库中的数据独立于应用程序而不依赖

于应用程序。也就是说,数据的逻辑结构、存储结构与存取方式的改变不会影响应用程序,即物理独立性和逻辑独立性。当然,数据独立性的实现需要模式间的映射关系作为保障。

(4) 用户接口

数据库管理系统作为用户与数据库的接口,对数据进行集中管理,并支持其他程序设计语言,方便用户编程需要。

1.1.2 数据、数据库、数据库管理系统

1. 数据 (Data)

数据是数据库系统研究和处理的对象,从本质上讲是描述事物的符号记录。符号不仅是指数字、字母和文字,而且包括了图形、图像、声音等。因此数据有多种表示形式,都是经过数字化处理后存入计算机并能够反映或描述事物的特征。

2. 数据库 (DataBase, 简称 DB)

数据库是将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、较高的数据独立性和易扩展性,并可为各种用户共享的数据集合。

通常,收集并抽取一个应用所需要的大量数据之后,应该将其保存起来以供进一步加工处理和抽取有用信息。保存方法有多种,尤其以保存在数据库中最佳。因为它一般由相互关联的数据表组成,能使数据冗余度尽可能得小。

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

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,属于系统软件。它是数据库系统的一个重要组成部分,是使数据库系统具有数据共享、并发访问、数据独立等特性的根本保证,主要提供以下功能:

- (1) 数据定义功能。
- (2) 数据操纵及查询优化。
- (3) 数据库的运行管理。
- (4) 数据库的建立和维护。

Microsoft Access 就是一个关系型数据库管理系统(简称 RDBMS),它提供一个软件环境,利用它用户可以方便快捷地建立数据库,并对数据库中的数据实现查询、编辑、打印等操作。

4. 数据库管理员 (DataBase Administrator, 简称 DBA)

由于数据库的共享性,因此对数据库的规划、设计、维护、监视等需要有专人管理,从事这方面工作的人员称为数据库管理员。其主要工作如下:

(1) 数据库设计。DBA 主要任务之一是进行数据库设计,具体地说就是进行数据模式的设计。由于数据库的集成与共享性,因此需要有专门人员对多个应用的数据需求进行全面的规划、设计与集成。

(2) 数据库维护。DBA 必须对数据库中的数据安全性、完整性、并发控制及系统恢复、数据定期转储等进行实施与维护。

(3) 改善系统性能,提高系统效率。DBA 必须随时监视数据库运行状态,不断调整内部结构,使系统保持最佳状态与最高效率。当数据库运行效率下降时,DBA 需采取适当的措施,如进行数据库的重组、重构等。

5. 数据库系统 (DataBase System, 简称 DBS)

数据库系统通常是指带有数据库的计算机应用系统。它一般由数据库(DB)、数据库管理

系统（及其开发工具）（DBMS）、软件（应用）系统、硬件系统、数据库管理员和用户几部分组成。在不引起混淆的情况下常把数据库系统简称为数据库。

6. 数据库应用系统（DataBase Application System, 简称 DBAS）

利用数据库系统进行应用开发可构成一个数据库应用系统，数据库应用系统是数据库系统再加上应用软件和用户界面组成，具体包括数据库、数据库管理系统、数据库管理员、硬件平台、软件平台、应用软件、应用界面，其结构如图 1-1 所示。

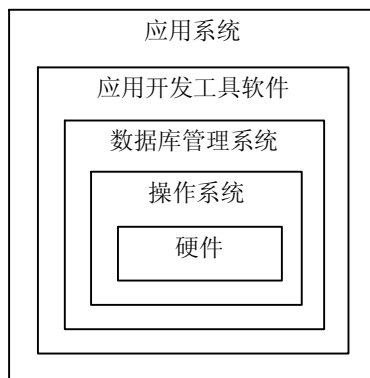


图 1-1 数据库系统软硬件层次结构图

1.1.3 数据库系统发展的主要特征

计算机技术迅速发展，其主要应用领域从数值计算转移到数据事务处理，从而出现了数据库技术，它是数据管理的最新技术，是计算机科学中发展最快、应用最广泛的重要分支之一。在短短的三十几年里，数据库技术的发展经历了三代：第一代为层次、网状数据库系统，第二代为关系数据库系统，第三代为以面向对象模型为主要特征的数据库系统。目前，数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算机技术等相互渗透，成为数据库系统发展的主要特征。

1. 第一代数据库系统——层次、网状数据库系统

数据库发展阶段的划分是以数据模型的发展为主要依据的。数据模型的发展经历了格式化数据模型（包括层次数据模型和网状数据模型）、关系数据模型两个阶段，并正向面向对象的数据模型等非传统数据模型阶段发展。实际上层次数据模型是网状数据模型的特例，层次数据库系统和网状数据库系统在体系结构、数据库语言和数据存储管理上均具有相同特征，并且都是在 20 世纪 60 年代后期研究和开发的，属于第一代数据库系统。

第一代数据库系统具有如下特点：

（1）支持三级模式的体系结构。三级模式通常指外模式、模式、内模式，模式之间具有转换功能。

（2）用存取路径来表示数据之间的联系。数据库系统不仅存储数据，而且存储数据之间的联系。在层次和网状数据库系统中，数据之间的联系是用存取路径来表示和实现的。

（3）独立的数据定义语言。层次数据库系统和网状数据库系统有独立的数据定义语言，用以描述数据库的外模式、模式、内模式以及相互映像。三种模式一经定义，就很难修改。这就要求数据库设计人员在建立数据库应用系统时，不仅要充分考虑用户的当前需求，还要充分

了解可能的需求变化和发展。

(4) 导航的数据操纵语言。层次数据库和网状数据库的数据查询和数据操纵语言是一个记录的导航式的过程化语言。这类语言通常嵌入某一种高级语言如 COBOL、Fortran、PL/I 中,其优点是存取效率高;缺点是编程繁琐,应用程序的可移植性较差,数据的逻辑独立性也较差。

第一代数据库系统的代表是:

(1) 1969 年,IBM 公司开发的层次模型的数据库系统 (Information Management System, IMS), 它可以多个程序共享数据库。

(2) 1969 年 10 月,美国数据库系统语言协会 (Conference On Data System Language, CODASYL) 的数据库研制者提出了网状模型数据库系统规范报告,称为 DBTG (Data Base Task Group) 报告,使数据库系统开始走向规范化和标准化。它是数据库网状模型的典型代表。

2. 第二代数据库系统——关系数据库系统

1970 年美国 IBM 公司 San Jose 研究室的高级研究员埃德加·考特 (E. F. Codd) 发表了论文《大型共享数据库数据的关系模型》,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,奠定了关系数据库技术的理论基础,为数据库技术开辟了一个新时代。

