

第 1 章 数据库概论



数据管理在信息系统的研究中一直都是一个非常重要的课题。数据库技术是数据管理的最新技术,是计算机科学技术中发展最快的领域之一。本章主要介绍数据库的理论基础知识,并重点介绍当前应用最广泛的关系模型和关系数据库标准语言(SQL)。



- 掌握数据、数据库、数据库管理系统和数据库系统的基本概念
- 了解数据库技术的发展历程
- 了解概念模型和常见的几种数据模型,掌握关系模型
- 掌握 SQL 语言的基本语法和主要语句

任务 1 数据库的基本概念



本任务主要介绍数据、数据库、数据库管理系统和数据库系统的基本概念。

阶段 1 数据

数据(Data)是数据库中存储的基本对象,也是数据库用户操作的对象。数据是描述事物的符号记录。文字、图像、声音、学生的档案记录、货物的运输情况等都是数据。在实际生活中,人们使用语言来描述事物。在计算机中,为了存储和使用这些事物,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,小明是计算机网络系的学生,今年 20 岁,性别男,学号 100001,我们就可以在计算机中这样记录小明的信息:100001,小明,男,20,计算机网络系。

阶段2 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。数据库中的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,并可以被各种用户共享。例如我们可以把学生的学号、姓名、课程和成绩的数据组成一个数据库,供学生和老师共享使用数据。

阶段3 数据库管理系统

数据库管理系统(Database Management System, DBMS)是为管理数据库而设计的软件系统,位于用户和操作系统之间。具有代表性的数据库管理系统有 Oracle、Microsoft SQL Server 和 My SQL 等。通常数据库管理员会使用数据库管理系统来建立数据库系统。

数据库管理系统的主要功能包括以下几个方面。

(1) 数据定义功能

DBMS 提供数据定义语言(Data Definition Language, DDL)。用户通过 DDL 可以方便地对数据库中的数据对象进行定义,同时可以定义数据库的三级结构、两级映象,定义数据的完整性约束、保密限制等约束。因此,在 DBMS 中应包括 DDL 的编译程序。

(2) 数据操纵功能

DBMS 提供数据操纵语言(Data Manipulation Language, DML)。用户可以通过 DML 实现对数据库中数据的基本操作,如查询、插入、删除和修改等。

(3) 数据库的保护功能

数据库中的数据是信息社会的战略资源,对数据的保护非常重要。DBMS 对数据库的保护主要通过数据库的恢复、数据库的并发控制、数据库的完整性控制、数据库的安全性控制 4 个方面实现。

(4) 数据库的存储管理

DBMS 的存储管理子系统提供了数据库中数据和应用程序的一个界面,其职责是把各种 DML 语句转换成底层的文件系统命令,起到数据的存储、检索和更新的作用。

(5) 数据库的维护功能

DBMS 中实现数据库维护功能的实用程序主要有数据加载程序、备份程序、文件重组程序和性能监控程序。

(6) 数据字典

数据库系统中存放三级结构定义的数据库称为数据字典(Data Dictionary, DD),对数据库的操作都要通过访问数据字典才能实现。

阶段4 数据库系统

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。一般由数据库、数据库管理系统(及其开发工具)、应用系统和数据库管理员构成。数据库系统结构

