



第一部分

- ▶ 第 1 章 数据库系统概述
- ▶ 第 2 章 数据模型
- ▶ 第 3 章 关系数据库
- ▶ 第 4 章 数据库设计

数据库技术是一门信息管理自动化学科，是计算机学科的一个重要分支。数据库技术所研究的问题是如何科学地组织和存储数据，在数据库系统中减少数据冗余，实现数据共享，以及如何保障数据安全，有效地获取和处理数据。

本章主要介绍数据库技术的发展历史，数据库系统的基本组成、概念和功能，以及数据库的体系结构。

1.1 数据库技术发展史

从 20 世纪 60 年代末开始到现在，数据库技术已经发展了 50 多年。在这近 50 多年的历程中，人们在数据库技术的理论研究和系统开发上取得了辉煌的成就，数据库系统已经成为现代计算机系统的重要组成部分。数据库技术最初是在大公司或大机构中用作大规模事务处理；随着个人计算机的普及，数据库技术被移植到 PC 上供单个用户应用；接着，由于 PC 在工作组内连成网，数据库技术就移植到工作组级；如今，数据库技术正在 Internet 中被广泛使用。

1.1.1 数据处理技术

1. 数据

数据 (Data) 是描述现实世界中各种具体事物或抽象概念的符号记录。除了常用的数字数据外，文字 (如名称)、图形、图像、声音等信息，也都是数据。日常生活中，人们使用交流语言 (如汉语) 去描述事物；在计算机中，为了存储和处理这些事物，就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如，在学生管理中，可以对学生的学号、姓名、性别和年龄等情况这样描述：201601001，张三，男，19。

2. 数据处理

数据处理 (Data Process) 是指对数据的收集、分类、组织、编码、存储、加工、计算、检索、维护、传播以及打印等一系列的活动。数据处理的目的是从大量的数据中，根据数据自身的规律和它们之间固有的联系，通过分析、归纳、推理等科学手段，提取出有效的信息资源。

在数据处理中，通常数据的加工、计算等比较简单，而数据的管理比较复杂。数据管理是数据处理的核心，是指数据的收集、分类、组织、编码、存储、检索、维护等操作，这部分操作是数据处理业务的基本环节，是任何数据处理业务中必不可少的共有部分，因此学习和掌握数据管理技术，能对数据处理提供有力的支持。

1.1.2 数据库技术的三个发展阶段

随着计算机硬件和软件的发展，数据库技术也在不断地发展。从数据管理的角度来说，



扫一扫

视频讲解

扫一扫



视频讲解

数据库技术经历了人工管理、文件系统和数据库系统三个阶段。

1. 人工管理阶段

20 世纪 50 年代中期以前, 计算机主要用于科学计算。从硬件上看, 外部存储器只有磁带、卡片和纸带等, 还没有磁盘等直接存取存储设备; 软件没有操作系统, 没有管理数据的软件, 没有高级语言, 只有汇编语言。

这个时期数据管理的特点如下。

(1) 数据无独立性。数据由计算或处理它的程序自行携带, 程序和数据是一个不可分割的整体, 应用程序依赖于数据的物理组织, 数据脱离了程序就无任何价值。数据只有与相应的程序一起保存才有价值, 否则就毫无用处。所以, 所有程序的数据均不单独保存。

(2) 数据不能共享。不同的程序均有各自的数据, 这些数据对不同的程序通常是不相同的, 当然不可共享; 即使不同的程序使用了相同的一组数据, 这些数据也不能共享, 程序中仍然需要各自加入这组数据, 谁也不能省略。这种数据的不可共享性, 必然导致程序与程序之间存在大量的重复数据, 浪费了存储空间。

人工管理阶段应用程序与数据之间的关系如图 1-1 所示。

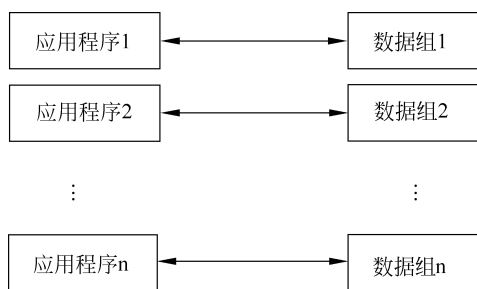


图 1-1 人工管理阶段应用程序与数据的对应关系

2. 文件系统阶段

20 世纪 50 年代后期至 20 世纪 60 年代中后期, 计算机的应用范围逐渐扩大, 不仅用于科学计算, 还大量用于管理。随着数据量的增加, 数据的存储、检索和维护问题成为紧迫的需要, 数据结构和数据管理技术迅速发展起来。硬件方面, 外部存储器已有磁盘、磁鼓等直接存取的存储设备; 软件方面, 出现了高级语言和操作系统。操作系统中的文件系统是专门

