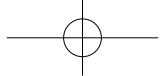
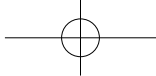


目 录

第1章 初识数据库	1
1.1 数据管理技术的发展	1
1.2 数据库管理系统	5
1.3 数据库系统	6
1.4 数据模型	9
1.5 数据库应用系统体系结构	13
1.6 数据库技术新发展	15
第2章 关系数据库原理	19
2.1 关系数据库概述	19
2.2 关系代数	20
2.3 关系模式的分解	27
2.4 关系模型的完整性约束	33
第3章 安装SQL Server 2008	37
3.1 SQL Server简介	37
3.2 SQL Server 2008的安装	38
第4章 设计和管理数据库	54
4.1 设计数据库	55
4.2 SQL Server数据库概述	60
4.3 创建数据库	64
4.4 修改数据库	74
4.5 删除数据库	81
4.6 查看数据库信息	83
4.7 收缩数据库	88
4.8 数据库快照	91
第5章 创建与管理学生选课数据库表	95
5.1 SQL Server表概述	95
5.2 数据类型	98
5.3 表的创建与维护	103
5.4 表中数据的维护	112
5.5 数据完整性	117



第6章 关系数据库标准语言SQL	134
6.1 单表查询	135
6.2 多表连接查询	140
6.3 嵌套查询	142
第7章 索引与视图	153
7.1 索引概述	153
7.2 索引的操作	155
7.3 视图概述	161
7.4 视图的操作	163
7.5 视图的应用	169
第8章 T-SQL 程序设计	174
8.1 T-SQL 编程基础	174
8.2 程序控制流语句	179
8.3 函数	185
8.4 游标	193
第9章 存储过程与触发器的应用	200
9.1 存储过程	200
9.2 触发器	208
9.3 批处理和事务	218
第10章 数据库备份与恢复	223
10.1 数据库的备份	223
10.2 数据库的恢复	239
10.3 分离与附加数据库	247
10.4 导入导出数据	250
10.5 复制数据库	257
第11章 数据库安全管理与维护	263
11.1 数据库安全性概述	263
11.2 管理SQL Server服务器安全性	266
11.3 管理角色	274
11.4 管理权限	279
第12章 JSP访问SQL Server 2008 数据库	284
12.1 JSP开发环境搭建	285
12.2 JDBC 技术详解	290
12.3 JSP 实现“学生选课”数据库应用实例	294
参考文献	306



第1章 初识数据库

【学习情境】

在计算机发展初期，计算机主要用于科学计算。后来随着人类社会进入信息化时代，人们需要面临和处理的信息量急剧膨胀，迫切需要利用计算机完成对大量数据的组织、管理、存储和维护，数据库技术应运而生。数据库有很多种类型，从最简单的存储各种数据的表格到能够进行海量数据存储的大型数据库系统，都在各个方面得到了广泛的应用。

数据库系统是指在计算机系统中引进数据库技术的系统，在数据库系统中，通过数据库管理系统对数据进行统一管理。有了数据库管理系统，我们可以做到：在图书馆浩如烟海的图书中，读者可以轻松地找到想要的图书及其存放位置；企业可以通过搜集用户的需求信息，有针对性地开发新的产品满足用户的需求，同时减少旧产品的库存量。通过对信息资源的开发利用，可以使各种社会资源得到最大限度地节约和合理运用。

为了能开发出适用的数据库应用系统，就需要掌握数据库原理，熟悉一门数据库语言。SQL Server 是目前使用较为广泛的大型数据库管理系统，本书以SQL Server 2008为背景，介绍数据库的基本操作和数据库应用系统开发的方法。

【知识技能】

1.1 数据管理技术的发展

1.1.1 数据与数据处理

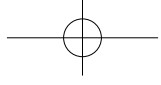
1.数据（Data）

数据是客观事物的反映和记录，是用于记录事物情况的物理符号。为了描述客观事物而用到的数字、字符以及所有能输入计算机中并能被计算机处理的符号都可以称作数据。数据不等同于数字，它包括两大类，即数值型数据和非数值型数据。例如：张三同学身高1.8米，这里的“张三”“1.8”就是数据，其中“张三”为字符型数据，“1.8”是数值型数据。此外，在计算机中，所有能被计算机存储并处理的数字、字符、图形和声音等统称为数据，如张三同学的照片。

2.信息（Information）

信息是有意义的信息，是经过加工处理并对人类社会实践和生产活动产生决策影响的数据。

数据与信息既有区别，又有联系。数据表示信息，但并非任何数据都能表示信息，



信息只是加工处理后的数据，是数据所表达的内容。另外，信息不随表示它的数据形式的改变而改变，它是反映客观现实世界的知识，而数据则具有任意性，用不同的数据形式可以表示同样的信息。例如，武汉今天的天气为28℃，这是一条信息，而“武汉”“28”“℃”则是数据。

3. 数据处理

数据处理是指将数据转换成信息的过程。数据处理的目的是从大量的数据中，根据数据自身的规律及其相互联系，通过分析、归纳、推理等科学方法，利用计算机技术、数据库管理等技术手段，提取有效的信息资源，为进一步分析、管理、决策提供依据。例如，学生各门课成绩为原始数据，经过计算得出平均成绩和总成绩等信息，计算处理的过程就是数据处理。数据处理也称信息处理。

1.1.2 数据描述

在实际应用中，任何数据的处理都是在反映客观事物属性和运行状态以及满足用户需求的基础上进行的。所谓数据描述，就是以“数据”符号的形式，从满足用户需求出发，对客观事物属性和运动状态进行描述。数据的描述既要符合客观现实，又要适应数据库原理与结构。进一步说，由于计算机不能够直接处理现实世界中的具体事物，所以人们必须将客观存在的具体事物进行有效的描述与刻画，转换成计算机能够处理的数据，这一转换过程可分为三个数据范畴：现实世界、信息世界和计算机世界。从客观现实到计算机的描述，数据转换过程如图1-1所示。

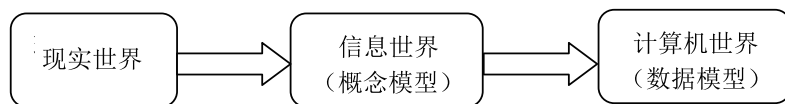


图1-1 数据的转换过程

1. 现实世界

现实世界是指客观存在的事物及其相互间的联系。在现实世界中，人们可以通过事物不同的属性和运动状态加以区别，描述事物的性质和运动规律。事物可以是个体的特殊事物，也可以是集体的共同事物；事物可以是具体的实物，也可以是抽象的概念。

