

数据库原理

苏一丹 刘智斌 刘玉萍 编著

重庆大学出版社

内 容 提 要

本书系统、全面地介绍了数据库系统的基本理论、技术及其方法;讲述了数据库系统的基本概念、数据库系统的性质和目标,关系数据模型、关系数据理论,SQL语言,数据库的完整性约束及其在数据库设计中的应用;讨论了数据库系统的保护技术,数据库的恢复技术、并发控制、完整性及安全性;介绍了数据库系统新技术,集中展示数据库研究中的新进展和新的数据库应用,包括面向对象的数据库系统、分布式数据库系统、并行数据库系统等。每章后均附有习题。

本书可作为高等院校计算机科学与技术、信息管理与信息系统、自动化、信息工程、应用数学、系统工程专业本、专科教材,也可作为有关工程技术人员学习数据库系统原理的参考书。

图书在版编目(CIP)数据

数据库原理/苏一丹,刘智武,刘玉萍编著.—重庆:
重庆大学出版社,2002.4

计算机科学与技术专业本科系列教材

ISBN 7-5624-2341-5

I.数.. II.①苏...②刘...③刘... III.数据库系统—
理论—高等学校—教材 IV.TP311.13

中国版本图书馆 CIP 数据核字(2001)第 094736 号

数据库原理

苏一丹 刘智斌 刘玉萍 编著

责任编辑 梁 涛

*

重庆大学出版社出版发行

新华书店经销

重庆大学建大印刷厂印刷

*

开本 787×1092 1/16 印张·12 5 字数·312 千字

2002 年 4 月第 1 版 2002 年 4 月第 1 次印刷

印数 1—6 000

ISBN 7-5624-2341-5/TP·297 定价·18.00 元

前言

数据库技术是数据管理的技术,是计算机科学的重要组成部分。近年来,数据库技术发展迅速,各种应用领域对数据管理的需求越来越多,数据库技术的重要性越来越为人们所熟知。目前,不仅在大、中、小、微各种机型上都配有数据库管理系统,各行各业的信息系统以及互联网上的信息系统都离不开数据库的支持。因此,数据库已成为信息社会的重要基础设施。而 Internet 的发展及各种应用技术的交叉与融合,又给数据库提供了更多的机遇。

本书是作者在多年从事数据库科研和教学基础上,进行归纳、总结、提高,并参考有关文献编写而成。全书共分 7 章:第 1 章介绍了数据库的基本概念;第 2 章介绍了数据模型,包括网状模型、层次模型和关系模型、实体联系模型、关系代数与关系演算;第 3 章介绍了数据库标准语言,包括 SQL 语言的基本概念、数据定义、查询、数据更新、视图、数据控制、嵌入式 SQL;第 4 章介绍了数据依赖和关系模式的规范化理论,包括关系模式设计中的数据语义问题、函数依赖、关系范式、关系模式的分解特性、联接依赖(JD)与投影联接范式(PJNF);第 5 章介绍了数据库应用系统设计的全过程,包括数据库设计的步骤、需求分析、概念结构设计、物理设计、数据库实施及数据库的运行与维护;第 6 章介绍了数据库系统的保护技术,包括恢复、并发控制、完整性和安全性等;第 7 章介绍了数据库的研究与发展。

本书由苏一丹、刘智斌、刘玉萍编著。第 1 章、第 2 章、第 7 章由苏一丹编著,第 4 章、第 6 章由刘智斌编著,第 3 章、第 5 章由刘玉萍编著。全书由苏一丹统稿。

本书可作为高等院校计算机科学与技术、信息管理与信息系统、自动化、信息工程、应用数学、系统工程专业本、专科教材,也可作为有关工程技术人员学习数据库系统原理的参考书。

由于编著者水平有限,书中可能存在疏漏和不当之处,恳请专家及广大读者不吝赐教。

编著者

2001 年 10 月

目录

第 1 章 数据库系统概述	1
1.1 引言	1
1.2 计算机数据处理和数据库系统的发展	1
1.3 数据库的定义	2
1.4 数据库中数据的特性	4
1.5 数据库管理系统	4
1.6 数据库系统的组成	6
1.7 使用数据库管理系统的原因	8
1.8 数据库管理系统的应用领域	8
1.9 数据库技术的新进展	9
习题	13
第 2 章 数据模型	14
2.1 数据模型的定义	14
2.2 实体-联系(E-R)模型	14
2.3 层次数据模型	19
2.4 网状数据模型	22
2.5 关系数据模型	24
习题	34
第 3 章 关系数据库标准语言 SQL 语言	36
3.1 SQL 概述	36
3.2 数据定义	39
3.3 查询	41
3.4 数据更新	57
3.5 视图	59
3.6 数据控制	64
3.7 嵌入式 SQL	66
习题	75