管理外存的数据管理软件, 该系统把计算机中的数据组织成相互独立的数据文件。数据处理方式有批处理, 也有联机实时处理。

这个阶段有如下几个特点。

(1) 数据以“文件”形式可长期保存在外部存储器的磁盘上。对数据的操作以记录为单位, 文件的建立、存取、查询、插入、删除、修改等所有操作, 都要用程序来实现。

(2) 程序与数据之间具有了一定的独立性——“设备独立性”, 即程序只需用文件名就可与数据打交道, 不必关心数据的物理位置。

(3) 文件结构的设计仍然是基于特定的用途, 程序基于特定的物理结构和存取方法, 因此程序与数据结构之间的依赖关系并未根本改变。由于文件之间缺乏联系, 造成每个应用程序都有对应的文件, 有可能同样的数据在多个文件中重复存储, 数据冗余度大。

文件系统阶段应用程序与数据之间的关系如图 1-2 所示。

3. 数据库系统阶段

20 世纪 60 年代后期以来, 计算机应用越来越广泛, 数据量急剧增加, 而且数据的共享要求越来越高。计算机的硬件和软件都有了进一步的发展, 硬件方面, 有了大容量的磁盘; 软件方面, 传统的文件系统已经不能满足人们的需求, 能够统一管理和共享数据的数据库管

扫一扫



视频讲解

理系统 (Database Management System, DBMS) 应运而生。所以, 此阶段将数据集中存储在各数据库中, 由 DBMS 进行统一组织和管理。从处理方式上讲, 联机实时处理要求更多了, 并开始提出和考虑分布处理。

数据库系统阶段的特点也是优点如下。

1) 数据结构化

数据结构化是数据库系统与文件系统的根本区别。有了 DBMS 后, 数据库中的数据不再针对某一应用, 而是面向整个应用系统, 它是对整个组织的各种应用 (包括将来可能的应用) 进行通盘考虑后建立起来的总的数据结构。

2) 较高的数据共享性

数据共享是指允许多个用户同时存取数据而互不影响, 该特征正是数据库技术先进性的体现。数据库系统从整体角度描述数据, 数据不再面向某个应用而是面向整个系统, 因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余, 节约存储空间。

3) 较高的数据独立性

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。应用程序不随数据存储结构的变化而变化, 因为应用程序以简单的逻辑结构操作数据而无须考虑数据的物理结构, 简化了应用程序的编制和程序员的工作负担。

4) 数据由 DBMS 统一管理和控制

数据库的共享是并发的共享, 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数据。因此, DBMS 还必须提供数据的统一管理和控制功能。

DBMS 加入了安全保密机制, 可以防止数据被非法存取; DBMS 的数据完整性保护可以保障数据的正确性、有效性和相容性, 完整性检查将数据控制在有效的范围内或保证数据之间满足一定的关系; 当多个用户的并发进程同时存取、修改数据库时, 可能会发生相互干扰而得到错误的结果, 或使得数据库的完整性遭到破坏, 因此 DBMS 必须对多用户的并发操作加以控制和协调; 另外, DBMS 还采取了一系列措施, 以实现对数据库破坏后的恢复。

数据库系统阶段应用程序与数据之间的关系如图 1-3 所示。

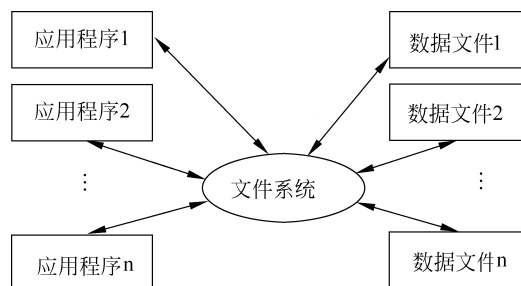


图 1-2 文件系统阶段应用程序与数据的对应关系

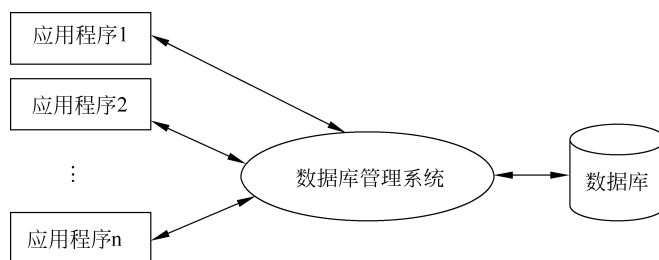


图 1-3 数据库系统阶段应用程序与数据的对应关系

1.2 数据库系统的组成

数据库系统 (Database System, DBS) 是指在计算机系统中引入数据库后的系统, 通常由软件、数据库和人员组成。软件包括操作系统、基于某种宿主语言的数据库开发工具、数据库应用系统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; 人员包括数据管理员、程序员和终端用户。

