

第 1 章 数据库概述



数据库是各种信息系统、应用程序及软件的核心和关键，也是信息科学技术的重要组成部分。数据库技术解决了信息处理过程中大量数据的组织、存储、管理、查询和统计等应用问题，使得在数据库系统中可以减少数据冗余，实现数据共享，保障数据一致性，实现数据快速检索及统计分析等。本章主要介绍数据库的概念、数据模型、关系运算等。

1.1 数据库的基本概念

数据库是便捷、高效、科学管理数据的重要平台，在大数据发展的背景下，数据库有效解决了海量数据管理、存储与应用的难题，因此被广泛应用于各行业领域。本节主要介绍信息与数据库、SQL 语言、数据库管理系统等。

1.1.1 信息与数据库

数据（data）是指描述事物属性的元素，包括金融数据、通信数据、教育数据等多种类型，同时数据具有多种形式，包括文本、图表、图像、声音、语言、视频等。而信息是按特定条件筛选的数据的集合，不仅具有存储、加工、传播等属性，还具有重要的使用价值。

数据库（database，DB）是由具有某些特性相互关联的数据组成的集合，可视为存储和管理数据的仓库，用户可对仓库中的数据进行增加、删除、修改等操作。同时，数据库具有数据结构化、共享化、独立性、统一性等特点。数据库不仅实现了面向整体的数据结构化，同时能够被多用户、多系统共享使用，并且数据库由数据管理员统一管理，因此极大地保障了数据的安全性、准确性与一致性，有效降低了数据丢失风险。

在实际应用中，可根据需求创建不同类型的数据库，例如银行可以创建一个用户数据库，将用户的姓名、出生日期、身份证号、账户金额等信息存放在一起，可以通过查询生成用户信息表、流水账单、交易明细表等；高校可以创建一个学生数据库，将学生的身份信息、上课时间、课程成绩等信息存放在一起，可以通过查询生成学生档案表、上课课表、成绩单等；同学可以建立一个就业信息数据库并发布，供同学们共享。

1.1.2 SQL 语言

SQL（structured query language，SQL）是一种用于程序设计以及数据查询、存取、更新的语言，不仅具有操作一体化、使用方式灵活、非过程化等特点，而且语言简洁、语法格式简单。SQL 语言主要由以下 6 部分组成。

（1）数据定义语言（data definition language，DDL），包括 CREATE、ALTER 和 DROP

等语句。DDL 的操作对象是数据库和表，主要用于创建和修改数据库、增加或删除表以及修改表的列或字段等。

(2) 数据操作语言 (data manipulation language, DML), 包括 SELECT、INSERT、UPDATE 和 DELETE 等语句。DML 的操作对象是表中数据, 主要用于插入、修改、删除表中数据等。其中, INSERT 语句用于向表中插入数据, UPDATE 语句用于修改表中数据, DELETE 语句用于删除表中一行、多行或全部数据。

(3) 数据控制语言 (data control language, DCL), 包括 REVOKE、DENY 和 GRANT 等语句。DCL 的操作对象是数据库用户, 主要用于设置和分配数据库用户或角色的操作权限。其中, REVOKE 语句用于回收数据库中当前用户或角色的权限, DENY 语句用于禁止数据库中当前用户或角色的权限, GRANT 语句用于分配数据库用户或角色权限。

(4) 数据查询语言 (data query language, DQL), 主要为 SELECT 语句, 常用 SELECT 语句与 FROM 和 WHERE 子句组合, 用于查询表中数据。

(5) 指针控制语言 (cursor control language, CCL), 包括 DECLARE CURSOR、FETCH INTO 和 UPDATE WHERE CURRENT 等语句。CCL 的操作对象是数据表, 主要用于对一个或多个表进行单独的操作。

(6) 事务处理语言 (transaction processing language, TPL), 包括 BEGIN TRANSACTION、COMMIT 和 ROLLBACK 等语句。TPL 的主要作用是保障每一条 DML 语句成功执行, 从而使事务完成执行。