2. 信息世界

信息世界是人们对客观存在的事物及其相互间的联系的反映。人们将对客观事物的反映通过符号记录下来，事实上是对现实世界的一种抽象描述。

3. 计算机世界

计算机世界是信息世界的数字化。客观存在的事物及其相互间的联系的反映，这里都可用数据模型来表示。也就是说，计算机世界的的数据模型将信息世界的概念模型进一步抽象，形成便于计算机处理的表现形式。数据模型的具体内容详见1.4节。

1.1.3 数据管理技术

数据处理的一个重要方面就是如何对数据进行管理。数据管理技术是指对数据进行分类、组织、编码、存储、检索和维护的技术。计算机数据管理技术主要经历了人工管理、

文件管理、数据库管理三个阶段。

1.人工管理阶段

在20世纪50年代中期以前，计算机主要用于科学计算。由于受计算机软、硬件技术发展水平的限制，这个阶段的硬件只有纸带、卡片和磁带，没有磁盘等可直接存取的设备；在软件方面，没有操作系统，没有数据管理的软件。数据处理方式是批处理。

在这个阶段数据管理任务，包括存储结构、存取方法、输入输出方式等都是针对每个具体应用，由编程人员单独设计解决的。数据被包含在程序之中，一组数据对应一个应用程序，一个程序中的数据无法被其他程序使用，因此，程序之间有着大量的不必要的重复、冗余数据，数据也无法被长期保存，数据存储结构一旦改变，则必须修改相应程序。

人工管理数据具有如下特点：

- （1）数据无法保存。
- （2）数据和应用程序一一对应。
- （3）数据无法共享。
- （4）数据不具有独立性。

下面以某高校的信息管理为例，应用程序和数据之间的关系如图1-2所示。学工部门需要处理的是学生的基本情况信息，包括学生的学号、姓名、性别、出生年月、电话等个人信息；教务部门需要处理的是学生的选课信息，包括学生的学号、姓名、性别、选修课程、成绩等个人信息，院系部门需要处理的是学生的专业信息，包括学生的学号、姓名、性别、所在班级、院系，专业编号、专业名称、所属学院等个人信息。从以上信息中分析出各部门都有很多重复的信息，由于在人工管理阶段，应用程序和数据是一一对应的，数据无法共享，因此造成了大量的冗余。

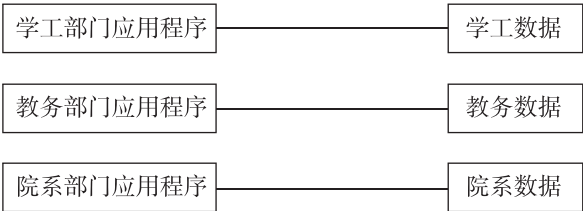


图1-2 人工管理阶段对学生数据进行管理

2.文件管理阶段

20世纪50年代后期到60年代中期，计算机应用领域拓宽，不仅用于科学计算，还大量用于数据管理。硬件方面出现了磁盘等可直接存储设备，软件方面出现了操作系统，可以利用文件存放大量数据，由专门的软件即文件系统对数据进行管理，这样数据可以长期保存。

由专门的软件即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，利用“按文件名访问，按记录进行存取”的管理技术，可以对文件中的数据进行修改、插入和删除操作。文件系统实现了记录内的结构化，但就文件整体而言是无结构的。程序和数据之间由文件系统提供的存取方法进行转换，使应用程序与数据之间有了一定的独立性。程序员可以不必过多地考虑物理细节，而将精力集中于算法。而且数据在存储上的改变不一定反映在程序上，这大大节省了维护程序的工作量。



文件系统使计算机在数据管理方面有了长足的进步。目前，文件系统仍是一般高级语言普遍采用的数据管理方式。然而当数据量增加、使用数据的用户越来越多时，文件系统便无法适应更有效地使用数据的需要了，其问题主要表现在3个方面：

(1) 数据的冗余度大、难以扩充。由于数据文件是根据应用程序的需要而建立的，当不同的应用程序所需要使用的数据有许多部分相同时也必须建立各自的文件，即数据不能共享，造成大量重复。这样不仅浪费存储空间，而且使数据修改变得非常困难，容易产生数据不一致，即同样的数据在不同的文件中所存储的数值不同，造成矛盾。例如图1-2所示的高校信息管理系统中，学工部门、教务部门、院系部门中都需要全校学生的学号、姓名、性别等信息，这些数据在学工数据、教务数据、院系数据中重复出现造成数据冗余。

(2) 数据独立性差。数据独立性是指应用程序和数据之间的独立性，即应用程序不应随着数据的逻辑结构或物理存储结构的改变而改变（分别称为逻辑独立性和物理独立性）。文件按名存储在文件系统中，数据文件之间是孤立的，不能反映现实世界事物之间的相互联系，使数据间的对外联系无法表达。因此，数据独立性差，不利于系统扩充、移植。

(3) 缺乏对数据的统一控制管理。在同一个应用项目中的各个数据文件没有统一的管理机构，数据完整性和安全性很难得到保证。数据的保护等均交给应用程序去解决，使得应用程序的编制相当烦琐。如何能较好地解决文件系统存在的问题，正是数据库系统的目标。

在文件管理阶段，高校信息管理中应用程序和数据文件之间的关系如图1-3所示。

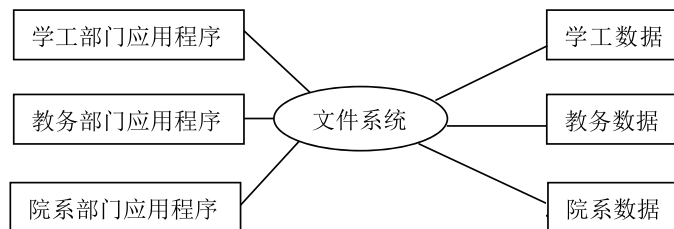


图1-3 用文件系统对学生数据进行管理

3. 数据库管理阶段

20世纪60年代后期以来，用计算机管理的数据量大增，同时各种应用、语言相互覆盖地共享数据的要求越来越强烈，处理方式上实时处理要求也更多。这时文件系统已不能满足应用的需求，于是为了多用户、多应用共享数据的需求，数据库技术应运而生，出现统一管理数据的软件系统：数据库系统。