第 4 章 数据依赖和关系模式的规范化	77
4.1 关系模式设计中的一些数据语义问题	77
4.2 函数依赖(Functional dependency)	79
4.3 关系模式的范式	85
4.4 关系模式的分解特性	90
4.5 联接依赖(JD)与投影联接范式(PJNF)	98
习题	99
第 5 章 数据库设计	101
5.1 数据库设计的步骤	101
5.2 需求分析	103
5.3 概念结构设计	106
5.4 逻辑结构设计	116
5.5 数据库物理设计	121
5.6 数据库实施	124
5.7 数据库的运行与维护	126
习题	127
第 6 章 数据库保护	128
6.1 数据库的恢复	128
6.2 数据库的并发控制	137
6.3 数据库的完整性	148
6.4 数据库的安全	156
习题	160
第 7 章 数据库技术的新进展	163
7.1 数据库技术的三个发展阶段	163
7.2 分布式数据库系统和并行数据库系统	170
7.3 面向对象的数据库系统	180
7.4 数据仓库及数据挖掘技术	186
7.5 数据库系统的研究与发展趋势	190
习题	191
主要参考文献	192

第 1 章

数据库系统概述

1.1 引 言

数据库技术产生于20世纪60年代,是数据管理的最新技术,是计算机科学的重要分支。在计算机的三大应用(科学计算、数据处理与过程控制)中,数据处理所占比重为70%左右。数据库技术就是作为数据处理中一门最重要的技术而发展起来的。随着各种应用领域对数据管理的进一步需求,数据库技术的重要性已愈来愈为人们所熟知。目前,不仅大、中、小、微各种机型上都配有数据库管理系统,各行各业的信息系统,包括因特网上的信息系统也大都离不开数据库的支持。因此,数据库已成为信息社会的重要基础设施。

在本书中,我们将研究有关数据库的各种问题:数据库的基本理论,如何逻辑表示数据库所管理的信息,数据库如何管理存储的数据,如何正确有效地保护数据和维护数据,数据库管理系统的发展过程,以及数据库如何随着对其要求的提高而继续发展。

1.2 计算机数据处理和数据库系统的发展

数据库管理系统的发展紧跟计算机及其操作系统的发展。随着ENIAC计算机系统的开发,在20世纪40年代末出现了计算机。早期的ENIAC及其后续计算机由真空管构成,其体积非常庞大,足以填满一个大房间。这些早期的计算机系统只能用来完成单项任务,它们还没有操作系统或数据库管理系统。

这些早期计算机的功能还比不上今天用的手掌大的计算器。这些计算机主要用于弹道导弹发射轨道的计算和军事研究。这些早期计算机的体系结构基于冯·诺伊曼存储程序式的单指令流系统结构(见图1.1)。当今大多数计算机系统仍在使用这个基本体系结构和原理。

这些早期的计算机系统没有数据库和复杂的操作系统来简化其操作。它们将程序指令和计算所需的数据存储在同一个地方。计算机每次从内存读取一条指令,大多数指令都与将程序数据从内存装载或存储到寄存器中有关,数据操作在寄存器上进行。早期计算机中的这些数据不被多个程序共享。如果一个程序所需要的数据是由另一个程序生成的,这些数据通常是被拷贝到程序空间末端附近的区域,为让嵌入这些数据的应用程序使用它们,末端地址被硬编码(hard-coded)。

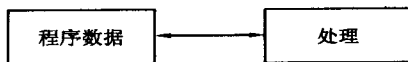


图1.1 冯·诺伊曼编程模式

随着对共享数据需求的增长,数据库系统诞生和发展了。操作系统的发展允许多个进程共享计算机资源,这促进了数据库系统的发展(见图1.2)。多进程操作系统的出现,使得程序可共享信息。在早期的系统中,数据要通过操作系统共享,或明确地通过程序员指令程序以后使用。但这样做的问题是无法控制数据在程序间传递。