20 世纪 70 年代,关系方法的理论研究和软件系统的研制均取得了很大成果。IBM 公司的 San Jose 实验室研制出关系数据库实验系统 System R。与 System R 同期,美国 Berkeley 大学也研制了 INGRES 数据库实验系统,并发展成为 INGRES 数据库产品,使关系方法从实验室走向了市场。

关系数据库产品一经问世,就以其简单清晰的概念、易懂易学的数据库语言,使得用户不需了解复杂的存取路径细节,不需说明“怎么干”,只需指出“干什么”,就能操作数据库,从而深受广大用户喜爱。

20 世纪 80 年代以来,大多数厂商推出的数据库管理系统的产品都是关系型的,如 FoxPro、Access、DB2、Oracle 及 Sybase 等都是关系型数据库管理系统 (简称 RDBMS),使数据库技术日益广泛地应用到企业管理、情报检索、辅助决策等各个方面,成为实现和优化信息系统的基本技术。

关系数据库是以关系模型为基础的,具有以下特点:

(1) 关系数据库对实体及实体之间的联系均采用关系来描述,对各种用户提供统一的数据结构形式,使用户容易掌握和应用。

(2) 关系数据库语言具有非过程化特性,将用户从数据库记录的导航式检索编程中解脱出来,降低了编程难度,可面向非专业用户。

(3) 数据独立性强,用户的应用程序、数据的逻辑结构与数据的物理存储方式无关。

(4) 以关系代数为基础,数据库的研究更加科学化,尤其在关系操作的完备性、规范化及查询优化等方面,为数据库技术的成熟奠定了很好的基础。

3. 第三代数据库系统

第一代和第二代数据库技术基本上是处理面向记录、以字符表示为主的数据,能较好地满足商业事务处理的需求,但远远不能满足多种多样的信息类型处理需求。新的数据库应用领域如计算机辅助设计/制造 (CAD/CAM)、计算机集成制造 (CIM)、办公信息系统 (OIS) 等需要数据库系统能支持各种静态和动态的数据,如图形、图像、语音、文本、视频、动画、音

乐等,并且还需要数据库系统具备处理复杂对象、实现程序设计语言和数据库语言无缝集成等能力。这种情况下,原有的数据库系统就暴露出了多种局限性。正是在这种新应用的推动下,数据库技术得到进一步发展。

1990 年高级 DBMS 功能委员会发表了《第三代数据库系统宣言》,提出了第三代数据库应具有的三个基本特征,并从三个基本特征导出了 13 个具体特征和功能。

经过多年的研究和讨论,对第三代数据库系统的基本特征已有了如下共识:

(1) 第三代数据库系统应支持数据管理、对象管理和知识管理。以支持面向对象数据模型为主要特征,并集数据管理、对象管理和知识管理为一体。

(2) 第三代数据库系统必须保持或继承第二代数据库系统的技术,如非过程化特性、数据独立性等。

(3) 第三代数据库系统必须对其他系统开放,如支持数据库语言标准、在网络上支持标准网络协议等。

4. 数据库技术的新进展

20 世纪 80 年代以来,数据库技术经历了从简单应用到复杂应用的巨大变化,数据库系统的发展呈现出百花齐放的局面,目前在新技术内容、应用领域和数据模型三个方面都取得了很大进展。

数据库技术与其他学科的有机结合,是新一代数据库技术的一个显著特征,从而出现了各种新型的数据库,例如:

- 数据库技术与分布处理技术相结合,出现了分布式数据库。
- 数据库技术与并行处理技术相结合,出现了并行数据库。
- 数据库技术与人工智能技术相结合,出现了知识库和主动数据库系统。
- 数据库技术与多媒体处理技术相结合,出现了多媒体数据库。
- 数据库技术与模糊技术相结合,出现了模糊数据库等。
- 数据库技术应用到其他领域中,出现了数据仓库、工程数据库、统计数据库、空间数据库及科学数据库等多种数据库技术,扩大了数据库应用领域。

数据库技术发展的核心是数据模型的发展。数据模型应满足三方面的要求:一是能比较真实地模拟现实世界;二是容易为人们所理解;三是便于在计算机上实现。目前,一种数据模型要很好地满足这三方面的要求是很困难的。新一代数据库技术则采用多种数据模型,例如面向对象数据模型、对象关系数据模型、基于逻辑的数据模型等。

1.1.4 数据库系统的体系结构

数据库系统的体系结构包括三级模式和两级映射,三级模式分别为外模式、概念模式和内模式;两级映射分别为外模式与概念模式间的映射以及概念模式与内模式间的映射,其抽象结构关系如图 1-2 所示。

1. 数据库系统的三级模式

数据模式是数据库系统中数据结构的一种表示形式,它具有不同的层次与结构方式。

(1) 外模式 (External Schema)。外模式又称为用户模式或子模式,是某个或某几个数据库用户所看到的数据库的数据视图。外模式是与某一应用有关的数据的逻辑结构和特征描述,也就是前面所介绍的局部结构,它由概念模式推导而来。概念模式给出了系统全局的数据描述,而外模式则给出每个用户的局部数据描述。对于不同的数据库用户,由于需求的不同,外模式

的描述也互不相同。一个概念模式可以有若干个外模式，每个用户只关心与其有关的外模式，这样有利于数据保护，对数据所有者和用户都极为方便。

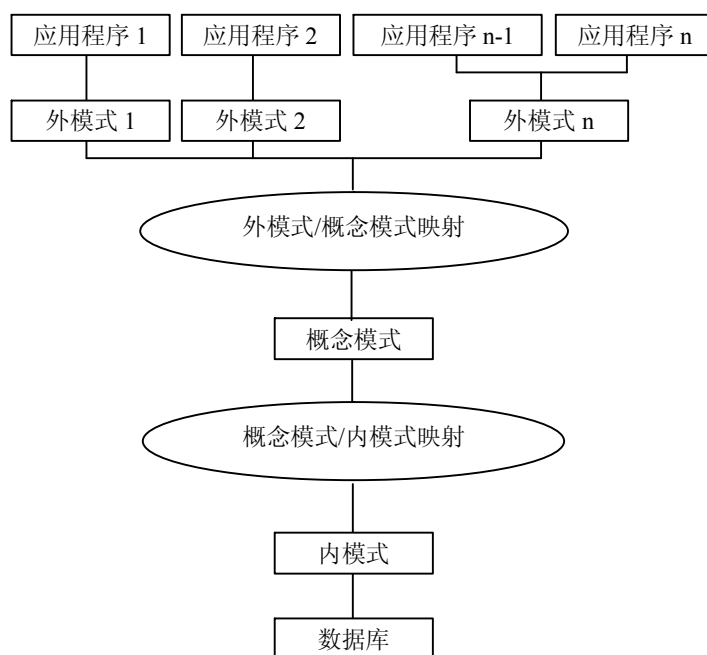


图 1-2 数据库三级模式结构与两级映射

（2）概念模式（Conceptual Schema）。概念模式又称为模式或逻辑模式，它介于内模式与外模式之间，是数据库设计者综合各用户的数据，按照统一的需求构造的全局逻辑结构，是对数据库中全部数据的逻辑结构和特征的总体描述，是所有用户的公共数据视图。外模式涉及的仅是局部的逻辑结构，通常是概念模式的子集。概念模式是用模式描述语言来描述的，在一个数据库中只有一个概念模式。