用数据库系统管理数据的基本思想是：将整个系统涉及的所有数据按一定的结构存放在数据库中，由数据库管理系统负责统一的数据管理和数据控制，用户或应用程序通过数据库管理系统操作数据库，存取其中的数据。

例如，用数据库系统管理学生数据。包括学工部门、院系部门、教务部门在内的学校各个部门都需要处理的学生数据。其中包括了所有学生的学号、姓名、班级、联系方式、选修的课程号、课程名、成绩、专业编号、专业名称等。所有这些数据组成一个学生数

数据库，通常包括四张表：学生表、课程表和选课表、院系表。其中：学生表存放所有学生的个人信息，如学号、姓名、性别等；课程表存放课程的基本信息，如课程编号、课程名称、学分等，选课表存放学生选课信息，如学号、选修的课程号和成绩；院系表存放院系编号、专业编号、专业名称等信息。该学生数据库在数据库管理系统的统一管理和统一控制下供整个系统中所有部门及其应用程序共享。它们之间的关系如图1-4所示。

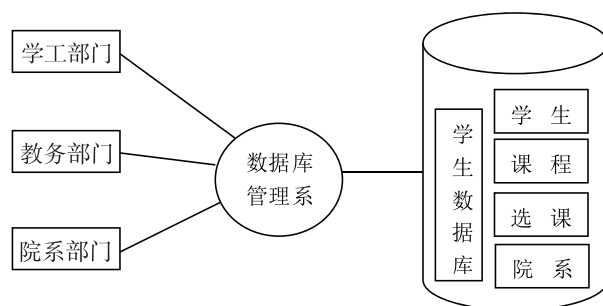


图1-4 用数据库系统对学生数据进行管理

在数据库系统中，由数据库管理系统（Database Management Systems，DBMS）的系统软件对数据进行统一的控制和管理，从而有效地减少了数据冗余，实现了数据共享，解决了数据独立性问题，并提供统一的安全性、完整性和并发控制功能。

数据库系统与文件系统的区别是：数据库中数据的存储是按同一结构进行的，不同的应用程序都可直接操作使用这些数据，应用程序与数据间保持高度的独立性；数据库系统提供一套有效的管理手段，保持数据的完整性、一致性和安全性，使数据具有充分的共享性；数据库系统还为用户管理、控制数据的操作，提供了功能强大的操作命令，使用户直接使用命令或将命令嵌入应用程序中，都简单方便地实现数据库的管理、控制操作。

和文件系统相比，数据库系统主要有以下特点：

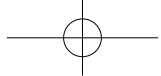
- （1）数据的结构化。
- （2）数据共享性高、冗余度低、易扩充。
- （3）数据由DBMS统一管理和控制。
- （4）数据独立性高。

综上所述，数据库（Database，DB）是长期存储在计算机内，可供不同用户共享的、按一定结构组织在一起的相关数据的集合，它具有最低的数据冗余度和较高的数据独立性，由数据库管理系统负责统一管理和统一控制，可保证数据的安全性和完整性，并在多用户同时使用数据库时进行并发控制，在发生故障时对数据库进行恢复。

1.2 数据库管理系统

为了科学地组织和存储数据，以及高效地获取和维护数据，需要一个专门的系统软件对数据库中的大量数据进行管理，这就是数据库管理系统。

数据库管理系统是数据库系统的核心，是为数据库的建立、使用和维护而配置的软件。它建立在操作系统的基础上，是位于操作系统与用户之间的一层数据管理软件，负责



对数据库进行统一的管理和控制。用户发出的或应用程序中的各种操作数据库中数据的命令，都要通过数据库管理系统来执行。数据库管理系统还承担着数据库的维护工作，能够按照数据库管理员所规定的要求，保证数据库的安全性和完整性。

数据库管理系统主要有如下几方面的具体功能：

1. 数据定义功能

数据定义包括定义构成数据库结构的外模式、模式和内模式，定义各个外模式与模式之间的映射，定义模式与内模式之间的映射，定义有关的约束条件（例如，为保证数据库中数据具有正确语义而定义的完整性规则，为保证数据库安全而定义的用户口令和存取权限等）。

2. 数据操纵功能

数据操纵包括对数据库数据的检索、插入、修改和删除等基本操作。

3. 数据库运行管理

对数据库的运行进行管理是DBMS运行时的核心部分，包括对数据库进行并发控制、安全性检查、完整性约束条件的检查和执行、数据库的内部维护（如索引、数据字典的自动维护）等。所有访问数据库的操作都要在这些控制程序的统一管理下进行，以保证数据的安全性、完整性、一致性以及多用户对数据库的并发使用。

4. 数据控制功能

数据库中需要存放多种数据，如数据字典、用户数据、存取路径等，DBMS负责分门别类地组织、存储和管理这些数据，确定以何种文件结构和存取方式物理地组织这些数据，如何实现数据之间的联系，以便提高存储空间利用率以及提高随机查找、顺序查找、增加、删除、修改等操作的时间效率。

5. 数据通信功能

DBMS需要提供与其他软件系统进行通信的功能。例如，提供与其他DBMS或文件系统的接口，从而能够将数据转换为另一个DBMS或文件系统能够接受的格式，或者接收其他DBMS或文件系统的数据库。

1.3 数据库系统

1.3.1 数据库系统的组成

数据库系统（Database System，DBS）泛指引入数据库后的系统。通常由软件、数据库和数据管理员（Database Administrator，DBA）组成。其软件主要包括操作系统（Operating System，OS）、各种宿主语言、实用程序以及数据库管理系统；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。数据库系统是一个有机整体，其在整个计算机系统中的地位如图1-5所示。

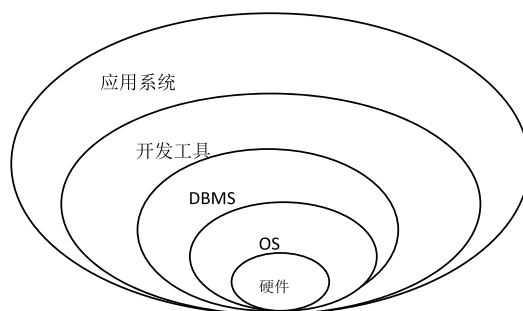


图1-5 数据库系统在计算机系统中的地位

1.3.2 数据库的三级模式结构

为了有效地组织、管理数据，提高数据的逻辑独立性和物理独立性，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，包括外模式、概念模式、内模式，其结构如图1-6所示。