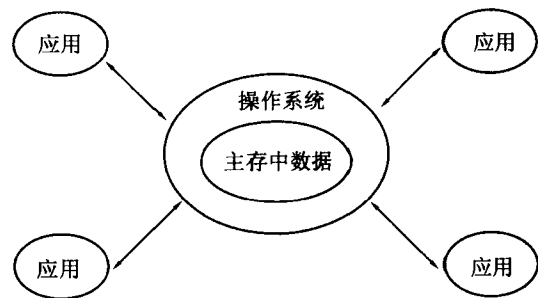


图1.2 早期的数据访问控制

随着这些系统中所出现的问题的逐渐积累,以及对稳定数据存储的需求增长,用户需要一种通用的方案——通过某种操作系统服务来解决内存和数据的共享。内存管理系统和文件管理系统的发展解决了内存共享的问题。在有多个程序需要共享数据时,这两个系统可协同工作,存取特定的信息,并将信息从一处移动到它处(见图1.3)。另外,由于存储数据的控制从程序员直接控制转为操作系统控制,其他数据安全措施可被用于数据的管理。文件服务器发展了数据存取机制,能检测出一个程序是否有权存取共享的数据,程序是否只有读权限或写权限,或者程序是否有改变数据基本结构的授权。系统开发人员和用户看到了这种技术的优点,其使用被大大地推广。用户和程序员需要从文件系统获得更多服务,这些需求促进了基本网络或CODASYL数据库语言的早期数据库系统的发展。早期的数据库系统提供了对存取的控制、构造和定义数据库的能力,以及保证数据库使用的正确性和一致性的能力。

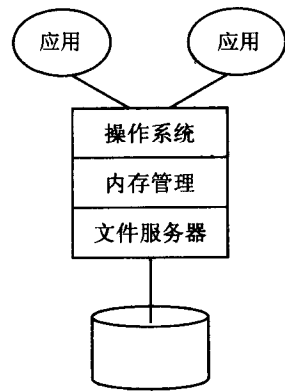


图1.3 计算机存储的层次结构

1.3 数据库的定义

定义数据库管理系统之前,必须首先定这种系统的基本成分,数据库或数据存储仓库。

一个数据库有四个主要成分:数据、联系(Relationship)、约束(Constraint)和模式(Schema)(见图1.4)。数据是用所存储的逻辑实体在计算机中的二进制表示;联系表示数据项之间的某

种对应;约束是定义正确数据状态的规则;模式描述数据库中数据的组织和联系。

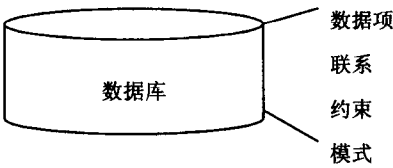


图14 数据库组成

模式为数据库管理系统各部件的使用和应用的安全定义数据库的各种视图。一种模式将数据存储的物理表示与逻辑表示分开(见图15)。内部模式定义数据在物理数据存储区中如何组织以及放在何处。概念模式模型按照适当的数据库数据模型(如关系模型或对象模型)定义所存储数据的结构。外部模式为特定用户(们)定义数据库的一个或多个视图。一个数据模型是数据库的一种抽象定义或模型,它提供一种描述数据的表示法。