1.1.3 数据库管理系统

数据库管理系统 (database management system, DBMS) 是一种用于数据库管理的系统平台, 由相互关联数据的集合和管理数据的程序模块组成。数据库管理系统设计的主要目的是方便、高效地管理海量数据, 极大地提高了数据使用和维护的效率, 不仅可以根据实际需求创建或删除数据库, 还可对数据库对象进行管理。随着数据管理技术的不断发展, 出现了多种数据库管理系统, 较为常用的有 MySQL、SQL Server、Oracle 等。

数据库管理系统程序模块主要由存储管理程序、查询处理程序、事务管理程序、恢复管理程序组成。存储管理程序用于存储空间分配和数据存储, 可自动分配服务器硬盘空间, 将数据存入服务器的数据库中。查询处理程序主要用于处理用户输入的查询语句, 一般为 SQL 语句。通常一个事务对应一个具体的工作任务, 例如银行的面向用户的储蓄业务, 事务管理程序主要用于数据库的事务处理, 可控制和管理事务并发问题, 有效保障数据库管理系统事务执行。恢复管理程序主要用于数据库管理系统操作日志管理, 可记录用户在数据库上进行的操作, 当发生异常情况时, 可将数据库中的数据恢复到发生异常情况前的状态。DBMS 架构如图 1.1 所示。

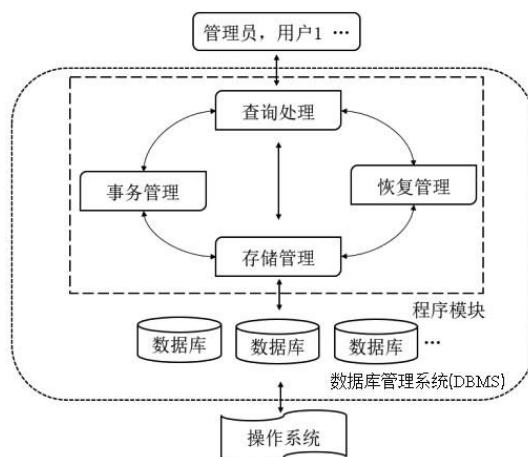


图 1.1 DBMS 架构

1.1.4 数据库系统

数据库系统 (database system, DBS) 是指引入数据库后的计算机应用系统。数据库系统通常包括硬件和软件, 由数据库、数据库管理系统、应用程序及开发工具、数据库管理员 (DBA)、用户等组成。

数据库、数据库管理系统、数据库系统的关系如图 1.2 所示, 数据库系统的构成图示如图 1.3 所示。

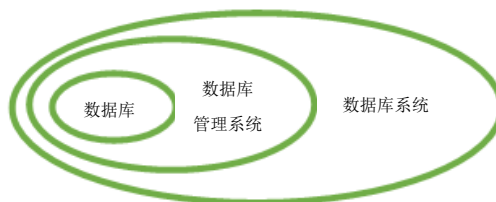


图 1.2 数据库、数据库管理系统、数据库系统的关系

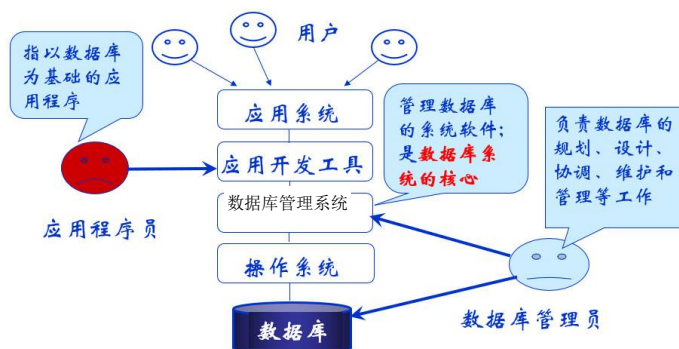


图 1.3 数据库系统的构成图示

1.2 数据模型

数据模型是数据库的基础，包括网状型、层次型、关系型、面向对象型等类型，其中关系型数据模型是目前数据库系统常用的一种模型。本节主要介绍数据模型的类型、概念模型及其表示方法等。