美国国家标准协会（American National Standards Institute, ANSI）的数据库管理系统研究小组于1978年提出了标准化的建议，将数据库结构分为3级：面向用户或应用程序员的用户级、面向建立和维护数据库人员的概念级、面向系统程序员的物理级。用户级对应外模式，概念级对应模式，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。所谓视图，就是指观察、认识和理解数据的范围、角度和方法，是数据库在用户“眼中”的反映，很显然，不同层次（级别）用户所“看到”的数据库是不同的。

1. 模式

模式也称逻辑模式，对应于概念级。它是由数据库设计者综合所有用户的数据，按照统一的观点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（全局视图）。它是由数据库管理系统提供的数据库模式描述语言（Data Description Language, DDL）来描述、定义的。

2. 外模式

外模式又称子模式或用户模式，对应于用户级。它是某个或某几个用户所看到的数据库的数据视图，是与某一应用有关的数据的逻辑表示。

外模式是从模式导出的一个子集，包含模式中允许特定用户使用的那部分数据。用户可以通过外模式描述语言来描述、定义对应于用户的数据记录（外模式），也可以利用数据操纵语言（Data Manipulation Language, DML）对这些数据进行记录。外模式反映了数据库的用户观。

3. 内模式

内模式又称存储模式，对应于物理级，它是数据库中全体数据的内部表示或底层描述，是数据库最低一级的逻辑描述，它描述了数据在存储介质上的存储方式和物理结构，对应实际存储在外存储介质上的数据库。内模式由内模式描述语言来描述、定义，它是数据库的存储观。

在一个数据库系统中，只有唯一的数据库，因而作为定义、描述数据库存储结构的内



模式和定义、描述数据库逻辑结构的模式，也是唯一的，但建立在数据库系统之上的应用则是非常广泛、多样的，所以对应的外模式不是唯一的，也不可能唯一。

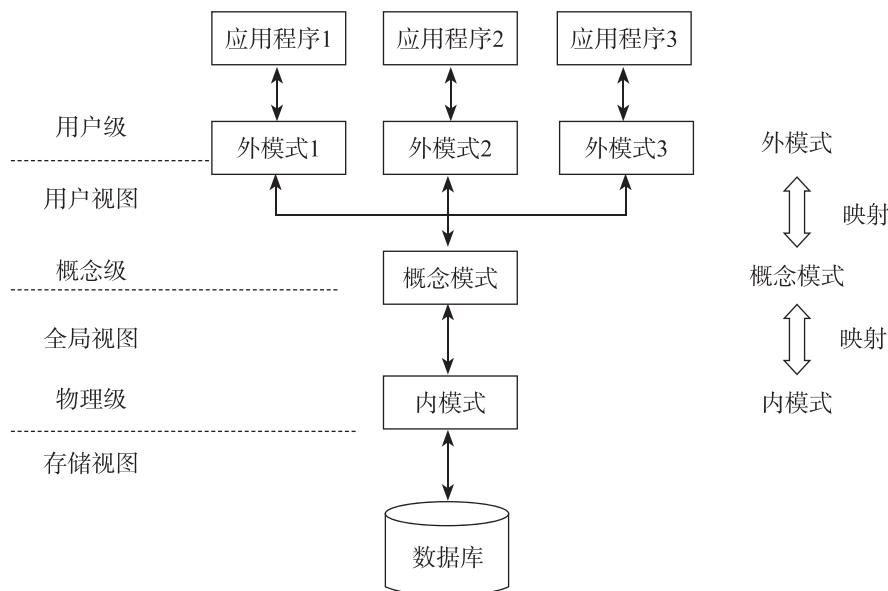


图1-6 数据库的三级模式结构

4.数据库系统三级模式间的两级映射

为了能够在内部实现数据库的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层映像。

(1) 外模式/模式映像。对应于同一个模式可以有任意多个外模式。对于每一个外模式，数据库系统都有一个外模式/模式映像，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/模式映像做相应的改变，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，从而应用程序可以不必修改，保证了数据与程序的逻辑独立性。

(2) 模式/内模式映像。数据库中只有一个模式，也只有一个内模式，所以模式/内模式映像是最唯一的，它定义了数据库的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/内模式映像做相应改变，可以使模式保持不变，从而应用程序也不必修改。保证了数据与程序的物理独立性。

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。因此，涉及数据库模式结构时应首先确定数据库的逻辑结构。

1.3.3 数据库系统的特点

1.数据结构化

数据库系统实现了整体数据的结构化，这是数据库的最主要的特征之一。这里所说的“整体”结构化，是指在数据库中的数据不再仅针对某个应用，而是面向全组织；不仅数据内部是结构化，而且整体式结构化，数据之间有联系。

2.数据的共享、冗余度低

数据共享是指多个用户可以同时存取数据而不相互影响，数据共享包括以下3个方面：所有用户可以同时存取数据；数据库不仅可以为当前的用户服务，也可以为将来的新用户服务；可以使用多种语言完成与数据库的接口。

数据冗余就是数据重复，数据冗余既浪费存储空间，又容易产生数据的不一致。在非数据库系统中，由于每个应用程序都有自己的数据文件，所以数据存在着大量的重复。

数据库从全局观念来组织和存储数据，数据已经根据特定的数据模型结构化，在数据库中用户的逻辑数据文件和具体的物理数据文件不必一一对应，从而有效地节省了存储资源，减少了数据冗余，增强了数据的一致性。

3.数据独立性高

所谓数据独立是指数据与应用程序之间的彼此独立，它们之间不存在相互依赖的关系。

在数据库系统中，数据库管理系统通过映像，实现了应用程序对数据的逻辑结构与物理存储结构之间较高的独立性。数据库的数据独立包括以下两个方面。

（1）物理数据独立：数据的存储格式和组织方法改变时，不影响数据库的逻辑结构，从而不影响应用程序。

（2）逻辑数据独立：数据库逻辑结构的变化（如数据定义的修改，数据间联系的变更等）不影响用户的应用程序。

4.数据由DBMS统一管理和控制

数据库的共享是并发的（concurrency）共享，即多个用户可以同时存取数据库中的数据，甚至可以同时存取数据库中的同一个数据。

DBMS必须提供以下几方面的数据控制功能：

- （1）数据的安全性保护（security）。
- （2）数据的完整性检查（integrity）。
- （3）数据库的并发访问控制（concurrency）。
- （4）数据库的故障恢复（recovery）。