（3）内模式（Internal Schema）。内模式又称为存储模式或物理模式，是数据库中全体数据的内部表示，它描述了数据的存储方式和物理结构，即数据库的“内部视图”。它是数据库的底层描述，定义了数据库中各种存储记录的物理表示、存储结构与物理存取方式，如数据存储文件的结构、索引、集簇等存取方式和存取路径等。内模式是用模式描述语言严格定义的，在一个数据库中只有一个内模式。

在数据库系统体系结构中，三级模式是根据所描述的三层体系结构的三个抽象层次来定义的，外模式处于最外层，它反映了用户对数据库的实际要求；概念模式处于中间层，它反映了设计者对数据的全局逻辑要求；内模式处于最内层，它反映了数据的物理结构和存取方式。

2. 数据库系统的两级映射

数据库系统的三级模式是数据的三个级别的抽象，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表示和存储。为实现三个抽象层次间的联系和转换，数据库系统在三个模式间提供了两级映射。

（1）外模式与概念模式间的映射。该映射定义了外模式与概念模式之间的对应关系，保

证了逻辑数据的独立性，即外模式不受概念模式变化的影响。

(2) 概念模式与内模式间的映射。该映射定义了内模式与概念模式之间的对应关系，保证了物理数据的独立性，即概念模式不受内模式变化的影响。

1.2 数据模型

数据模型是表示事物以及事物之间联系的模型，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。数据模型是数据库系统的核心和基础，现有的数据库系统均是基于某种数据模型的。

数据模型有三个基本组成要素：数据结构、数据操作和完整性约束。

(1) 数据结构：是目标类型的集合，用于描述系统的静态特性，研究的对象包括两类，一类是与数据类型、内容、性质有关的对象；另一类是与数据之间的联系有关的对象。

(2) 数据操作：是指对数据库中各种对象（型）的实例（值）允许执行的所有操作，即操作的集合，包括操作及有关的操作规则。数据库主要有检索和更新两类操作。

(3) 完整性约束：是给定的数据模型中数据及其联系所具有的制约和依存规则，用以限定数据库的状态及状态的变化，以保证数据的正确、有效和相容。

数据模型按不同的应用层次分成三种类型：概念数据模型、逻辑数据模型、物理数据模型。

概念数据模型简称概念模型，它是一种面向客观世界、面向用户的模型，与具体的数据库管理系统和计算机平台无关。概念模型着重于对客观世界复杂事物的结构及它们之间的内在联系的描述。概念模型是整个数据模型的基础，设计概念模型常用的方法是 E-R 方法，也就是 E-R 模型（实体—联系模型）。

逻辑数据模型又称为数据模型，它是一种面向数据库系统的模型，该模型着重于数据库系统级别的实现。概念模型只有在转换成数据模型后才能在数据库中得以表示。数据库领域中过去和现在最常见的数据模型有四种：层次模型（Hierarchical Model）、网状模型（Network Model）、关系模型（Relational Model）和面向对象模型（Object Oriented Model），其中层次模型和网状模型统称为非关系模型。在关系模型出现以前，它们是非常流行的数据模型。

物理数据模型又称为物理模型，它是一种面向计算机物理表示的模型，此模型给出了数据模型在计算机上物理结构的表示。

1.2.1 E-R 模型

概念模型的表示方法很多，其中最为著名和被广泛使用的是 P.P.S.Chen 于 1976 年提出的实体—联系方法（Entity-Relationship approach, E-R），简称 E-R 模型。该模型将现实世界的要求转化成实体、联系、属性等几个基本概念，以及它们之间的两种基本联接关系，E-R 模型可用图形直观地表示，称为 E-R 图。

(1) 实体（集）用矩形来表示，在框内标注实体（集）的名字，如图 1-3 中的学生、课程实体（集）。

(2) 属性用椭圆形来表示，在框内标注属性名，并用无方向箭头线段连接属性框与实体（集）框，如图 1-3 中的学号、姓名等属性。

(3) 联系用菱形表示，在框内标注联系名，并用无方向箭头线段连接联系框和实体（集）框，同时在连线段上标注联系的类型，如图 1-3 中的选课联系。

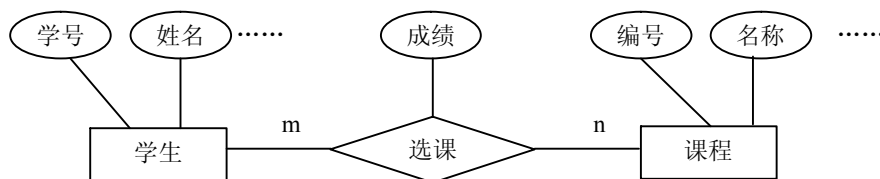


图 1-3 学生选课 E-R 图

1.2.2 层次模型

层次模型（Hierarchical Model）用树形结构来表示数据间的从属关系结构，其主要特征如下：

- 有且仅有一个无双亲结点，这个结点称为根结点。
- 其他结点向上仅有一个双亲结点，向下有若干子结点。

如图 1-4 所示的层次模型就像一棵倒置的树，根结点在上，层次最高；子结点在下，逐层排列。同一双亲的子结点称为兄弟结点，没有子结点的结点称为叶结点。

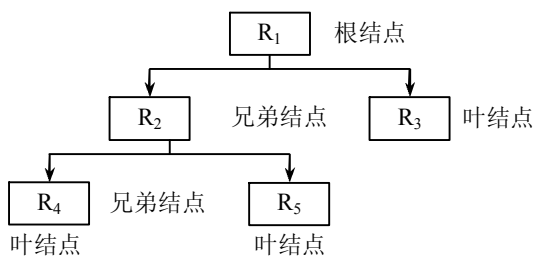


图 1-4 层次模型示例

层次数据模型比较简单，结构清晰，容易理解。但由于现实世界中很多联系是非层次的，采用层次模型表示这种非层次的联系很不直接，只能通过冗余数据或创建非自然的数据组织来解决。