如图 1-1 所示。

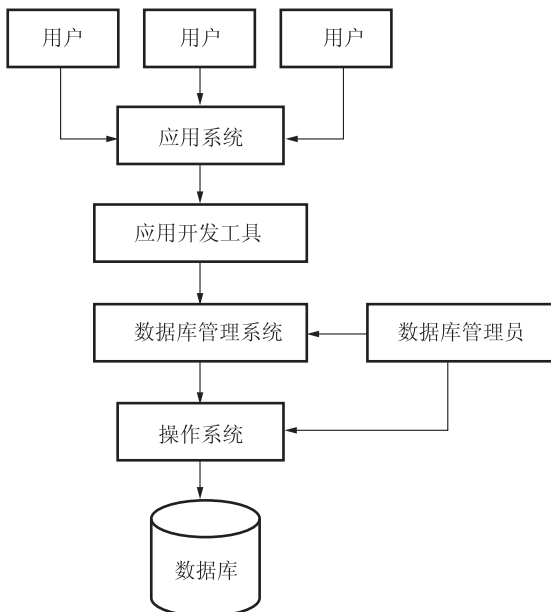


图 1-1 数据库系统

任务 2 数据库技术的发展历程



数据库技术的发展经历了人工管理阶段、文件系统阶段、数据库系统阶段 3 个阶段。本任务主要介绍数据库技术发展的这 3 个阶段。

阶段 1 人工管理阶段

20 世纪 50 年代中期以前为人工管理阶段,是计算机数据管理的初级阶段。这一阶段计算机主要用于科学计算,硬件中的外存只有卡片、纸带、磁带,没有磁盘等直接存取设备;软件只有汇编语言,没有操作系统,更无统一的管理数据的软件;对数据的管理完全在程序中进行,数据处理的方式基本上是批处理;程序员编写应用程序时,要考虑具体的数据物理存储细节,即每个应用程序中都要包括数据的存储结构、存取方法、输入方式、地址分配等。如果数据的类型、格式或输入/输出方式等逻辑结构或物理结构发生变化,则必须对应用程序作相应的修改。另外,数据是面向程序的,一组数据只能对应一个程序,很难实现多个应用程序共享数据资源。程序之间有大量的冗余数据。因此,这个阶段数据管理的特点体现在

以下4个方面。

- (1)数据不保存在计算机中。需要时把数据输入,用完就撤走。
- (2)没有软件系统管理数据。数据和程序都不具有独立性,程序不仅要规定数据的逻辑结构,还要设计其物理结构、存取方法、输入/输出方式。
- (3)数据的组织方式必须由程序员自行设计。
- (4)数据是面向应用的。一组数据只对应于一个程序,即使两个应用程序都涉及某些相同数据,也必须各自定义,无法互相利用。不仅在程序之间有大量重复数据,还易产生数据的不一致性。

人工管理阶段的特点如图1-2所示。

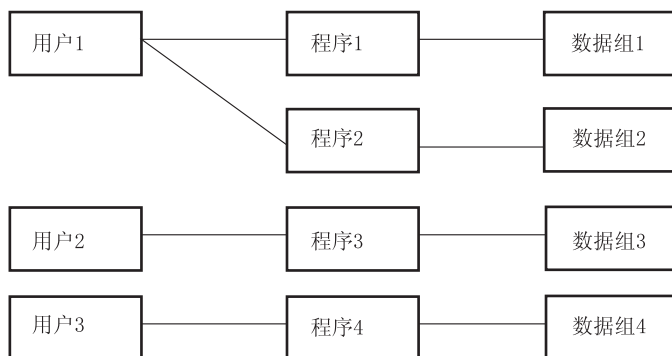


图 1-2 人工管理阶段

阶段2 文件系统阶段

20世纪50年代后期至60年代中期为文件系统阶段,计算机应用领域开始拓宽。不仅用于科学计算,还大量用于数据管理。硬件方面出现了可以直接存取的外部存储设备如磁盘、磁鼓等。软件方面有了操作系统中专门管理数据的文件系统。数据的管理是以独立的数据文件形式存放,并可按记录存取。在该阶段,一个应用程序可以处理多个数据文件,文件系统程序与数据间起到接口的作用,使程序和数据有了一定的独立性,如图1-3所示。但文件系统的致命缺陷是不能有效地共享相同的数据,从而造成了数据的冗余度大且不一致,数据独立性差,同时给数据的修改和维护带来了困难。