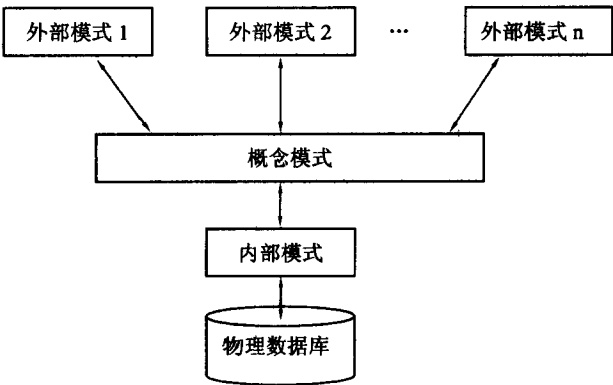


图15 数据库模式的概念

数据库管理系统为访问数据库提供服务,同时维护存储数据所要求的正确性和一致性(见图1.6)。

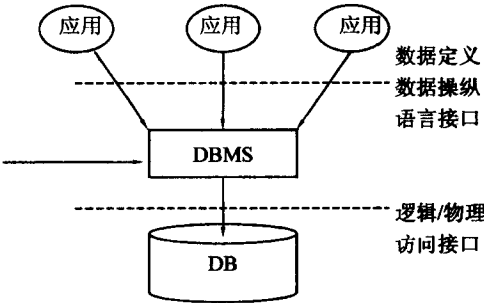


图16 数据库管理系统的服务

在数据库管理系统中,运行的工作单元是事务(Transaction)单元,在此之上定义了一致性和正确性。事务应该支持ACID属性。ACID属性包括:事务运行的原子性(Atomic)、一致性(Consistent)、独立性(Isolation)和事务执行的持久性(Durability)。原子性确保事务被当成一个不可分割(all-or-nothing)的操作单元处理。事务操作的一致性确保数据库从原来的一致性状态

正确转移到一个新的一致性状态,此处的一致性由数据库中数据项上的谓词定义。事务的独立属性定义了允许它们可以见到什么。一个被孤立的事务只“看到”数据库的事务,其结果就持久存在且不能从数据库中消除。

1.4 数据库中数据的特性

数据库中数据必须具有下列的各种特性:

- 1)共享:数据库中的数据可在几个用户和应用程序间共享。
- 2)持续:数据库中的数据持久存在,即数据项可在创建它的进程作用范围外存在。
- 3)安全:避免数据库中数据不授权的泄露、更改或破坏。数据库系统管理员根据各个用户或用户组成员的用途和级别,决定数据库合法访问的判断准则。
- 4)有效:也指数据完整性或正确性,数据库中的数据应正确表示它们所代表的现实世界实体。
- 5)一致:当用多个数据库中的数据表示相关现实世界的某些值时,数据的取值应与它们所对应的联系相一致。
- 6)不冗余:数据库中不存在两个数据项表示同一个现实世界实体。
- 7)独立:在三级模式模型中的三个层次应相互独立,某一层模式的改变不会影响其他两层。物理数据独立性是指改变内部模式时不用改变概念或外部模式。因此,访问数据的程序与数据存在何处以及如何放都无关,即数据库物理存储的变动不会影响访问数据的应用程序。逻辑数据独立性是指根据子概念模式的修改不需要改变外部模式(或者可能只需重定义映射)。当然,如果修改涉及到数据的删除(不管在哪一层),则其他层次都会受到影响,不过这种情况可加以控制。

1.5 数据库管理系统

数据库管理系统(DBMS)是为数据库访问提供服务的软件(见图1.7),同时维护所有数据必需的特性。数据库管理系统为支持应用程序访问和操作数据库数据提供下列服务:

1)事务处理:一个事务是数据库操作的编序(partially ordered)系列,它代表一个逻辑工作单元和访问共享数据库的单元。事务将使数据库从一个一致状态转移到另一个一致状态。数据库操作被分成两大类:数据访问操作和事务操作。有三种特定的事务操作:启动(start),指示将开始一个新事务;提交(commit),指示事务已正常终止且其作用结果将持久存在;以及放弃(abort),指示事务被异常终止,其所有结果将被放弃。

事务通常需要具有前面提到过的ACID属性。

一个调度或历史是指若干事务中的操作执行的编序,其中不同事务的执行相互关联。

事务处理包括将出自不同来源(用户、应用程序等)的事务操作作用到数据库中,同时还要维护事务所需的属性。作为事务处理的一部分,对资源(如CPU)和数据项的访问被调度,以满足事务的需求。以下两个服务——并发控制和恢复,与事务处理紧密相关,因为它们有助于维护事务的ACID属性。

2)并发控制:并发控制是一种数据库管理活动,它协调数据库操作进程的并发操作和对共

享数据的访问,并且解决它们之间可能发生的潜在冲突。并发控制机制的目标是允许并发维护共享数据的一致性。数据库系统中的并发单元是事务。

数据库系统可能允许或要求不同级别的并发性。大多数技术沿用某个正确性指标,由此指标决定事务间所允许的最大并发数。并发的数量由并发事务间的交错程度和并发运行的事务数定义。最简单的正确性指标是互斥(mutual exclusion),它要求每个事务的运行期间不能被另一个事务中断。这个运行指标产生顺序调度——即一个事务的调度中所有操作即这些技术产生串行调度。一个调度可被串行化,是指按照该调度进行事务处理,产生的输出和对数据库的影响与一个顺序调度一样。但是,不太严格地说,为了增加并发性,我们建议正确性标准应谨慎地放松串行化的要求。