1.2.3 网状模型

网状模型（Network Model）是层次模型的扩展，呈现一种交叉关系的网络结构，可以表示较复杂的数据结构。其主要特征如下：

- 可以有一个以上的结点无双亲结点。
- 一个结点可以有多个双亲结点。

在网状模型中，子结点与双亲结点的联系可以不唯一。因此，要为每个联系命名，并指出与该联系有关的双亲记录和子记录。

学生选课数据库可以采用网状模型。学生选课时一个学生可以选修多门课程，一门课程也可以被多个学生选修，学生与课程是多对多的联系。尽管网状模型不支持多对多联系，但由于一个多对多联系可以转化为两个一对多联系，所以网状模型可以间接地描述多对多联系。图 1-5 是学生选课系统数据库网状模型。

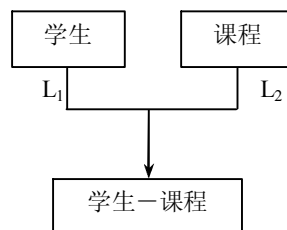


图 1-5 学生/学生-课程/课程网状模型

同层次模型相比，网状模型能更好地描述复杂的现实世界，但网状模型结构比较复杂，到达一个结点

的路径有多条, 用户必须了解系统结构的细节, 对于开发人员的要求也较高。

1.2.4 关系模型

1970 年美国 IBM 公司 San Jose 研究室的高级研究员埃德加·考特提出了数据库的关系模型(Relational Model)。由于他的杰出贡献, 他于 1981 年获得了计算机科学领域的最高奖项——图灵奖。

在我们的社会生活中, 关系无处不在。如数之间的大小关系、人之间的亲属关系、商品流通中的购销关系等。在关系模型中表示实体间联系的方法与非关系模型不同。关系模型中数据的逻辑结构就是一张二维表。如表 1-1 所示的教师档案表就是一个关系模型的例子。

表 1-1 教师档案表

教师编号	教师姓名	所属院系名称	所属专业名称
0973	黄军	信息科学与工程	计算机科学与技术
1302	李立名	信息科学与工程	计算机科学与技术
1704	刘义明	数学与计量经济	应用数学
1870	王国华	数学与计量经济	应用数学

1.2.5 面向对象数据模型

面向对象模型(Object Oriented Model, 简称 OO 模型)是近几年来发展起来的一种新兴的数据模型。该模型是在吸收了以前的各种数据模型优点的基础上, 借鉴了面向对象程序设计方法而建立的一种模型。OO 模型是用面向对象观点来描述现实世界实体(对象)的逻辑组织、对象间限制、联系等的模型。这种模型具有更强的表示现实世界的能力, 是数据模型发展的一个重要方向。目前对于 OO 模型还缺少统一的规范说明, 尚没有一个统一的严格的定义。但在 OO 模型中, 面向对象核心概念构成了面向对象数据模型的基础。OO 模型的基本概念如下:

(1) 对象(Object)与对象标识(OID)。现实世界中的任何实体都可以统一地用对象来表示。每一个对象都有它唯一的标识, 称为对象标识, 对象标识始终保持不变。一个学生是一个对象, 他的姓名、性别、年龄等构成了这个对象的属性, 属性描述的是对象的静态特性。对象的动态特性可以用操作来描述, 对象对某一事件所做出的反应就是操作, 也称为方法(Method)。每一个对象可以认为是其本身的一组属性和它可以执行的一组操作。

(2) 类(Class)。所有具有相同属性和操作集的对象构成一个对象类(简称类)。任何一个对象都是某一对象类的一个实例(instance)。例如学生是一个类, 每个学生如李刚、王磊、刘小红等都是学生类中的对象。他们是这个对象类的具体实例, 具有一些相同的属性如班级、学号等, 但有不同的属性值如属于不同的班级、学号不同等。

(3) 事件。客观世界是由对象构成的, 客观世界中的所有行动都是由对象发出且能够被某些对象感受到, 我们把这样的行动称为事件。在关系数据库应用系统中, 事件分为内部事件和外部事件。系统中对象的数据操作和功能调用命令等都是内部事件, 而鼠标的移动、点击等都是外部事件。

此外还有封装(Encapsulation)、类层次、消息(Message)等概念, 这里不再详细介绍。

1.3 关系数据库系统

尽管数据库领域中存在多种组织数据的方式，但关系数据库是效率最高的一种数据库系统。关系数据库系统（Relational DataBase System，简称 RDBS）采用关系模型作为数据的组织方式，Access 就是基于关系模型的数据库系统。关系数据模型之所以重要，是因为它是用途广泛的关系数据库系统的基础。

1.3.1 关系模型的基本概念

关系模型的主要概念如下：

（1）关系（Relation）

一个关系可以看成是一张二维表，通常将一个没有重复行，重复列的二维表看成一个关系，每个关系都有一个关系名。如表 1-1 所示的教师档案表。

（2）元组（Tuple）

二维表的每一行（记录的值得）在关系中称为元组，Access 中称为记录（Record）。在关系数据库中，一个元组对应表中一个记录。在关系表中不能有完全相同的两个元组。

（3）属性（Attribute）

二维表的每一列在关系中称为属性，Access 中称为字段（Field）。每个属性都有一个属性名，属性值则是各元组属性的取值。在关系数据库中，一个属性对应表中一个字段，属性名对应字段名，属性值对应于各个记录的字段值。在一个表中，不能有两个字段名完全相同。表 1-1 中有 4 列，则有 4 个属性（教师编号，教师姓名，所属院系名称，所属专业名称）。

（4）域（Domain）

属性的取值范围称为域。域作为属性值的集合，其类型与范围由属性的性质及其所表示的意义具体确定。同一属性只能在相同域中取值。如表 1-1 中“所属院系名称”的域是该校所有院系名称的集合。

（5）主码或主关键字（Primary Key）

表中的某个属性或属性组，能够唯一确定一个元组。Access 中的主码称为主键。如表 1-1 中的“教师编号”可以唯一确定一名教师，即是本关系中的主码或主关键字。

（6）关系模式

对关系的描述称为关系模式。一般表示为：关系名（属性 1，属性 2，…，属性 n）。

一个关系模式对应一个关系的结构。例如表 1-1 中的关系可描述为：

教师档案（教师编号，教师姓名，所属院系名称，所属专业名称）

关系模型的主要特点有：

（1）关系中每一数据项不可再分，也就是不允许表中还有表。如表 1-2 所示的模型就不符合关系模型的要求。工资又被分为基本工资、岗位工资和补贴，这相当于大表中又有一张小表。

（2）每一列中的各个数据项具有相同的属性。

（3）每一行中的记录由一个事物的多种属性项构成。

（4）每一行代表一个实体，不允许有相同的记录行。

（5）行与行、列与列的次序可以任意交换，不会改变关系的实际意义。

表 1-2 “表中有表”示例

编号	姓名	出生日期	系别	职称	工资		
					基本工资	岗位工资	补贴
0973	黄军	1965.3	信息科学与工程	副教授	3200	1858	280
1302	李立名	1960.6	信息科学与工程	教授	4400	2620	420