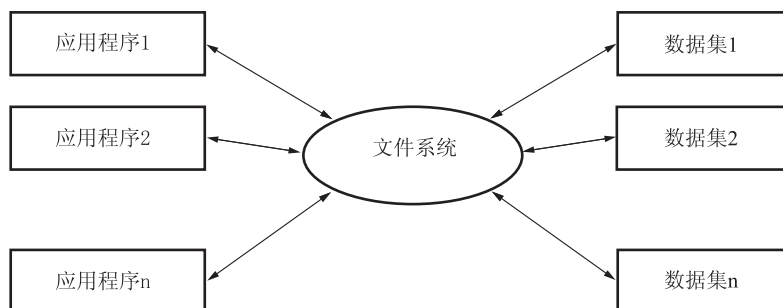


图 1-3 文件系统阶段程序与数据之间的联系

阶段3 数据库系统阶段

20世纪60年代后期以来,计算机用于管理的规模更为庞大,应用越来越广泛,数据量急剧增长,同时多种应用、多种语言互相覆盖共享数据集合的要求越来越强烈。这时硬件有大容量磁盘,硬件价格下降,软件价格上升,为编制和维护系统软件及应用程序所需的成本相对增加。在处理方式上,联机实时处理要求更多,并开始提出和考虑分布处理。在这种背景下,以文件系统作为数据管理手段已经不能满足应用的需求,于是为解决多用户、多应用共享数据的需求,使数据尽可能多地为应用服务,出现了数据库技术和统一管理数据的专门软件系统——数据库管理系统。

数据库技术从20世纪60年代中期产生到现在仅仅30余年的历史,但其发展速度之快,使用范围之广是其他技术所不及的。20世纪60年代末出现了第一代数据库——网状数据库、层次数据库,20世纪70年代出现了第二代数据库——关系数据库。目前关系数据库系统已逐渐淘汰了网状数据库和层次数据库,成为当今最流行的商用数据库系统。

任务3 概念模型和数据模型



计算机不能直接处理现实世界中的具体事物,所以人们需要把具体事物抽象并转换成计算机能够处理的数据。模型是现实世界特征的模拟和抽象。本任务主要介绍概念模型和数据模型。

阶段1 概念模型

当前常用的概念数据模型是实体联系模型(Entity Relationship Model, E-R模型)。

E-R模型,是P. P. Chen于1976年提出的。该模型是用E-R图来描述概念模型的一种常用的表示方法。E-R模型的基本语义单位是实体与联系,它可以形象地用图形表示实体及其关系。E-R模型有3要素:实体、属性、实体间的联系。

1. 实体

客观存在并可相互区别的事物称为实体(Entity)。例如学生、一次选课等都是实体。实体表示的是一类事物,其中的一个具体事物称为该实体的一个实例,例如C01,“数据库原理”,“4”则具体表示了一门课程,它是课程实体的一个实例。在E-R模型中,用矩形框表示实体,框内标注实体名称。

2. 属性

实体所具有的某一特性称为属性(Attribute)。实体可以具有若干个属性,例如,学生实体有学号、姓名、性别、年龄等属性。在E-R模型中,用椭圆形表示属性,并用连线与实体或联系连接起来。

3. 实体间的联系

实体不是孤立的,实体与实体之间有着密切的联系。实体间的联系分为一对一联系(1:1)、一对多联系(1:n)、多对多联系(m:n)3种类型。在E-R模型中,用菱形框表示联系,框内标注联系名称,并用连线将菱形框分别与有关实体相连。

(1) 一对一联系(1:1)。如果实体集A和B中的每一个实体至多和另一个实体集中的一个实体有联系,那么实体集A和B的联系称为一对一联系,记为1:1。例如,班长和班级之间。

(2) 一对多联系(1:n)。若实体集A中每个实体与实体集B中多个任意实体($n \geq 0$)有联系,而实体集B中每个实体至多与实体集A中一个实体有联系,那么称从A到B是一对多联系,记为1:n。例如,班级与学生之间。

(3) 多对多联系(m:n)。若实体集A和实体集B中允许每个实体都和另外一个实体集中多个任意实体有联系,那么称A和B为多对多联系,记为m:n。例如,学生和课程之间。