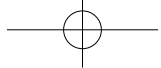
1.4 数据模型

模型是现实世界特征的模拟和抽象。在数据库技术中，用数据模型的概念描述数据库的结构和语义，是对现实世界的数据库抽象。

数据模型是研究数据库技术的核心和基础。数据库技术中研究的数据模型分为两个层面：一层是面向用户的，称为概念模型；另一层是面向计算机系统的，称为结构模型。

1.4.1 概念模型

概念模型是现实世界到机器世界的一个中间层次。现实世界的事物反映到人的脑子中，人们把这些事物抽象为一种既不依赖于具体的计算机系统又不为某一DBMS支持的概念模型，然后再把概念模型转换为计算机上某一DBMS支持的数据模型。



概念模型的表示方法最常用的是P.P.Chen于1976年提出的“实体—联系图方法（Entity-Relationship Approach）”，简称“E-R模型”。E-R实体联系图是直观表示概念模型的工具，其中包含了实体、联系、属性三个成分，联系的方法为一对一（1:1）、一对多（1:n）、多对多（m:n）三种方式。

1. 概念模型的主要概念

（1）实体：客观存在并相互区别的事物及其事物之间的联系。例如，一个学生、一门课程、学生的一次选课等，都是实体。

（2）属性：实体所具有的某一特性。例如，学生的学号、姓名、性别、出生日期等。

（3）主码：也称主关键字，它能够唯一标识一个实体。码可以是一个属性或者属性组。例如，学号是学生实体的主码。主码必须满足以下两个条件。

①主码不能为空。

②主码不能有重复值。

（4）域：属性的取值范围。例如，年龄的域为大于0小于100的整数，性别的域为（男，女）。

（5）实体型：用实体名及其属性名集合来描述同类实体，称为实体型。例如，学生（学号，姓名，性别，出生日期，入学成绩，专业号，电话号码）就是一个实体型。

（6）实体集：同型实体的集合称为实体集。例如，全体学生就是一个实体集。

（7）联系：实体间的联系是指一个实体集中可能出现的每一个实体与另一实体集中多少个具体实体存在联系。

实体之间有各种各样的联系，归纳起来有以下3种类型。

①一对一联系（1:1）。如果对于实体集A中的每一个实体，实体集B中有且只有一个实体与之联系，反之亦然，则称实体集A与实体集B具有一对一联系。例如：“班级”实体与“班长”实体之间属于1:1联系，如图1-7所示。



图1-7 1:1联系



图1-8 1:n联系



图1-9 m:n联系

②一对多联系（1:n）。如果对于实体集A中的每一个实体，实体集B中有多个实体与之联系，反之，对于实体集B中的每一个实体，实体集A中至多只有一个实体与之联系，则称实体集A与实体集B有一对多的联系。例如：“学校”实体与“教师”实体之间属于1:n联系，如图1-8所示。

③多对多联系（m:n）。如果对于实体集A中的每一个实体，实体集B中有多个实体与之联系，而对于实体集B中的每一个实体，实体集A中也有多个实体与之联系，则称实

体集A与实体集B之间有多对多的联系。例如：“教师”实体与“课程”实体之间属于m:n联系。如图1-9所示。

2.E-R图的表示方法（表1-1）

表1-1 E-R图的表示方法

对象类型	表示方法	E-R图表示图示
实体型	用矩形表示，矩形框内写明实体名称	实体名
属性	用椭圆形表示，并用无向线段将其与相应的实体连接起来。	属性名
联系	用菱形表示，菱形框内写明联系名，并用无向线段分别与有关实体连接起来，同时在无向线段旁标上联系的类型（1:1, 1:n或m:n）。	联系名

注：联系也可以有属性，若有，则属性也用连线与该联系连接。

例如：学生与课程之间存在着多对多的联系。如图1-10所示就是一个描述“学生”与“课程”联系的E-R图。

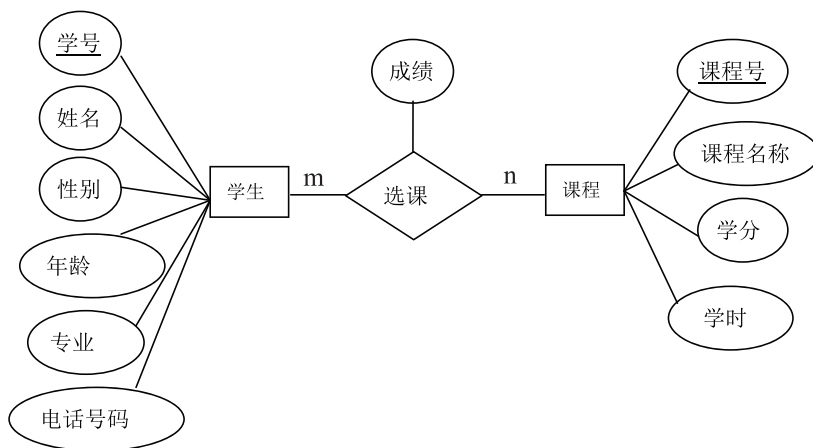


图1-10 “学生”与“课程”联系的E-R图

1.4.2 逻辑模型

概念模型设计结束后，就进入逻辑模型设计阶段。逻辑模型是将概念模型转化为具体的数据模型的过程，即按照概念结构设计阶段建立的基本E-R图，按选定的管理系统软件支持的数据模型（层次、网状、关系、面向对象），转换成相应的逻辑模型。这种转换要符合关系数据模型的原则。目前最流行就是关系模型（也就是对应的关系数据库）。

在数据库系统中，常用的逻辑模型有3种：分别为层次模型、网状模型、关系模型。

1.层次模型

层次模型是数据库系统中最早出现的数据模型。层次数据库系统的典型代表是IBM公司的IMS（Information Management System）数据库管理系统。