1.3.2 关系模型的组成

关系模型由关系数据结构、关系操作和关系完整性约束三部分组成。

(1) 关系数据结构。关系模型中数据的逻辑结构是一张二维表。在用户看来非常单一，但这种简单的数据结构能表达丰富的语义，可描述出现实世界的实体以及实体间的各种联系。如一个学校可以有一个数据库，在数据库中建立多个表，其中一个表用来存放教师信息，一个表用来存放学生信息，一个表用来存放课程设置信息等。

在关系数据库中，表与表的关系有三种：一对一关系、一对多关系、多对多关系。关联关系分别取决于相关列的定义方式。若表 A 和表 B 之间存在一对一的关系，那么表 A 中的一条记录只能与表 B 中的一条记录匹配，反过来表 B 中的一条记录只能与表 A 中的一条记录匹配；若表 A 和表 B 之间存在一对多的关系，且表 B 为多的一方，那么表 A 中的一条记录与表 B 中的多条记录匹配，反过来表 B 中的一条记录只能与表 A 中的一条记录匹配；若表 A 和表 B 之间存在多对多的关系，那么表 A 中的一条记录能与表 B 中的多条记录匹配，反过来表 B 中的一条记录也能与表 A 中的多条记录匹配。

(2) 关系操作。关系操作采用集合操作方式，即操作的对象和结果都是集合。关系模型中常用的关系操作包括如下两类：

查询操作：选择 (Select)、投影 (Project)、连接 (Join)、除 (Divide)、并 (Union)、交 (Intersection)、差 (Difference) 等。

数据维护操作：增加 (Insert)、删除 (Delete)、修改 (Update) 操作。

(3) 关系完整性约束。关系模型中的完整性是指数据库中数据的正确性和一致性，关系数据模型的操作必须满足关系的完整性约束条件。

关系的完整性约束条件包括实体完整性、参照完整性和用户定义的完整性。其中实体完整性和参照完整性是关系模型必须满足的完整性约束条件，适用于任何关系数据库系统。用户定义的完整性是针对某一具体领域的约束条件，它反映某一具体应用所涉及的数据必须满足的语义要求。

1.3.3 关系运算的基本概念

关系运算的对象是关系，运算结果也为关系。关系的基本运算有两类，一类是传统的集合运算如并、差、交等，另一类是专门的关系运算如选择、投影、连接等。

假设有两个关系 R 和 S，它们具有相同的结构。

(1) 并 (Union)

R 和 S 的并是由属于 R 或属于 S 的元组组成的集合，运算符为 “ $\cup$ ”，记为 $R \cup S$ 。

(2) 差 (Difference)

R 和 S 的差是由属于 R 但不属于 S 的元组组成的集合，运算符为 “ $-$ ”，记为 $R - S$ 。

(3) 交 (Intersection)

R 和 S 的交是由既属于 R 又属于 S 的元组组成的集合，运算符为 “ $\cap$ ”，记为 $R \cap S$ 。

(4) 广义笛卡儿积 (Extended Cartesian Product)

关系 R (假设为 n 列) 和关系 S (假设为 m 列) 的广义笛卡儿积是一个 (n+m) 列元组的集合。每一个元组的前 n 列是来自关系 R 的一个元组，后 m 列是来自关系 S 的一个元组。

(5) 选择 (Selection)

在关系中选择符合某些条件的元组的操作称为选择运算。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。如要在教师档案表 (如表 1-1 所示) 中查询计算机专业的所有教师数据，就可以对教师档案表做选择操作，条件是 “所属专业名称” = “计算机科学与技术”。运算结果如表 1-3 所示。

表 1-3 选择运算举例

教师编号	教师姓名	所属院系名称	所属专业名称
0973	黄军	信息科学与工程	计算机科学与技术
1302	李立名	信息科学与工程	计算机科学与技术

(6) 投影 (Projection)

在关系中选择某些属性列组成新的关系称为投影运算。这是从列的角度进行的运算，相当于对关系进行垂直分解。

如要查询所有教师的姓名和所属院系名称，则可以对教师档案表做投影操作。结果如图 1-6 (a) 所示。进行投影运算之后不仅会取消原关系中的某些列，而且还可能取消某些元组，因为取消了某些属性列后，就可能出现重复行，应取消这些完全相同的行。例如查询教师档案表中有哪些院系，即查询教师档案表在 “所属院系名称” 属性上的投影，结果如图 1-6 (b) 所示。表中原来有 4 个元组，而投影结果取消了重复元组，因此只有两个元组。

教师姓名	所属院系名称
黄军	信息科学与工程
李立名	信息科学与工程
刘义明	数学与计量经济
王国华	数学与计量经济

(a)

所属院系名称
信息科学与工程
数学与计量经济

(b)

图 1-6 投影运算举例

(7) 连接 (Join)

选择和投影运算的操作对象是一个关系，而连接运算需要两个关系作为操作对象。连接运算也是使用比较频繁的一种二元运算，它用来根据一定的关系把两个关系连接成一个关系。连接是成对进行处理的，无论要组合多少表，连接总是在两个表间进行，这些连接的结果称为中间结果。然后中间结果通过使用连接算法连接到另一个表。对于每个连接，查询优化器将确定要使用的适当的连接算法。从原理上讲，连接运算是一种比较特殊的选择运算，即连接是按照两个关系表中相同字段间的一定条件对两个关系表中的记录进行选择而形成新的记录集。最常用的连接运算有等值连接 (Equal Join) 和自然连接 (Natural Join) 两种。

1.3.4 关系数据库管理系统的功能

关系数据库管理系统主要有 4 方面的功能：数据定义、数据处理、数据控制和数据维护。

(1) 数据定义功能

关系数据库管理系统一般均提供数据定义语言 (Data Definition Language, DDL)，可以允许用户定义数据在数据库中存储时所使用的类型 (例如文本或数字类型)，以及各主题之间的数据如何相关。

(2) 数据处理功能

关系数据库管理系统一般均提供数据操纵语言 (Data Manipulation Language, DML)，让用户可以使用多种方法来操作数据。例如只显示用户关心的数据。

(3) 数据控制功能

可以管理工作组中使用、编辑数据的权限，完成数据安全性、完整性及一致性的定义与检查，还可以保证数据库在多个用户间正常使用。

(4) 数据维护功能

包括数据库中初始数据的装载，数据库的转储、重组、性能监控、系统恢复等功能，它们大都由 RDBMS 中的实用程序来完成。

1.3.5 常见的关系数据库管理系统及分类

关系数据库有很多优点，包括有严格的理论基础、提供单一的数据结构、存取路径对用户透明等，因此关系数据库的使用非常普遍。目前，关系数据库管理系统 (RDBMS) 的种类很多，常见的有 Oracle、DB2、Sybase、Informix、Ingres、RDB、SQL Server、Access、FoxPro 等系统。