图1-4描述了两个实体之间的3种联系(1:1,1:n,m:n)。

E-R模型有两个明显的优点:接近于人的思维,容易理解;与计算机无关,用户容易接受。因此,E-R模型已成为软件工程的一个重要设计方法。

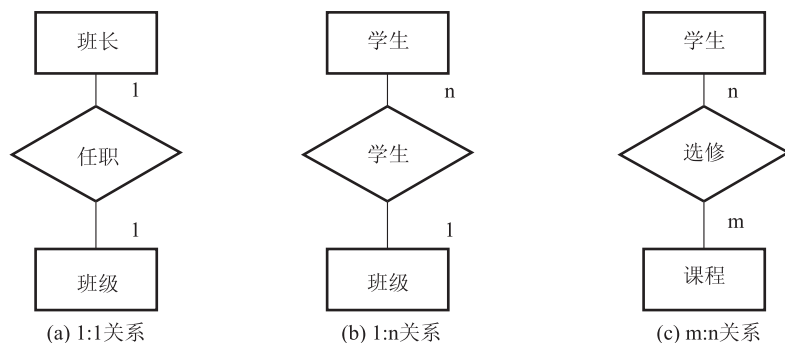


图1-4 实体联系图

【例1-1】 用E-R图来表示一个部门管理的概念模型。某部门的劳保用品管理涉及的实体有:职工,部门,工种,劳保用品。其中各实体的属性和联系如下。

各实体间的属性如下。

- (1) 职工的属性有:职工号、姓名、年龄、性别、部门名;
- (2) 部门的属性有:部门号、部门名、电话;

(3) 工种的属性有:工种名;

