

# 第 1 章

## 数据库入门

数据库是依照某种数据模型组织起来并存放在二级存储器中的数据集合,可以将其视为电子化的文件柜。数据库具有不重复、以最优方式提供多种应用服务、数据结构独立于应用程序、对数据的操作由统一软件进行管理和控制等特点。从数据管理技术的发展历程来看,数据库是由文件管理系统发展起来的,是数据管理的高级阶段。

本章重点内容:

- 了解数据库的发展与组成
- 掌握数据库体系结构
- 掌握数据库的数据模型
- 了解常见的数据库

## 1.1 数据库系统概述

数据库(DataBase, DB)是按照数据结构来组织、存储和管理数据的仓库,产生于距今 60 多年前。随着信息技术和市场的发展,特别是 20 世纪 90 年代以后,数据管理不再仅仅用于存储和管理数据,出现了用户所需要的各种数据管理的方式。从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属于数据库的范畴,并在各个方面得到了广泛的应用。

在信息化社会充分有效地管理和利用各类信息资源是进行科学研究和决策管理的前提条件。数据库技术是管理信息系统、办公自动化系统、决策支持系统等各类信息系统的核心部分,是进行科学研究和决策管理的重要技术手段。

### 1.1.1 数据库技术的发展

使用计算机后,随着数据处理量的增长,产生了数据管理技术。数据管理技术的发展与计算机硬件(主要是外部存储器)、系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了 4 个阶段:人工管理阶段、文件系统阶段、数据库阶段和高级数据库技术阶段。其中,数据库阶段和高级数据库技术阶段可以统称为系统阶段,即由数据库系统进行管理数据的阶段。

#### 1. 人工管理

20 世纪 50 年代中期之前,计算机的软硬件均不完善。硬件存储设备只有磁带、卡片和纸



带，软件方面还没有操作系统，当时的计算机主要用于科学计算。人工管理阶段由于还没有软件系统对数据进行管理，程序员在程序中不仅要规定数据的逻辑结构，还要设计其物理结构，包括存储结构、存取方法、输入输出方式等。当数据的物理组织或存储设备改变时，用户程序就必须重新编制。由于数据的组织面向应用，不同的计算程序之间不能共享数据，使得不同的应用之间存在大量的重复数据，很难维护应用程序之间数据的一致性。这一阶段的主要特征可归纳为如下几点：

- 计算机中没有支持数据管理的软件。
- 数据组织面向应用，数据不能共享，数据重复。
- 在程序中要规定数据的逻辑结构和物理结构，数据与程序不独立。
- 数据处理方式——批处理。

## 2. 文件系统

这一阶段处于 20 世纪 50 年代中期到 60 年代中期，其主要标志是计算机中有了专门管理数据库的软件——操作系统。操作系统文件管理功能的出现标志着数据管理步入一个新的阶段。

在文件系统阶段，数据以文件为单位存储在外存，由操作系统统一管理，而操作系统为用户使用文件提供友好界面。该阶段中的文件逻辑结构与物理结构脱钩，程序和数据分离，使数据与程序有了一定的独立性。用户的程序与数据可分别存放在外存储器上，各个应用程序可以共享一组数据，实现了以文件为单位的数据共享。

由于数据的组织仍然是面向程序的，因此仍存在大量的数据冗余。同时，由于数据的逻辑结构不能方便地修改和扩充，因此数据逻辑结构的每一点微小改变都会影响应用程序。此外，由于文件之间互相独立，因此不能反映现实世界中事物之间的联系，而操作系统不负责维护文件之间的联系信息。如果文件之间有内容上的联系，那么只能由应用程序去处理，这加大了程序设计人员的工作量。

## 3. 系统阶段

20 世纪 60 年代后，随着计算机在数据管理领域的普遍应用，人们对数据管理技术提出了更高的要求：希望面向企业或部门，以数据为中心组织数据，减少数据的冗余，提供更高的数据共享能力，同时要求程序和数据具有较高的独立性，当数据的逻辑结构改变时，不涉及数据的物理结构，也不影响应用程序，以降低应用程序研制与维护的费用。数据库技术正是在这样的应用需求基础上发展起来的。

数据管理技术经历了人工管理阶段和文件阶段后，获得了大量的技术积累，这为数据库的诞生奠定了基础。具体来说，数据库技术有如下特点：

（1）面向企业或部门。数据库以数据为中心进行数据的组织，形成综合性的数据库，从而为各应用共享。

（2）采用一定的数据模型。数据模型不仅描述了数据本身的特点，而且描述了数据之间的联系。



(3) 数据冗余小, 易修改、易扩充。数据库技术阶段中, 不同的应用程序根据处理要求从数据库中获取需要的数据, 这样就减少了数据的重复存储, 也便于增加新的数据结构, 便于维护数据的一致性。