数据库系统一般由数据库、数据库管理系统、数据库开发工具、数据库应用系统和人员构成。数据库系统可以用图 1-4 来表示。

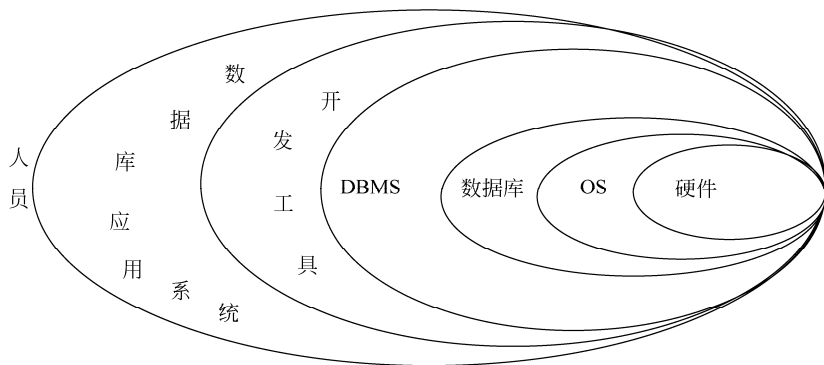


图 1-4 数据库系统的层次结构

1. 数据库

数据库 (Database, DB) 是指长期存储在计算机内有组织的、可共享的数据集合, 即在计算机系统中按一定的数据模型组织、存储和使用的相关联的数据集合。它不仅包括描述事物的数据本身, 还包括相关事物之间的联系。数据库中的数据以文件的形式存储在存储介质上, 它是数据库系统操作的对象和结果。

2. 数据库管理系统

DBMS 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。它是建立在操作系统的基础上, 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, 包括数据库的创建、查询、更新及各种数据控制等。数据库中数据的插入、修改和检索均要通过数据库管理系统进行, 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。

一般说来, 数据库管理系统的功能主要包括以下几方面。

1) 数据定义功能

DBMS 提供数据定义语言 (Data Definition Language, DDL) 对数据库中的对象进行定义, 使用户能够定义构成数据库结构的各级模式, 包括定义表结构、定义索引以及定义视图, 也能够定义数据库的完整性、安全性等。这些定义存储在数据字典中, 是 DBMS 运行的基本依据。

2) 数据操纵功能

DBMS 提供数据操纵语言 (Data Manipulation Language, DML) 操纵数据库中的数据, 实

扫一扫



视频讲解

扫一扫



视频讲解

现对数据库的基本操作, 包括对数据库中的数据进行检索、插入、修改和删除等基本操作。

3) 数据库运行控制功能

对数据库的运行进行管理是数据库管理系统运行时的核心部分, 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以及数据库的内部维护等。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) 数据库的组织、存储和管理

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, 数据库管理系统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、删除和查改等操作的时间效率。

5) 建立和维护数据库

建立数据库包括数据库初始数据的输入与数据转换等。维护数据库包括数据库的备份和还原、数据库的重组与重构造、性能的监视与分析等。

常见的数据库管理系统有 Access、SQL Server、MySQL、Oracle、DB2 等。

3. 数据库应用系统

凡使用数据库技术管理其数据的系统都称为数据库应用系统 (Database Application System, DBAS)。数据库应用系统的应用非常广泛, 它可以用于事务管理、计算机辅助设计、计算机图形分析和处理及人工智能等系统中。

4. 人员

1) 终端用户

终端用户 (End User) 是数据库的使用者, 通过应用程序与数据库进行交互。他们不需要具有数据库的专业知识, 只是通过应用程序的用户接口存取数据库的数据, 使用数据库来完成其业务活动, 直观地显示和使用数据。