(4) 劳保用品的属性有:用品名、库存量、价格。

各实体间的联系如下。

(1) 职工和部门:一个部门有多个职工,而一个职工仅属于一个部门。所以部门和职工具有一对多的联系。

(2) 职工和工种:一个工种有多个职工,而一个职工仅属于一个工种。所以工种和职工也具有一对多的联系。

(3) 工种和劳保用品:一种劳保用品可以发给若干工种的职工,而一个工种的职工可以得到多种劳保用品,所以工种和劳保用品具有多对多的联系。

作 E-R 图的步骤如下。

(1) 确定实体和实体的属性,如图 1-5 所示。

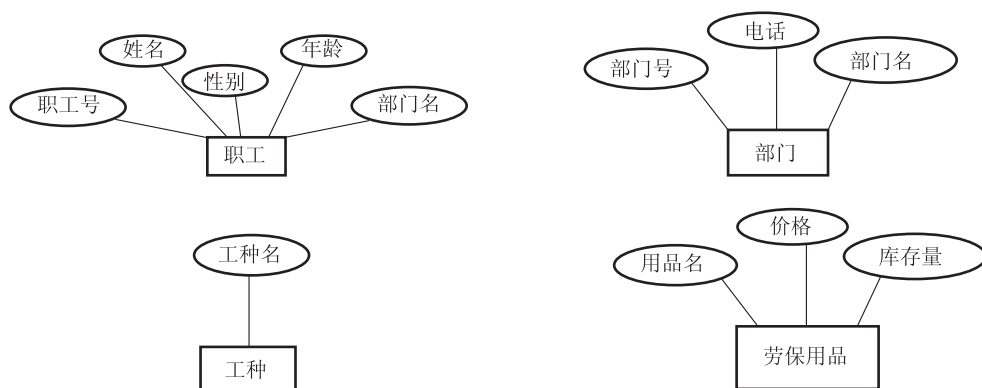


图 1-5 实体及其属性图

(2) 确定实体之间的联系及联系的类型,如图 1-6 所示。

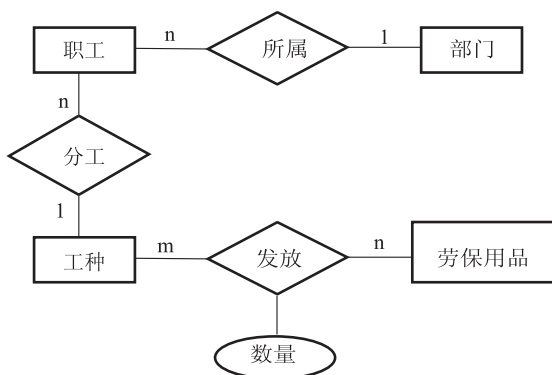


图 1-6 实体及其联系图

(3) 给实体和联系加上属性,如图 1-7 所示。

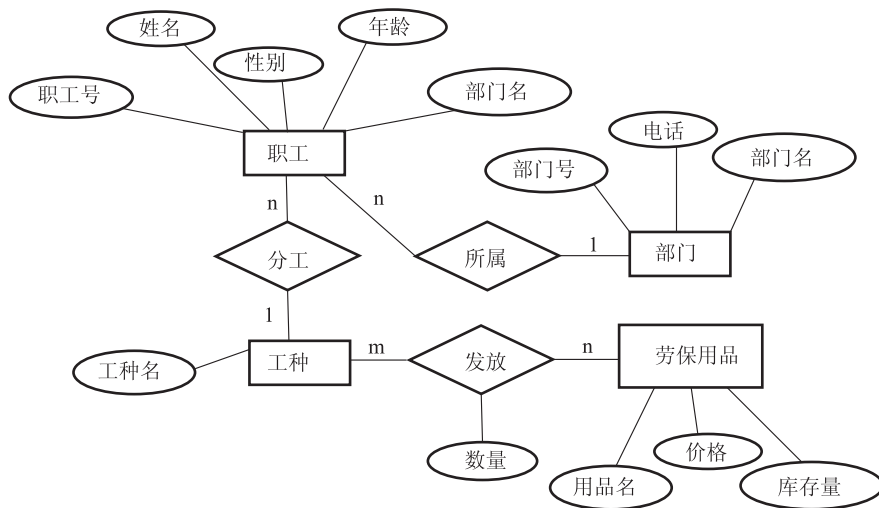


图 1-7 完整的实体联系图

阶段 2 数据模型

数据模型是数据库中用来对现实世界进行抽象的工具,是数据库中用于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,每个 DBMS 软件都是基于某种数据模型的。

数据模型有 3 个要素:数据结构、数据操作和完整性规则。数据结构用于描述系统的静态特性,人们通常按照其数据结构的类型来命名数据模型。数据操作用于描述数据的动态特征。完整性规则是给定的数据模型中数据及其联系所具有的制约和储存规则,用以限定符合数据模型的数据库状态以及状态的变化。

目前,数据库领域采用的数据模型有层次模型、网状模型和关系模型。其中应用最广泛的是关系模型。

1. 层次模型

用树型结构表示实体之间联系的模型称为层次模型。层次模型是最早用于商品数据库管理系统的数据库模型。其典型代表是于 1969 年问世、由 IBM 公司开发的数据库管理系统 IMS(Information Management System)。在现实世界中,许多实体集之间的联系就是一个自然的层次关系。例如,行政机构、家族关系等都是层次关系。图 1-8 就是学校中系的层次模型。

层次模型的代表方法是:树的节点表示实体集(记录类型),节点之间的连线表示相连两实体集之间的关系,这种关系只能是 1: N 的。通常把表示 1 的实体集放在上方,称为父节点,表示 N 的实体集放在下方,称为子节点。层次模型的结构特点是:

(1) 有且仅有一个根节点;

(2)根节点以外的其他节点有且仅有一个父节点。

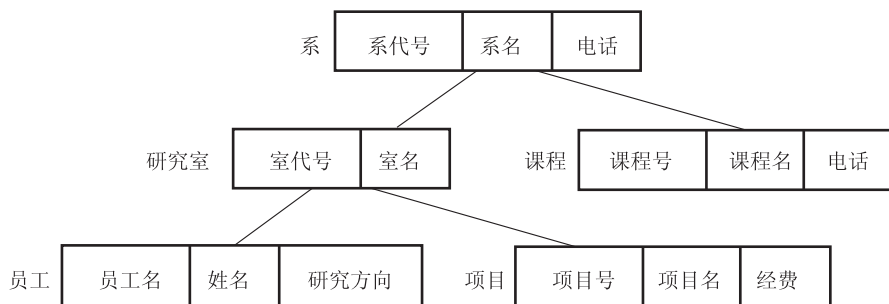


图 1-8 层次模型示例

因而层次模型只能表示 1: N 关系,而不能直接表示 M: N 关系。

在层次模型中,一个节点称为一个记录类型,用来描述实体集。每个记录类型可以有一个或多个记录值,上层一个记录值对应下层一个或多个记录值,而下层每个记录值只能对应上层一个记录值。