层次模型用树形结构来表示实体及其之间的联系。在这种模型中，数据被组织成由“根”开始的“树”，每个实体由根开始沿着不同的分支放在不同的层次上。树中的每一



个结点代表实体型，连线则表示它们之间的关系。根据树形结构的特点，建立数据的层次模型需要满足两个条件：

- (1) 有一个结点没有父结点，这个结点即为根结点；
- (2) 其他结点有且仅有一个父结点。

层次模型具有层次清晰、构造简单、易于实现等优点。但由于受到如上所述的两个条件的限制，它虽然可以比较方便地表示出一对一和一对多的实体联系，但是不能直接表示出多对多的实体，对于多对多的联系，必须先将其分解为几个一对多的联系，才能表示出来。因而，对于复杂的数据关系，实现起来较为麻烦，这就是层次模型的局限性。

采用层次模型设计的数据库称为层次数据库。图1-11给出了一个学校的层次模型。

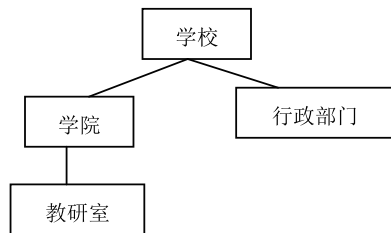


图1-11 层次模型

2.网状模型

网状模型用以实体型为结点的有向图来表示各实体及其之间的联系。其特点是：

- (1) 可以有一个以上的结点无父结点。
- (2) 至少有一个结点有多于一个的父结点。

由于树形结构可以看成是有向图的特例，所以网络模型要比层次模型复杂，但它可以直接用来表示“多对多”联系。然而由于技术上的困难，一些已实现的网状数据库管理系统（如DBTG）中仍然只允许处理“一对多”联系。

在以上两种数据模型中，各实体之间的联系是用指针实现的，其优点是查询速度快。但是当实体集和实体集中实体的数目都较多时（这对数据库系统来说是理所当然的），众多的指针使得管理工作相当复杂，对用户来说使用也比较麻烦。图1-12给出了一个学生选课的网络数据模型图。

例如：一个学生可以选修若干门课程，某一课程可以被多个学生选修，学生与课程之间是多对多联系。

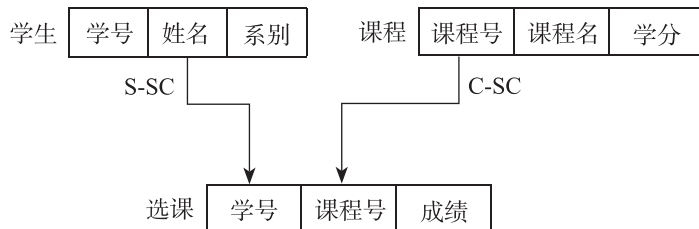


图1-12 学生选课的网络数据模型

3.关系模型

关系模型与层次模型和网状模型相比有着本质的差别，它是用二维表格来表示实体及

其相互之间的联系。在关系模型中，把实体集看成一个二维表，每一个二维表称为一个关系。每个关系均有一个名字，称为关系名。关系模型的结构如表1-2所示。

表1-2 关系模型

学号	姓名	性别	年龄	电话号码
1001	王小月	女	19	13871368300
1002	陈明	男	19	15387131230
1003	张盼	男	22	15397136830
1004	李丹	女	24	13387136820

虽然关系模型比层次模型和网状模型发展得晚，但是因为它建立在严格的数学理论基础之上，所以是目前比较常用的一种数据模型。自20世纪80年代以来，新推出的数据库管理系统几乎都支持关系模型。

1.4.3 物理模型

物理模型就是根据逻辑模型对应到具体的数据模型的机器实现的。物理模型是对真实数据库的描述。如关系数据库中的一些对象为表、视图、字段、数据类型、长度、主键、外键、索引、约束、是否可为空、默认值。

1.5 数据库应用系统体系结构

一个数据库应用系统通常包括数据存储层、应用层与用户界面三个层次。数据存储层一般由DBMS来承担对数据库的各种维护操作；应用层是使用某种程序设计语言实现用户要求的各项工作的程序；用户界面层是提供用户的可视化图形操作界面，便于用户与数据库系统之间的交互。

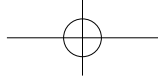
从最终用户角度看，数据库系统可分为单机结构、主从式结构、分布式结构、客户—服务器结构和浏览器—服务器结构五种，下面分别介绍。

1.5.1 单机结构

单机结构是一种比较简单的数据库系统。在单机系统中，整个数据库系统包括的应用程序、DBMS和数据库都安装在一台计算机上，由一个用户独占，不同机器之间不能共享数据。这种数据库系统也称桌面系统。

在这种桌面型DBMS中，数据的存储层、应用层和用户的界面层的所有功能都存储在单机上，因而适合于未联网的用户、个人用户及移动用户。若将这种系统应用于企事业单位中，容易造成大量的数据冗余。目前比较流行的桌面型DBMS有Visual FoxPro和Access等。

例如，一个企业的各个部门都使用本部门的机器来管理本部门的数据，各个部门的机器是独立的。由于不同部门之间不能共享数据，因此企业内部存在大量的冗余数据。



1.5.2 主从式结构

这是一个主机带多个终端的多用户结构。数据库系统，包括应用程序、DBMS、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数据库，共享数据资源。

这种结构的优点：易于管理、控制与维护；缺点：当终端用户数目增加到一定程度后，主机的任务会过于繁重，成为瓶颈，从而使系统性能下降。系统的可靠性依赖主机，当主机出现故障时，整个系统都不能使用。主从式结构示意图如图1-13所示。

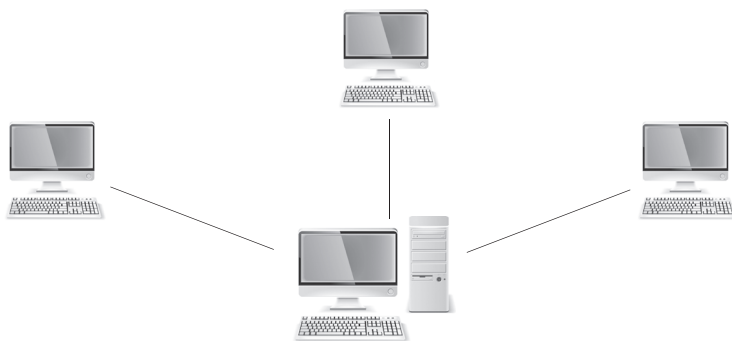


图1-13 主从式结构示意图

1.5.3 分布式结构

分布式结构是指地理上或物理上分散而逻辑上集中的数据库系统。管理这种结构的软件称为分布式DBMS，分布式数据库系统通常由计算机网络连接起来，被连接的逻辑单位（包括计算机、外部设备等）称为结点。