一个数据库管理系统可定义为关系系统，它至少支持关系数据结构及选择、投影和连接运算，这是对关系数据库系统的最低要求。按照 E.F.Codd 衡量关系系统的准则，可以把关系数据库系统分为如下三类：

(1) 半关系型系统。这类系统大都采用关系作为基本数据结构，仅支持三种关系操作，但不提供完备的数据子语言，数据独立性差。如 FoxBase、FoxPro 就属于这类。

(2) 基本关系型系统。这类系统均采用关系作为基本数据结构，支持所有的关系代数操作，有完备的数据子语言，有一定的数据独立性，并有一定的空值处理能力，有视图功能，它满足 E.F.Codd 衡量关系系统的准则的大部分条件。目前，大多数关系数据库产品均属于此类。如 DB2、Oracle、Sybase 等。

(3) 完全关系型系统。这是一种理想化的系统，这类系统支持关系模型的所有特征。虽然 DB2、Oracle 等系统已经接近这个目标，但尚不属于完全关系型系统。

1.3.6 关系数据库管理系统——Access

Microsoft Access 是 Microsoft Office 组件中重要的组成部分，是目前较为流行的关系数据库管理系统。Access 把数据库引擎的图形用户界面和软件开发工具结合在一起，它具有大型数据库的一些基本功能，支持事务处理功能，具有多用户管理功能，支持数据压缩、备份和恢复功能，能够保证数据的安全性。

Access 不仅是数据库管理系统，而且还是一个功能强大的开发工具，具有良好的二次开

发支持特性，有许多软件开发者把它作为主要的开发工具。与其他的数据库管理系统相比，Access 更加简单易学，一个普通的计算机用户即可掌握并使用它。

1.4 关系数据库设计

在关系数据库应用系统的开发过程中，数据库设计是核心和基础。数据库设计是指针对一个给定的应用环境，构造最优的数据模式，建立数据库及其应用系统，有效存储数据，以满足用户的信息存储和处理要求。数据库设计阶段主要包括需求分析、概念设计、逻辑设计和物理设计，如逻辑设计的主要工作是将 E-R 图转换为关系模式。针对一个具体问题，应该如何构造一个符合实际的恰当的数据模式，即应该构造几个关系，每个关系应该包括哪些属性，各个元组的属性值应符合什么条件等，这些都是应全面考虑的问题。在关系数据库设计时要遵守一定的规则，下面介绍数据库关系完整性设计和数据库规范化设计。

1.4.1 关系的键

1. 候选键（Candidate key）

能唯一标识关系中元组的一个属性或属性集称之为候选键，也称为候选关键字或候选码。例如“学生”关系（见表 1-4）中的学号能唯一标识每一个学生，则属性“学号”是“学生”关系的候选键。在“选课”关系（见表 1-5）中，只有属性的组合“学号+课程号”才能唯一地区分每一条选课记录，则属性集“学号+课程号”是“选课”关系的候选键。

表 1-4 学生表

学号	姓名	院系
0001	蒋民军	信息科学与工程
0002	李兵	信息科学与工程
0003	王小山	信息科学与工程
0004	张强明	信息科学与工程

表 1-5 选课表

学号	课程号	成绩
0001	12	94
0001	13	89
0002	15	97
0003	12	90
0003	14	88

候选键将满足唯一性和最小性的条件。例如，“学生”关系中的每一个学生的学号都是唯一的，“选课”关系中“学号+课程号”的组合也是唯一的。在属性集“学号+课程号”中去掉任一属性，都无法唯一标识选课记录。

2. 主关系键（Primary key）

如果一个关系中有多个候选键，可以从中选择一个作为查询、插入或删除元组的操作变量，被选用的候选键称为主关系键，或简称为主键、主码、关键字等。

例如，假设在“学生”关系中没有重名的学生，则学号和姓名都可以作为“学生”关系的候选键。如果选定“学号”作为数据操作的依据，则“学号”为主关系键；如果选定“姓名”作为数据操作的依据，则“姓名”为主关系键。

3. 主属性与非主属性

主属性：包含在主关系键中的各个属性称为主属性。

非主属性：不包含在任何候选键中的属性称之为非主属性。

在最简单的情况下，一个候选键只包含一个属性，如“学生”关系中的“学号”，“课程”关系中的“课程号”。在最极端的情况下，一个关系的所有属性的组合是关系的候选键，这时称为全码（All-Key）。

4. 外部关系键（Foreign Key）

如果一个关系 R_2 的一个或一组属性 X 不是 R_2 的主码，而是另一个关系 R_1 的主码，则该属性或属性组 X 称为关系 R_2 的外部关系键或外码，并称关系 R_2 为参照关系（Referencing Relation），关系 R_1 为被参照关系（Referenced Relation）。

例如，在“选课”表（见表 1-5）中，“学号”属性与“学生”关系中的主码“学号”相对应，“课程号”与“课程”关系中的主码“课程号”相对应。因此，“学号”和“课程号”属性是“选课”关系的外部关系键，“学生”关系和“课程”关系为被参照关系，“选课”关系为参照关系。

1.4.2 数据库关系完整性设计

关系数据库设计是对数据进行组织化和结构化的过程，核心问题是关系模型的设计。关系模型的完整性规则是对关系的某种约束条件，是指数据库中数据的正确性和一致性。现实世界的实际存在决定了关系必须满足一定的完整性约束条件，这些约束表现在对属性取值范围的限制上。完整性规则就是防止用户使用数据库时，向数据库中插入不符合语义的数据。关系模型中有三类完整性约束：实体完整性、参照完整性和用户定义的完整性。其中实体完整性和参照完整性是关系模型必须满足的完整性约束条件，被称作关系的两个不变性。

1. 实体完整性规则

实体完整性是指基本关系的主属性，即主码的值都不能取空值。

在关系系统中一个关系通常对应一个表，实际存储数据的表称为基本表，而查询结果表、视图表等都不是基本表。实体完整性是针对基本表而言的，指在实际存储数据的基本表中，主属性不能取空值。例如在教师表中，“教师编号”属性为主码，则“教师编号”不能取空值。

一个基本关系对应现实世界中的一个实体集，如教师关系对应教师集合，学生关系对应学生集合。现实世界中实体是可区分的，即每个实体具有唯一性标识。在关系模型中用主码做唯一性标识时，若主码取空值，则说明这个实体无法标识，即不可区分。这显然与现实世界相矛盾，现实世界不可能存在这样的不可标识的实体，基于此引入了实体完整性规则。