2. 网状模型

网状模型的典型代表是 DBTG(Database Task Group)数据模型。数据库任务小组 DBTG 是美国 CODASYL(Conference of Dam System Language)组织的下属机构。它于 1969 年提出 DBTG 报告。DBTG 报告是一个网状模型的数据描述语言和数据操纵语言的规范文本。

网状模型(Network Model)是一种更具有普遍性的结构。从图论的角度讲,网状模型是一个不加任何条件限制的无向图。

网状模型是以记录为节点的网状结构,它满足以下条件:

- (1)可以有任意个节点无双亲;
- (2)允许节点有一个以上的双亲;
- (3)允许两个节点之间有一种或两种以上的联系。

在网状模型的 DBTG(Database Task Group)标准中,基本结构是简单二级树称为系,系的基本数据单位是记录,它相当于 E-R 模型中的实体集。记录又有若干数据项组成,它相当于 E-R 模型中的属性。图 1-9 为教师授课数据库的网状数据库模式。

网状模型明显优于层次模型。网络模型在一定程度上支持数据的重构,具有一定的数据独立性和共享特性,并且运行效率较高。但它应用时存在以下问题。

- (1)网状结构的复杂,增加了用户查询和定位的困难。它要求用户熟悉数据的逻辑结构,知道自身所处的位置。
- (2)网状数据操作命令具有过程式性质。
- (3)不直接支持对于层次结构的表达。

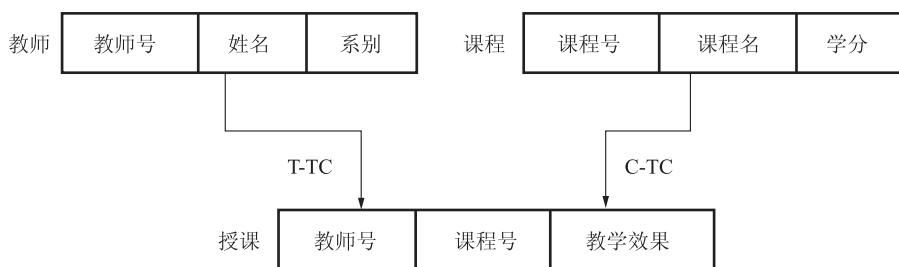


图 1-9 网状模型示例

3. 关系模型

关系模型是目前最常用的一种数据模型。关系数据库系统采用关系模型作为数据的组织方式。1970年美国IBM公司San Jose研究室的研究员E. F. Codd首次提出了数据库系统的关系模型,开创了数据库关系方法和关系数据理论的研究,为数据库技术奠定了理论基础。现在流行的数据库系统大都是基于关系模型的关系数据库系统。我们要学习的Microsoft SQL Server就是典型的关系数据库。

(1) 关系模型的数据结构

关系模型建立在严格的数学概念的基础上。一个关系模型的逻辑结构是一张二维表,它由行和列组成。在关系模型中,实体以及实体间的联系都用关系来表示。关系模型要求关系必须是规范化的。最基本的条件是关系的每一个分量必须是一个不可分的数据项,即不允许表中还有表。例如,表1-1中的学生学籍表是一个关系模式,它涉及以下概念。

表 1-1 学生学籍表

学号	姓名	性别	年龄	所在系
S00001	王立	男	20	计算机系
S00002	莉莉	女	19	数学系
⋮	⋮	⋮	⋮	⋮
S00005	张晓潇	男	21	英语系