分布式数据库系统由多台计算机组成，每台计算机都配有各自的本地数据库。在分布式DBS中，大多数处理任务由本地计算机访问本地DB完成局部应用；对于少量本地计算机不能胜任的处理任务，通过网络同时存取和处理多个异地DB中的数据，执行全局应用。分布式DBS是计算机网络发展的必然产物。它适应了地理上分散的组织对于数据库应用的需求。

这种结构的优点：适应了地理上分散的公司、团体和组织对于数据库应用的需求；缺点：数据的分布存放给数据的处理、管理与维护带来困难。当用户需要经常访问远程数据时，系统效率会明显地受到网络传输的制约。分布式结构示意图如图1-14所示。

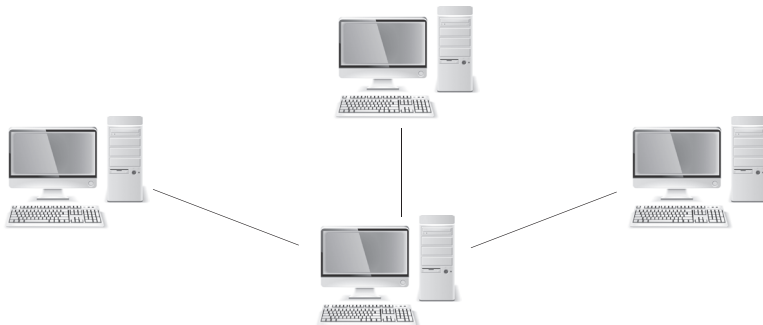


图1-14 分布式结构示意图

1.5.4 客户—服务器结构 (Client/Server, C/S)

主从式结构DBS中的主机和分布式DBS中的每个结点机都是一台通用计算机,既执行DBMS功能又执行应用程序。

随着工作站功能的增强和广泛使用,人们开始把DBMS功能和应用分开,网络中专门用于执行DBMS功能的计算机,称为数据库服务器,简称服务器 (Server);其他安装数据库应用程序的计算机称为客户机 (Client),这种结构称为客户—服务器 (C/S) 结构。它是当前非常流行的数据库系统结构。

在C/S结构的DBS中,数据存储层处于服务器上,而应用层和用户界面层处于客户机上。C/S结构的优点:一是可以减少网络流量,提高系统的性能、吞吐量和负载能力;二是使数据库更加开放,客户机和服务器可以在多种不同的硬件和软件平台上运行。C/S结构的缺点是系统的客户端程序更新升级有一定困难。

1.5.5 浏览器—服务器结构 (Browser/Server, B/S)

在C/S结构的DBS中,数据存储层处于服务器上,而应用层和用户界面层处于客户机上。C/S结构的缺点是系统的客户端程序更新升级有一定困难。而且对客户机的要求较高,将客户机上的应用层从客户机中分离出来,集中于一台高性能的计算机上,成为应用服务器。而客户机上的用户界面层由安装在客户机上的浏览器软件充当,这样就形成了现今流行的B/S结构数据库系统。

应用服务器也称Web服务器,它充当了客户端与数据库服务器的中介,架起了用户界面与数据库之间的桥梁。B/S结构有效克服了C/S结构的缺陷,客户机只要能运行浏览器软件即可。

1.6 数据库技术新发展

数据库技术最初产生于20世纪60年代中期,根据数据模型的发展,可以划分为以下3个阶段。

(1) 第一代的网状、层次数据库系统。层次模型和网状模型代表产品是1969年IBM公司研制的层次模型数据库管理系统IMS。

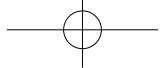
(2) 第二代的关系数据库系统。目前大部分数据库采用的是关系型数据库 (Oracle、SQL Server)。1970年IBM公司的研究员E.F.Codd提出了关系模型。其代表产品为system R和INGRES。

(3) 第三代的以面向对象模型为主要特征的数据库系统。第三代数据库产生于80年代,随着科学技术的不断进步,各个行业领域对数据库技术提出了更多的需求,关系型数据库已经不能完全满足需求,于是产生了第三代数据库。主要具有以下特征。

①支持数据管理、对象管理和知识管理。

②保持和继承了第二代数据库系统的技术。

③对其他系统开放,支持数据库语言标准,支持标准网络协议,有良好的可移植性、可连接性、可扩展性和互操作性等。第三代数据库支持多种数据模型 (比如关系模型和面向对象的模型),并和诸多新技术相结合 (比如分布处理技术、并行计算技术、人工



智能技术、多媒体技术、模糊技术），广泛应用于多个领域（商业管理、GIS、计划统计等），由此也衍生出多种新的数据库技术。

下面简单介绍几种新型数据库系统和数据管理技术。

1.6.1 面向对象数据库

面向对象数据库系统（Object-Oriented Database System, OODBS）是将面向对象的模型、方法和机制与先进的数据库技术有机地结合而形成的新型数据库系统。它从关系模型中脱离出来，强调在数据库框架中发展类型、数据抽象、继承和持久性。它的基本设计思想是，一方面把面向对象语言向数据库方向发展，使应用程序能够存取并处理对象，另一方面扩展数据库系统，使其具有面向对象的特征，提供一种综合的面向对象数据建模概念集，以便对现实世界中复杂应用的实体和联系建模。因此，面向对象数据库系统首先是一个数据库系统，具备数据库系统的基本功能，其次是一个面向对象的系统，针对面向对象的程序设计语言永久性对象存储管理设计的，充分支持完整的面向对象概念和机制。

面向对象数据库的优点：能够表示复杂的数据模型，但由于没有统一的数据模式和形式化理论，因此缺少严格的数据逻辑基础。而演绎数据库虽有坚强的数学逻辑基础，但只能处理平面数据类型。因此，部分学者将两者结合，提出了一种新的数据库技术——演绎面向对象数据库，并指出这一技术有可能成为下一代数据库技术发展的主流。