实体完整性规则规定基本关系的所有主属性都不能取空值，而不仅是主码整体不能取空值。如学生选课表中，“学号”和“课程号”一起构成主码，则“学号”和“课程号”这两个属性的值均不能为空值，否则就违反了实体完整性规则。

2. 参照完整性规则

现实世界中的实体之间往往存在某种联系，在关系模型中实体及实体间的联系都是用关系来描述的。这样就存在着关系与关系间的引用。

参照完整性规则：是指在输入或删除记录时，为维持表之间已定义的关系而必须遵循的规则。实施了参照完整性后，对表中主关键字字段进行操作时系统会自动地检查主关键字字段，看看该字段是否被添加、修改或删除。如果对主关键字的修改违背了参照完整性的要求，那么系统会自动强制执行参照完整性。如设置主表中没有相关记录就不能将记录添加到相关表、相关表中的记录删除时主表的相关记录随之删除、相关表中的记录更新时主表的相关记录随之更新等参照完整性。

例如,“教师”关系和“院系”关系中主码分别是:教师编号、院系代码,用下划线标识。

教师(教师编号,教师姓名,院系代码,专业名称)

院系(院系代码,院系名称)

这两个关系之间存在属性的引用,即教师关系引用了院系关系的主码“院系代码”。按照参照完整性规则,教师关系中每个元组的“院系代码”属性只能取下面两类值:

- 空值,表示这位教师还未分配到任何一个院系工作。
- 非空值,此时取值必须和“院系”关系中某个元组的“院系代码”值相同,表示这个教师分配到该院系工作。

参照完整性规则规定不能引用不存在的实体。上例中如果“教师”关系中某个教师的“院系代码”取值不与“院系”关系中任何一个元组的院系代码一致,表示这个教师被分配到一个不存在的院系中,这与实际应用环境不相符,显然是错误的。

3. 用户定义的完整性

用户定义的完整性是针对某一具体关系数据库的约束条件,它反映某一具体应用所涉及的数据必须满足的语义要求。任何关系数据库系统都应该支持实体完整性和参照完整性。关系模型应提供定义和检验这类完整性规则的机制,其目的是用统一的方式由系统来处理它们,而不由应用程序来完成这项工作。

例如,在学生成绩表中规定成绩不能超过100;在教师档案表(教师编号,教师姓名,所属院系名称,所属专业名称)中,要求教师姓名的取值不能为空。

1.4.3 数据库规范化设计

在数据库设计中,如何把现实世界表示成合理的数据库模式,一直是人们非常重视的问题。关系数据库的规范化理论就是进行数据库设计时的有力工具。该理论是指导数据库设计的一项非常重要的原则,其基本思想是:每个关系都应该满足一定的规范,才能使关系模式设计合理,达到减少冗余、提高查询效率的目的。关系规范化理论要求在设计关系型数据表时,应该力求满足3个范式内容:

1. 第一范式(1NF)

要求表结构的关系中每一个数据项必须是不可再分的。

2. 第二范式(2NF)

在满足第一范式的关系中,要求所有非主属性都完全函数性地依赖于主关系键。

3. 第三范式(3NF)

对于满足第二范式的关系,要求每一个非主属性都不传递依赖于主关系键。传递依赖是指某些数据项间接依赖于主关系键。如在“课程设置”中,“职称”属于“任课教师”,主码“课程代码”不直接决定非主属性“职称”,“职称”是通过“任课教师”传递依赖于“课程代码”的,则此关系不满足第三范式,在某些情况下,会存在插入异常、删除异常和数据冗余等现象。为将此关系处理成满足第三范式的数据表,可以将其细分成“课程设置表”和“任课教师名单”两张数据表以满足第三范式。

对于数据库规范化设计的要求是应该保证所有数据表都能满足第二范式,力求绝大多数数据表满足第三范式。除以上介绍的三种范式外,还有BCNF(Boyce Codd Normal Form)、第四范式、第五范式。一个低一级范式的关系模式,通过模式分解可以规范化为若干个高一级范式的关系模式的集合。

1.4.4 Access 数据库应用系统设计实例

按照规范化理论和完整性规则设计出能够正确反映现实应用的数据模型后,还要进行系统功能的设计。对于系统功能设计应遵循自顶向下、逐步求精的原则,将系统必备的功能分解为若干相互独立又相互依存的模块,每一模块采用不同的技术,解决不同的问题,从而将问题局部化,这是数据库设计中的分步设计法。下面以一个学生成绩管理系统为例,简单介绍数据库系统开发的方法。

1. 需求分析

这是数据库应用系统开发的第一步。首先要详细调查要处理的对象,明确用户的各种要求,在此基础上确定数据库中需要存储哪些数据及系统需要具备哪些功能等。需求分析阶段可使用的工具有数据流图(DFD)、数据字典(DD)、判断树和判断表等。设计人员必须不断深入地与用户交流,才能逐步确定用户的实际需求,以确定设计方案。对于学生成绩管理系统来说,进行需求分析后,得到以下结果:

(1) 用户需要完成数据的录入。学校开设新课程、新生入学、增加新院系、增加新教师、重新选课、统计期末统计成绩时都需要进行数据录入,并提交数据库保存。因此系统要包括以下数据表:专业表、教师表、学生表、课程设置表、学生选课表等。

(2) 完成数据的修改。当学生、教师、课程等情况发生变化或数据录入错误时,用户要进行数据的修改,以保证数据表中数据的正确性。

(3) 实现信息查询。包括学生成绩查询、学生信息查询、学生选课查询、课程查询等。

2. 应用系统的数据库设计

这是在需求分析的基础上进行的。首先要弄清需要存储哪些数据,确定需要几个数据表,每一个表中包括几个字段等,然后在 Access 中建立数据表。这一过程要严格遵循关系数据库完整性和规范化设计要求。学生成绩管理系统要创建 6 个数据表:

(1) 专业表(专业编号,专业名称)

主键:专业编号

外键:无

(2) 学生表(学号,姓名,性别,民族,出生日期,专业编号,地址,团员否,照片)

主键:学号

外键:专业编号

(3) 选修表(学号,课程号,成绩)

主键:学号,课程号

外键:学号,课程号

(4) 课程表(课程号,课程名,总学分,总学时,课程性质,考核方式)

主键:课程号

外键:无

(5) 教学表(课程号,教师号)

主键:课程号,教师号

外键:课程号,教师号

(6) 教师表(教师号,教师姓名,职称,学历,工资)