2) 应用程序员

程序员 (Programmer) 负责分析、设计、开发、维护数据库系统中的各类应用程序, 数据库系统一般需要一个以上的程序员在开发周期内完成数据库结构设计、应用程序开发等任务, 在后期管理应用程序, 保证使用周期中对应用程序在功能及性能方面的维护、修改工作。

3) 数据库管理员

数据库管理员 (Database Administrator, DBA) 的职能是管理、监督、维护数据库系统的正常运行, 负责全面管理和控制数据库系统。数据库管理员的主要职责包括设计与定义数据库系统, 帮助最终用户使用数据库系统, 监督与控制数据库系统的使用和运行, 改进和重组数据库系统, 优化数据库系统的性能, 定义数据的安全性和完整性约束, 备份与恢复数据库等。

1.3 数据库的体系结构

虽然现在 DBMS 的产品多种多样, 在不同的操作系统支持下工作, 但是大多系统在总的体系结构上都具有三级模式的结构特征。





1.3.1 数据库的三级模式结构

为了保障数据与程序之间的独立性,使用户能以简单的逻辑结构操作数据而无须考虑数据的物理结构,简化应用程序的编制和程序员的负担,增强系统的可靠性,通常 DBMS 将数据库的体系结构分为三级模式:外模式、模式和内模式。三级模式结构如图 1-5 所示。

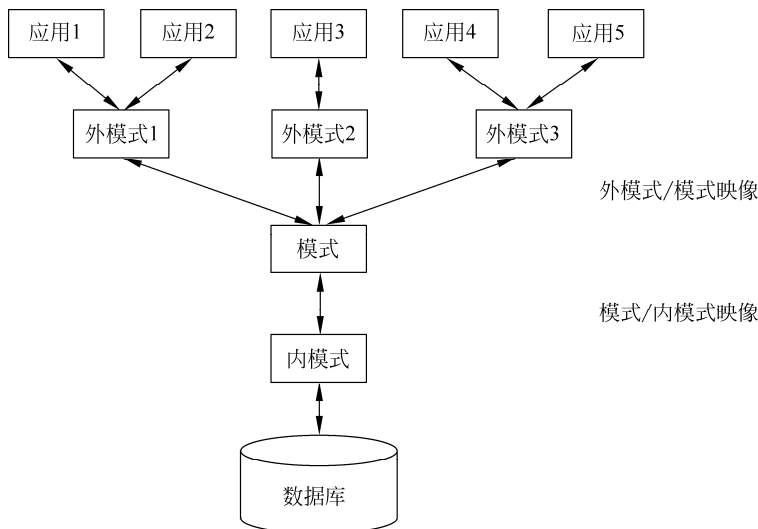


图 1-5 数据库系统的三级模式结构

1. 模式

模式 (Schema) 也称概念模式或逻辑模式,是对数据库中全部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。一个数据库只有一个模式,通常以某种数据模型为基础,统一综合地考虑了所有用户的需求,并将这些需求有机地结合成一个逻辑整体。定义模式时不仅要定义数据的逻辑结构,例如数据记录由哪些数据项构成,数据项的名称、类型、取值范围等,而且还要定义数据项之间的联系,定义不同记录之间的联系,以及定义与数据有关的完整性、安全性等要求。数据库管理系统提供模式描述语言 (模式 DDL) 来定义模式。

2. 内模式

内模式 (Internal Schema) 也称存储模式或物理模式,是对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。例如,记录的存储方式是顺序存储、按照 B+树结构存储还是按 Hash (哈希) 方法存储;索引按照什么方式组织;数据是否压缩存储,是否加密等。数据库管理系统提供内模式描述语言 (内模式 DDL) 来定义内模式。

3. 外模式

外模式 (External Schema) 也称子模式或用户模式,它是对数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述。外模式通常是模式的子集,一个数据库可以有多个外模式,但一个应用程序只能使用同一个外模式。外模式是保证数据库安全性的一个有效措施,每个用户只能看见或访问所对应的外模式中的数据,数据库中的其余数据是不可见的。数据库管理系统提供外模式描述语言 (外模式 DDL) 来定义外模式。