- 关系:对应通常所说的一张二维表。如表1-1就是一个关系。
- 元组:表中的一行称为一个元组,许多系统中把元组称为记录。如表1-1中有5行,即有5个元组,或5条记录。
- 属性:表中的一列称为一个属性。一个表中往往会有多个属性,为了区分属性,要给每一个列起一个属性名。同一个表中的属性应具有不同的属性名。如表1-1中有5列,对应5个属性(学号,姓名,性别,年龄,所在系)。
- 域:属性的取值范围称为域。例如,大学生的年龄属性的域是(16~40),性别的域是(男,女)。
- 候选键与主键:能唯一标识关系中元组的一个属性或属性集,称为候选键或关键字。

若一个关系有多个候选键,则选定其中一个作为主键。如表 1-1 中学号是候选键,同时也是主键。

(2) 关系模型的操作与完整性约束

关系模型的操作主要包括查询、插入、删除和更新数据。这些操作必须满足关系的完整性约束条件,即是对关系的某种约束条件。关系的完整性约束条件包括 3 类:实体完整性、参照完整性和用户定义完整性。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,应该由关系系统自动支持。

关系模型中的数据操作是集合操作。操作对象和操作结果都是关系,即若干记录的集合。

(3) 关系模型的优缺点

其优点包括以下两个方面。

第一,在关系模型中,无论实体还是实体之间的联系都用关系来表示。数据的查询结果也是关系(表)。因此,概念单一、数据结构简单、清晰。

第二,关系模型的存取路径对用户透明,从而具有更高的数据独立性、更好的安全保密性,也简化了程序员的工作和数据库的开发工作。

缺点:由于存取路径对用户透明,查询效率往往不如非关系数据模型。因此,为了提高性能,必须对用户的查询请求进行优化,从而增加了另外开发数据库管理系统的负担。

任务 4 关系数据库标准语言 SQL



使用数据库就需要使用一种语言对数据库进行各种各样的操作。本任务主要介绍在 SQL Server 中使用的 T-SQL 语言。

阶段 1 T-SQL 语言简介

SQL 语言是用于数据库查询的结构化语言。最早由 Boyce 和 Chamberlin 在 1974 年提出,称为 SEQUEL 语言。1976 年,IBM 公司的 San Jose 研究所在研制关系数据库管理系统 SystemR 时修改为 SEQUEL2,即目前的 SQL 语言。此后,SQL 开始在商品化关系数据库管理系统中应用。1982 年美国标准化组织 ANSI 确认 SQL 为数据库系统的工业标准。现在许多关系型数据库供应商都在自己的数据库中支持 SQL 语言,如: Access, Oracle, DB2 等。目前,最新的 SQL 语言是 ANSI SQL-99。

Transact-SQL,即 T-SQL,是微软公司在 SQL Server 数据库管理系统中 ANSI SQL-99 的实

现。在 SQL Server 数据库中,T-SQL 语言由以下几部分组成。

(1)数据定义语言(DDL)。用于执行数据库的任务,对数据库以及数据库中的各种对象进行创建、删除、修改等操作。DDL 的主要语句及功能如表 1-2 所示。

表 1-2 DDL 的主要语句及功能

语句	功能	说明
CREATE	创建数据库或数据库对象	不同数据库对象,其 CREATE 语句的语法形式不同
ALTER	对数据库或数据库对象进行修改	不同数据库对象,其 ALTER 语句的语法形式不同
DROP	删除数据库或数据库对象	不同数据库对象,其 DROP 语句的语法形式不同

(2)数据操纵语言(DML)。用于操纵数据库中各种对象,检索和修改数据。DML 包括的语句及功能如表 1-3 所示。

表 1-3 DML 的主要语句及功能

语句	功能	说明
SELECT	从表或视图中检索数据	它是使用最频繁的 SQL 语句之一
INSERT	将数据插入到表或视图中	
UPDATE	修改表或视图中的数据	既可修改表或视图的一行数据,也可修改一组或全部数据
DELETE	从表或视图中删除数据	可根据条件删除指定的数据