主键:教师号

外键:无

本章小结

本章概述了数据库系统的基本概念，并通过对数据库技术发展的情况介绍，概括了数据库技术发展的新方向。

数据模型是数据库系统的核心和基础。本章介绍了五种数据模型：E-R 模型、层次模型、网状模型、关系模型和面向对象数据模型。之后详细介绍了关系数据库及其设计的有关理论。

最后通过一个 Access 数据库应用系统设计（学生成绩管理系统）实例，简要介绍关系数据库设计理论，包括数据库关系完整性设计和数据库规范化设计理论及数据库系统开发的基本方法。

习题 1

一、选择题

- 要求主表中没有相关记录时就不能将记录添加到相关表中，则应该在表关系中设置（ ）。
 - 参照完整性
 - 有效性规则
 - 输入掩码
 - 级联更新相关字段
- 在超市营业过程中，每个时段要安排一个班组上岗值班，每个收款口要配备两名收款员配合工作，共同使用一套收款设备为顾客服务。在数据库中，实体之间属于一对一关系的是（ ）。
 - “顾客”与“收款口”的关系
 - “收款口”与“收款员”的关系
 - “班组”与“收款员”的关系
 - “收款口”与“设备”的关系
- 在人事管理数据库中工资与职工之间存在的关系是（ ）。
 - 一对一
 - 一对多
 - 多对一
 - 多对多
- 如果一个关系中，存在多个属性（或属性组）都能用来唯一标识该关系的元组，且其任何一个子集都不具有这一特性。这些属性（或属性组）都被称为关系的（ ）。
 - 连接码
 - 主码
 - 外码
 - 候选码
- 在数据库设计中，将 E-R 图转换为关系模式的过程属于（ ）。
 - 需求分析阶段
 - 逻辑设计阶段
 - 概念设计阶段
 - 物理设计阶段
- 设有如下关系表：

R		
A	B	C
1	2	3
4	5	6

S		
A	B	C
7	8	9

T		
A	B	C
1	2	3
4	5	6
7	8	9

则下列操作正确的是 ()。

- A. $T=R/S$ B. $T=R \times S$ C. $T=R \cap S$ D. $T=R \cup S$
7. 在关系运算中, 选择运算的含义是 ()。
- A. 在基本表中选择满足条件的记录组成一个新的关系
B. 在基本表中选择需要的字段 (属性) 组成一个新的关系
C. 在基本表中选择满足条件的记录和属性组成一个新的关系
D. 上述说法均是正确的
8. Access 中的参照完整性规则不包括 ()。
- A. 删除规则 B. 插入规则 C. 查询规则 D. 更新规则
9. 数据库管理员 (DBA) 是数据库系统的一个重要组成部分, 有很多职责。以下选项不属于 DBA 职责的是 ()。
- A. 定义数据的存储结构和存取策略
B. 定义数据库的结构
C. 定期对数据库进行重组和重构
D. 设计和编写应用系统的程序模块
10. 假设数据库中表 A 与表 B 建立了 “一对多” 关系, 表 A 为 “多” 的一方, 则下述说法正确的是 ()。
- A. 表 B 中的一个字段能与表 A 中的多个字段匹配
B. 表 B 中的一个记录能与表 A 中的多个记录匹配
C. 表 A 中的一个记录能与表 B 中的多个记录匹配
D. 表 A 中的一个字段能与表 B 中的多个字段匹配
11. 数据类型是 ()。
- A. 字段的另一种说法
B. 决定字段能包含哪类数据的设置
C. 一类数据库应用程序
D. 一类用来描述 Access 表向导允许从中选择的字段名称
12. 任何关系数据库系统都应该支持: ①实体完整性约束; ②参照完整性约束; ③用户自定义完整性约束 ()。
- A. 仅① B. 仅①、② C. 都是 D. 都不是
13. 在关系运算中, 投影运算的含义是 ()。
- A. 在基本表中选择满足条件的记录组成一个新的关系
B. 在基本表中选择需要的字段 (属性) 组成一个新的关系
C. 在基本表中选择满足条件的记录和属性组成一个新的关系
D. 上述说法均是正确的
14. 下列关于二维表的说法错误的是 ()。
- A. 二维表中的列称为属性 B. 属性值的取值范围称为值域
C. 二维表中的行称为元组 D. 属性的集合称为关系
15. 在三级模式之间引入两层映像, 其主要功能之一是 ()。
- A. 使数据与程序具有较高的独立性
B. 使系统具有较高的通道能力
C. 保持数据与程序的一致性

- D. 提高存储空间的利用率
16. Access 中表和数据库的关系是 ()。
- A. 一个表可以包含多个数据库 B. 一个数据库只能包含一个表
- C. 一个数据库可以包含多个表 D. 一个表只能包含一个数据库
17. 数据库 (DB)、数据库系统 (DBS)、数据库管理系统 (DBMS) 之间的关系是 ()。
- A. DB 包括 DBS 和 DBMS B. DBMS 包括 DB 和 DBS
- C. DBS 包括 DB 和 DBMS D. 以上三个都不是
18. 常见的数据模型有三种, 它们是 ()。
- A. 层次、关系和语义 B. 环状、层次和星形
- C. 字段名、字段类型和记录 D. 层次、关系和网状
19. 在 Access 数据库中, 表就是 ()。
- A. 数据库 B. 记录 C. 字段 D. 关系
20. 在软件开发中, 需求分析阶段不可以使用的工具是 ()。
- A. 数据字典 B. 数据流图 C. 判断树 D. 程序流程图

二、填空题

- 在关系模型中, 二维表的行称为_____。
- 在关系代数运算中, _____操作会消去关系中的某些列并删去重复的元组。
- 不改变关系表中的属性个数但能减少元组个数的是_____。
- 关系代数是关系操作语言的一种传统表示方式, 它以集合代数为基础, 它的运算对象和运算结果均为_____。
- 数据库系统其内部分为三级模式, 即外模式、_____和内模式。其中, _____是用户的数据视图, 也就是用户所见到的数据模式。
- 数据库系统的核心是_____。
- 数据的逻辑结构在计算机存储空间中的存放形式称为数据的_____。
- 关系模型的完整性规则是对关系的某种约束条件, 包括实体完整性、_____和自定义完整性。
- 若 $D1=\{a1,a2,a3\}$, $D2=\{b1,b2,b3\}$, 则 $D1 \times D2$ 集合中共有_____个元组。

三、简答题

- 数据库技术的发展经历了哪几代? 请简述每一代数据库系统的特点。
- 数据库技术有哪些新的进展?
- 简述数据、数据库、数据库管理系统、数据库系统的概念。
- 简述数据模型的概念和数据模型的两个要素。
- 简述关系模型的组成。
- 关系数据库系统有哪些主要功能?
- 解释以下术语: 实体完整性、参照完整性、用户定义的完整性。
- 简述第一范式、第二范式、第三范式的概念。