例如,图书出版公司数据库三级模式中,模式、内模式都只有一个,外模式有三个,如图1-6所示。

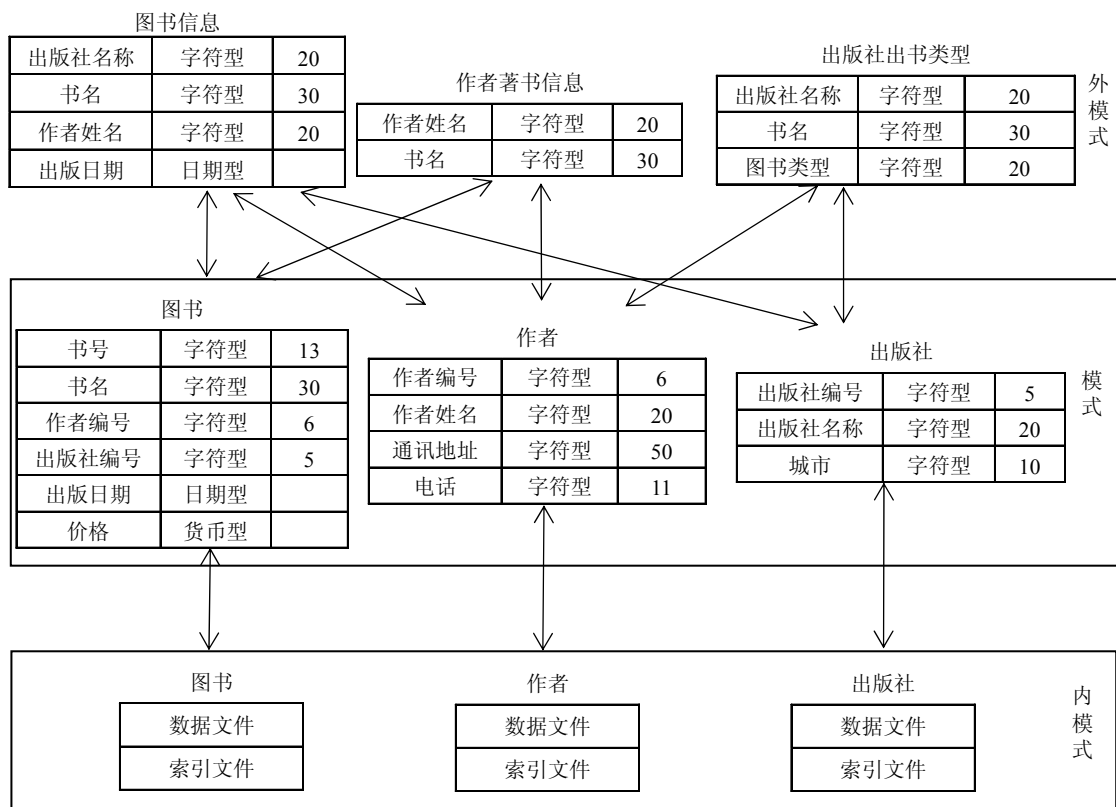


图 1-6 图书出版公司数据库三级模式实例

1.3.2 数据库的两级映像

数据库的三级模式结构是数据的三个抽象级别。它把数据的具体组织留给 DBMS 去做,用户只要抽象地处理数据,而不必关心数据在计算机中的表示和存储,这样就减轻了用户使用系统的负担。

三级模式结构之间差别往往很大,为了实现这三个抽象级别的联系和转换,DBMS 在三级模式结构之间提供了两级映像:外模式/模式映像,模式/内模式映像。

1. 外模式/模式映像

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,对应于同一个模式可以有任意多个外模式。对于每个外模式,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。这些映像定义通常包含在各自外模式的描述中。当模式改变时(如增加新的关系、新的属性、改变属性的数据类型等),由数据库管理员对各个外模式/模式映像做相应改变,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,因而应用程序不必修改,保证了数据与程序的逻辑独立性,简称逻辑数据独立性。

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。例如,说明逻辑记录和字段在内部是如何

扫一扫



视频讲解

表示的。该映像定义通常包含在模式描述中。当数据库的存储结构改变了（如选用了另一种存储结构），由数据库管理员对模式/内模式映像做相应改变，可以保证模式保持不变，因而应用程序也不必改变。保证了数据与程序的物理独立性，简称物理数据独立性。

数据与程序之间的独立性，使数据的定义和描述可以从应用程序中分离出去。另外，由于数据的存取由 DBMS 管理，用户不必考虑存取路径等细节，从而简化了应用程序的编写，大大减少了对应用程序的维护和修改。

习 题 1

1. 数据库的发展历史分为哪几个阶段？各有什么特点？
2. 简述数据、数据库、数据库管理系统、数据库应用系统的概念。
3. 简述数据库管理系统的功能。
4. 数据库的三级模式和两级映像是什么？
5. 名词解释：模式、内模式、外模式。
6. 简述数据库的逻辑独立性和物理独立性。