1.2.1 数据模型简介

数据模型是数据库系统的核心，大多数数据库管理系统都是基于数据模型而设计的。数据模型一般由数据结构、数据操作、完整性约束条件 3 部分组成。其中，数据结构针对的是数据的类型、内容、特性以及数据间的关系，因此被用于描述数据库的静态特性；数据操作是指数据库中对象允许执行的操作，例如对数据表执行查询、删除、新增、修改等操作；为保证数据的准确性和相容性，完整性约束条件可规定数据库的状态以及状态变化时所应满足的条件。

按数据模型的特性进行划分，可将数据模型分为网状型、层次型、关系型和面向对象型，其中关系型数据模型是目前数据库系统常用的数据模型。

网状型数据模型是最早被用于数据库管理系统的数据库模型，其采用网状结构表示实体以及实体间的联系，实体间的联系为多对多的联系 ($m:n$)，常用于处理数据类型为节点网状型数据库。

层次型数据模型是仅次于网状型数据库模型被用于数据库管理系统的数据库模型，其采用树状结构表示实体间的联系，实体间的联系为一对多的联系 ($1:n$)。层次型数据库模型包含一个根节点以及多个子节点和叶节点，每个节点代表一个实体，除了叶节点，其他子节点不能脱离根节点而单独存在。在实际应用中，如果删除了根节点的数据，则子节点和叶节点的数据也会被删除。

关系型数据库模型是使用数据表来表示实体间的联系，操作对象是二维表。关系型数据库模型是目前使用最多的数据库模型，许多数据库管理系统均使用该数据库模型，如 MySQL、Oracle、SQL Server 等。

面向对象型数据库模型是以面向对象的方式设计数据库系统，具有类和继承等特性。面向对象型数据库模型的数据结构面向的是对象，并以对象为单位进行存储，其中每个对象包含对象的属性和联系。

1.2.2 概念模型及其表示

关系型数据库是建立在关系模型上的数据库系统，而关系模型由表及其中的联系组成，因此关系型数据库是由表组成的集合。例如在表 1.1 中记录了每个学生的基本信息，将一行的学号、姓名、年级联系在一起即可描述某个具体的学生，因此表中每一行代表了一种联系。

在关系模型的概念中，关系指对应的二维表。每一行代表一个实体，用元组指代表中的行。每一列代表一种属性，用属性指代表中的列，每个实体都有一个属性集合，能够唯一标

识这个属性集合的属性或属性子集被称为主键。在数学领域中，某个元素的取值范围被称为值域，因此属性取值的集合被称为域，属性的取值可以是数字、文字、符号等。例如，在表 1.1 中每一行代表一个学生实体，有学号、姓名、年级 3 个属性，主键是“学号”，“学号”的域是所有学号的集合，“姓名”的域是所有学生姓名的集合，“年级”的域是学生所在年级的集合。

表 1.1 学生信息简况表

学 号	姓 名	年 级
2020001	张三	2020
2020002	李四	2020
2022003	王五	2022
...	...	...

设计数据库时，首先需要建立数据库的概念模型。用矩形表示实体，用椭圆表示实体的属性，并用直线将实体及其属性连接。在实际应用中，实体之间是存在联系的，分为一对一的联系（1:1）、一对多的联系（1:n）、多对多的联系（m:n），用菱形表示实体间的联系，用直线将菱形与有关实体连接，并标注联系类型。概念模型的表示如图 1.4 所示。

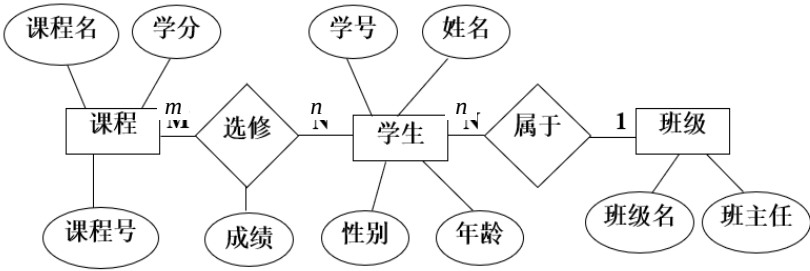


图 1.4 概念模型示例

1.3 关系运算

关系运算是指作用于单个关系或多个关系最终得到运算结果为单个关系的运算方法，它也可作用于运算结果上。由前面的内容可知，关系模型中用关系指代表，那么关系运算实际上就是表格或表格之间的运算。由于关系是集合，因此可在关系上使用集合运算，如集合的并、交、差运算。