(4) 程序和数据有较高的独立性。

(5) 具有良好的用户接口, 用户可方便地开发和使用数据库。

(6) 对数据进行统一管理和控制, 提供了数据的安全性、完整性以及并发控制。

数据管理技术从文件系统发展到数据库系统, 这在信息领域中具有里程碑的意义。在文件系统阶段, 人们在信息处理中关注的中心问题是系统功能的设计, 因此程序设计占主导地位; 而在数据库阶段, 数据开始占据了中心位置, 数据的结构设计成为信息系统首先关心的问题, 而应用程序则以既定的数据结构为基础进行设计。

#### 4. 发展趋势

随着信息管理内容的不断扩展, 出现了丰富多样的数据模型(层次模型、网状模型、关系模型、面向对象模型、半结构化模型等), 新技术也层出不穷(数据流、Web 数据管理、数据挖掘等)。每隔几年, 国际上一些资深的数据库专家就会聚集一堂, 探讨数据库现状、研究存在的问题和未来需要关注的新技术焦点。

数据库与学科技术的结合将会建立一系列新数据库, 如分布式数据库、并行数据库、知识库、多媒体数据库等, 这将是数据库技术重要的发展方向。未来数据库技术及市场发展的两大方向是数据仓库和电子商务, 数据管理技术将在数据仓库技术以及与之相关的数据挖掘和知识发现领域持续发展。

### 1.1.2 数据库系统组成

数据库系统(Database System, DBS)是指一个具体的数据库管理系统软件和用它建立起来的数据库, 通常由系统软件、数据库和数据管理员组成。系统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(DBMS); 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(DBA)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, 数据库管理员一般由业务水平较高、资历较深的人员担任。

数据库系统是软件研究领域的一个重要分支, 常称为数据库领域。数据库系统是为适应数据处理的需要而发展起来的一种较为理想的数据处理的核心机构, 具体来说由如下部分组成。

(1) 数据库: 长期存储在计算机内, 有组织、可共享的数据集合。数据库中的数据按一定的数学模型组织、描述和存储, 具有较小的冗余、较高的数据独立性和易扩展性, 并可为各种用户共享。

(2) 硬件: 构成计算机系统的各种物理设备, 包括存储所需的外部设备, 如物理硬盘、光盘等媒介。



**提示**

硬件的配置应满足整个数据库系统的需要。

(3) 系统软件：包括操作系统、数据库管理系统及应用程序。数据库管理系统（DataBase Management System, DBMS）是数据库系统的核心软件，在操作系统的支持下工作，是科学地组织和存储数据、高效获取和维护数据的系统软件。其主要功能包括数据定义、数据操纵、数据库的运行管理和数据库的建立与维护。

(4) 人员：主要包括如下 4 类。

- 第一类为系统分析员和数据库设计人员。系统分析员负责应用系统的需求分析和规范说明，他们和用户及数据库管理员一起确定系统的硬件配置，并参与数据库系统的概要设计。数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。
- 第二类为应用程序员，负责编写使用数据库的应用程序。这些应用程序可对数据进行检索、建立、删除或修改。
- 第三类为最终用户，他们利用系统的接口或查询语言访问数据库。
- 第四类是数据库管理员（Data Base Administrator, DBA），负责数据库的总体信息控制。DBA 的具体职责包括确定数据库中的信息内容和结构，决定数据库的存储结构和存取策略，定义数据库的安全性要求和完整性约束条件，监控数据库的使用和运行，负责数据库的性能改进、数据库的重组和重构，以提高系统的性能。

## 1.2 数据库体系结构

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、概念模式和内模式。数据库体系结构能够有效组织、管理数据，提高数据库的逻辑独立性和物理独立性。

### 1.2.1 什么是模式

虽然实际的数据库管理系统产品种类很多，支持不同的数据模式，使用不同的数据库语言，建立在不同的操作系统之上，数据的存储结构也各不相同，但它们在体系结构上通常具有相同的特征，即采用三级模式结构并提供两级映像功能。

模式是数据库中全体数据的逻辑结构和特征的描述，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，因为数据库中的数据是在不断更新的。模式反映的是数据的结构及其联系，而实例反映的是数据库某一时刻的状态。