(3)数据控制语言(DCL)。用于安全管理,确定哪些用户可以查看或修改数据库中的数据,DCL 包括的语句及功能如表 1-4 所示。

表 1-4 DCL 的主要语句及功能

语句	功能	说明
GRANT	授予权限	可把语句许可或对象许可的权限授予其他用户和角色
REVOKE	收回权限	与 GRANT 的功能相反,但不影响该用户或角色从其他角色中继承许可权限
DENY	拒绝权限	与 REVOKE 不同之处:除收回权限外,还禁止从其他角色继承许可权限

(4)T-SQL 增加的语言元素。这部分不是 ANSI SQL-99 所包含的内容,而是微软为了用户编程的方便增加的语言元素。这些语言元素包括变量、运算符、函数、流程控制语句和注释。这些 T-SQL 语句都可以在查询分析器中交互执行。

阶段 2 标识符

SQL 标识符用于标识数据库对象,如表、视图、列、索引、触发器、过程、约束、规则等。一般应该定义有意义的字符序列。用户定义 SQL 标识符时必须遵守以下规则。

(1)标识符最多可以包含 128 个字符,包括字母、数字、下划线(_)、at 符号(@)、数字符

号(#),以及美元符号(\$)。

- (2) 第一个字符可以是字母、下划线(_)、at 符号(@),或者数字符号(#)。
- (3) 不允许嵌入空格或其他特殊字符。
- (4) 不允许使用 T-SQL 的保留字,包括大写和小写形式。
- (5) 中文版的 SQL Server 允许使用汉字作为标识符。

阶段3 语法格式约定

表 1-5 列出了 Transact-SQL 参考的语法关系图中使用的约定。

表 1-5 语法约定表

约定	用于
(竖线)	分隔括号或大括号中的语法项,只能使用其中一项
[] (方括号)	可选语法项,不要键入方括号
{ } (大括号)	必选语法项,不要键入大括号
[, ..., n]	指示前面的项可以重复 n 次,各项之间以逗号分隔
[... n]	指示前面的项可以重复 n 次,每一项由空格分隔
< label > :: =	语法块的名称,此约定用于对可在语句中的多个位置使用的过长语法段或语法单元进行分组和标记;可使用的语法块的每个位置由括在尖括号内的标签指示: < label >

本章小结

本章主要介绍了数据库的基本概念、数据库技术的发展历程、概念模型和数据模型以及关系数据库查询语言 SQL。重点介绍了当前应用最广泛的关系模型和关系数据库查询语言 SQL。

数据是数据库中存储的基本对象,也是数据库用户操作的对象。数据库是指长期存储在计算机内,有组织的、可共享的数据集合。数据库管理系统是为管理数据库而设计的软件系统,位于用户和操作系统之间。数据库系统是指在计算机系统中引入数据库后的系统,一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员构成。

数据库技术的发展经历了人工管理阶段、文件系统阶段、数据库系统阶段 3 个阶段。

根据模型应用的目的,将数据模型分为两种类型:概念模型和数据模型。最常用的概念模型是实体—联系模型(E-R 模型)。常见的数据模型有层次模型、网状模型和关系模型。其中关系模型应用最广泛。

结构化查询语言 SQL 是关系数据库标准语言。T-SQL 是微软公司在 SQL Server 数据库管理系统中 SQL 的实现。T-SQL 语言包括数据定义语言(DDL)、数据操纵语言(DML)、数据控制语言(DCL)和 T-SQL 增加的语言元素。

本章习题

1. 简述数据、数据库、数据库管理系统和数据库系统的概念。
2. 什么是 E-R 模型, E-R 模型有什么作用?
3. 什么是关系模型? 关系模型具有什么优点?
4. 数据库技术经历了哪 3 个阶段? 简述每一个阶段的特点。
5. 某学校有若干个系, 每个系有若干名教师和学生; 每个教师可以教授若干门课程, 参加若干个项目; 每个学生可以同时选择若干门课程。试设计该学校的教学管理 E-R 模型, 并给出每个实体、联系的属性。