关系运算按运算符的不同可分为传统的集合运算和专门的关系运算两类。本节主要介绍专门的关系运算。

1.3.1 选择运算

选择运算是指从表中筛选出满足一定条件的记录组合为一张新表的运算，用符号 σ 表示。选择运算实质上是对表的行进行操作。

【例 1.1】从学生信息表（见表 1.2）中查询 IS 系的全体学生。

$\sigma_{Sdept='IS'}(Student)$
或 $\sigma_{S=IS}(Student)$

表 1.2 学生信息表 Student

Sno	Sname	Ssex	Sage	Sdept
95001	李勇	男	20	CS
95002	刘晨	女	19	IS
95003	王敏	女	18	MA
95004	张立	男	19	IS

1.3.2 投影运算

投影运算与选择运算类似，是从表中选出满足一定条件的列组合成一张新表，用符号 Π 表示。投影运算实质上是对表的列进行操作。

【例 1.2】从学生信息表（见表 1.2）中选出学号和姓名列。

$\Pi_{Sno, Sname}(Student)$

1.3.3 连接运算

连接运算实质上是将两张表的行连接起来。它就是根据给定的条件，从两个已知的关系 R 和 S 的笛卡儿积中，选取满足连接条件的若干元组，组成一个新的关系。连接运算具体又分为以下几种。

- （1）条件连接：选取满足条件的元组组成新关系。
- （2）等值连接：选取满足等值条件的元组组成关系。
- （3）自然连接：是一种特殊的等值连接，只不过它是选取满足公共属性等值条件的元组，组成关系。

【例 1.3】从学生信息表（见表 1.2）和选修表（见表 1.3）中查询李勇同学选修的课程号及成绩，结果如表 1.4 所示。

$\Pi_{Cno, Score}(\sigma_{Sname='李勇'}(Student \bowtie SC))$

表 1.3 选修表 SC

Sno	Cno	Score
95001	C1	89
95002	C2	88
95001	C3	90
95003	C2	92

表 1.4 连接运算的结果

Cno	Score
C1	89
C3	90

1.4 本章小结

本章介绍了数据库的基本概念、关系型数据库等，内容为数据库预备知识。通过本章的学习，读者可对数据库的基本概念和原理有较为清晰的认识。需要读者重点掌握的内容有 SQL 语言的基本组成、数据库系统的基本结构、关系运算、数据模型等。

1.5 本章练习

一、单选题

1. 数据库(DB)、数据库系统(DBS)和数据库管理系统(DBMS)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
2. () 是位于用户与操作系统之间的数据管理软件。
A. DBMS B. DBS C. DBA D. DB
3. 以下说法错误的是()。
A. 数据库中的数据是没有冗余的
B. 数据库中的数据是长期存储的
C. 数据库中的数据是按照一定方式组织的
D. 数据库中的数据是可共享的
4. 在关系数据库中，一个属性是指()。
A. 一个二维表 B. 表中的一行
C. 表中的一列 D. 表中列的取值范围
5. 以下选项中，不属于数据模型的是()。
A. 关系模型 B. 网状模型 C. 层次模型 D. 网络模型
6. 在关系数据库中，一个关系是指()。
A. 一个二维表 B. 表中的一行 C. 表中的一列 D. 表中列的取值范围
7. 数据模型通常由()三要素构成。
A. 网络模型、关系模型、面向对象模型
B. 数据结构、网状模型、关系模型
C. 数据结构、数据操纵、关系模型
D. 数据结构、数据操纵、完整性约束

二、简答题

1. 简述数据库系统的概念，并举例说明。
2. 简述三种关系运算的特点。