1.6.2 分布式数据库

分布式数据库是指利用高速计算机网络将物理上分散的多个数据存储单元连接起来组成一个逻辑上统一的数据库。

分布式数据库的基本思想是将原来集中式数据库中的数据分散存储到多个通过网络连接的数据存储节点上，以获取更大的存储容量和更高的并发访问量。近年来，随着数据量的高速增长，分布式数据库技术也得到了快速的发展，传统的关系型数据库开始从集中式模型向分布式架构发展，基于关系型的分布式数据库在保留了传统数据库的数据模型和基本特征下，从集中式存储走向分布式存储，从集中式计算走向分布式计算。分布式数据库有下列特点。

（1）数据的物理分布性：分布式数据库系统的数据分别存储在不同的计算机上，系统中每台计算机均称为一个数据结点。

（2）数据的逻辑相关性。

（3）区域自治性。

（4）数据结点之间通过通信网络进行联系。

分布式数据库系统（Distributed DataBase System, DDBS）通常使用较小的计算机系统，每台计算机可单独放在一个地方，每台计算机中都可能存有DBMS的一份完整拷贝副本，或者部分拷贝副本，并具有自己局部的数据库，位于不同地点的许多计算机通过网络互相连接，共同组成一个完整的、全局的逻辑上集中、物理上分布的大型数据库。

1.6.3 数据仓库技术

数据仓库技术（Data Warehousing, DW）是基于信息系统业务发展的需要，基于数据

库系统技术发展而来，并逐步独立的一系列新的应用技术。

随着90年代后期Internet 的兴起与飞速发展，我们进入了一个新的时代，大量的信息和数据迎面而来，用科学的方法去整理数据，从而从不同视角对企业经营各方面信息的精确分析、准确判断，比以往更为迫切，实施商业行为的有效性也比以往更受关注。

使用这些技术建设的信息系统我们称为数据仓库系统。随着数据仓库技术应用的不断深入，近几年数据仓库技术得到长足的发展。典型的数据仓库系统，比如：经营分析系统，决策支持系统等。随着数据仓库系统带来的良好效果，各行各业的单位已经能很好地接受“整合数据，从数据中找知识，运用数据知识、用数据说话”等新的关系到改良生产活动各环节、提高生产效率、发展生产力的理念。

数据仓库技术就是基于数学及统计学严谨逻辑思维的、并达成“科学的判断、有效的行为”的一个工具。数据仓库技术也是一种达成“数据整合、知识管理”的有效手段。

数据仓库是面向主题的、集成的、与时间相关的、不可修改的数据集合。这是数据仓库技术特征的定位。

1.6.4 大数据技术

大数据（Big Data），一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合。

大数据有4个基本特征：Volume（大量）、Velocity（高速）、Variety（多样）、Value（价值），即所谓的4V特征。

（1）数据容量大（Volume）。从TB级别，跃升到PB级别。

（2）数据类型繁多（Variety）。相对于以往便于存储的以文本为主的结构化数据，非结构化数据越来越多，包括网络日志、音频、视频、图片、地理位置信息等，这些多类型的数据对数据的处理能力提出了更高要求。

（3）商业价值高（Value）。价值密度的高低与数据总量的大小成反比。以视频为例，一部1小时的视频，在连续不间断的监控中，有用数据可能仅有一二秒。如何通过强大的机器算法更迅速地完成数据的价值“提纯”成为目前大数据背景下亟待解决的难题。

（4）处理速度快（Velocity）。1秒定律。这是大数据区别于传统数据挖掘的最显著特征。根据IDC的“数字宇宙”的报告，预计到2020年，全球数据使用量将达到35.2ZB。在如此海量的数据面前，处理数据的效率就是企业的生命。

随着云时代的来临，大数据也吸引了越来越多的关注。大数据需要特殊的技术，以有效地处理大量的数据。适用于大数据的技术，包括大规模并行处理数据库、数据挖掘、分布式文件系统、分布式数据库、云计算平台、互联网和可扩展的存储系统。

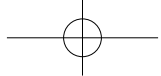


【思考练习】

一、选择题

1.在数据管理技术的发展过程中，经历了人工管理阶段、文件系统阶段和数据库系统阶段。在这几个阶段中，数据独立性最高的是（ ）阶段。

A.数据库系统 B.文件系统 C.人工管理 D.数据项管理



2. 下列叙述中正确的是 ()。
- A. 数据库是一个独立的系统, 不需要操作系统的支持
B. 数据库设计是指设计数据库管理系统
C. 数据库技术的根本目标是要解决数据共享的问题
D. 数据库系统中, 数据的物理结构必须与逻辑结构一致
3. 下列模式中, 能够给出数据库物理存储结构与物理存取方法的是 ()。
- A. 内模式 B. 外模式 C. 概念模式 D. 逻辑模式
4. 数据库系统的特点不包括 ()。
- A. 数据结构化
B. 数据独立性高
C. 数据冗余度低
D. 应用程序管理数据
5. 概念模型独立于 ()。
- A. DBMS B. E-R图 C. 信息世界 D. 现实世界
6. 在数据库系统中, 把可以相互区别的客观事物称为 ()。
- A. 属性 B. 字段 C. 文件 D. 实体
7. 学生社团可以接纳多名学生参加, 但每个学生只能参加一个社团, 从社团到学生之间的联系类型是 ()。
- A. 多对多 B. 一对多 C. 多对一 D. 一对一
8. () 是存储在计算机内有结构的数据的集合。
- A. 数据库系统 B. 数据库 C. 数据库管理系统 D. 数据结构
9. 数据库 (DB)、数据库系统 (DBS) 和数据库管理系统 (DBMS) 三者之间的关系是 ()。
- A. DBS 包括 DB 和 DBMS
B. DBMS 包括 DB 和 DBS
C. DB 包括 DBS 和 DBMS
D. DBS 就是 DB, 也就是 DBMS
10. 数据库管理系统能实现对数据库中数据的查询、插入、修改和删除等操作, 这种功能称为 ()。
- A. 数据定义功能 B. 数据管理功能 C. 数据操纵功能 D. 数据控制功能

二、问答题

1. 简述数据、信息、数据处理的含义与区别。
2. 目前比较流行的数据模型主要有哪几种? 各有何特别?
3. 简述 DB、DBMS、DBS 的含义。