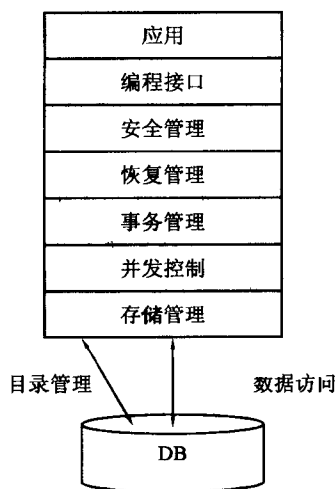


图17 数据库系统的管理部件

3)恢复:数据库中恢复的目标是确保异常终止或出错的事务不会对数据库或其他事务产生不利影响。异常终止的事务有两种影响:对数据的影响和对其他事务的影响。恢复可使得数据库在事务异常终止后返回某个一致状态。如果第二个事务读取被第一个事务修改的数据,它们就会被异常终止。由于允许越多的开发,一个异常终止的事务就越有可能影响其他事务,因此恢复与并发控制密切相关。

4)安全:安全是保护数据免受非授权的泄露、更改或破坏。每个用户和应用程序都有特定的数据访问特权。这些特权可以由外部模式定义,即根据各个用户被允许访问和/或修改的数据,给予它们不同的数据视图。安全系统提供一些方法,来决定每个用户或应用程序可访问什么视图。通过授权和身份鉴别过程,安全还具有限制初始访问数据库的功能。这些过程中最常用的是注册名和口令保护服务,大多数计算机用户对此都很熟悉。

5)语言接口:DBMS提供对用于定义和操作数据的语言的支持。概念模式是用数据定义语言(data definition language,DDL)说明的。这种数据库语言部件是用来描述数据、数据间联系和对数据和联系的约束的表示法。DDL首先用在数据库设计时,以后修改模式时还会用到DDL。

数据操纵语言(data manipulation language,DML)用于表达数据库上的操作。DML有时也称为查询语言。DBMS提供DML,以使用户和应用程序编写者可访问数据库中数据,而不必知道数据库如何存储数据或把数据存在何处。

6)容错性:不管发生什么故障仍能继续提供可靠DBMS服务的能力称为容错性。一个出错的数据库部件将使与其交互的其他部件产生故障。典型的数据库故障包括违反约束的事务超时错误。如上所述,恢复与容错性密切相关,因为恢复是一种机制,它能容许发生使事务异常终止的差错。

7)数据字典:数据字典是一个系统数据库,它含有主数据库中数据的描述(有时被称为元数据,metadata)。它还包含有关数据、联系、约束的信息,以及将这些特征组织到一个统一数据库中的所有模式的信息。通过查询字典可获得有关主要数据库结构的信息,因而被看成是一个数据库。

8)存储管理:DBMS提供数据持久存储的管理机制。内部模式定义数据应该如何用存储管理机制存储。为了访问物理存储,存储管理系统与操作系统间有接口。

1.6 数据库系统的组成

数据库系统是一个复杂的系统,它是采用数据库技术的计算机系统。数据库系统的含义不仅仅是对一组数据进行管理的软件(即数据库管理系统),也不仅仅是一个数据库。数据库系统是一个可运行的按照数据库方法存储、维护和向应用提供数据支持的系统,它是存储介质、处理对象和管理系统的集合体,由数据库、硬件、软件 and 用户四部分组成。

(1)数据库(DB)

数据库是与一个特定组织各项应用有关的全部数据的汇总。通常由两大部分组成:一部分是应用所需要的数据的集合,统称为物理数据库,它是数据库的主体;另一部分是关于各级数据库结构的描述,成为描述数据库,由数据字典系统管理。

(2)硬件支持系统

这一部分包括中央处理机、内存、外存、输入输出设备、数据通道等硬件设备。计算机系统一般是从市场上选购的,对数据库系统来说,特别要考虑内存、外存、输入输出存取速度、可支持终端数和性能稳定等指标,在许多应用中还要考虑系统支持联网的能力和配备后备存储器等因素。

数据库要存储大量的数据,现在磁盘容量已达每轴10 000MB,有了光盘的使用,大容量数据库的物质条件越来越优越。数据库还需要足够的磁带作备份。此外,还要求系统有较高的通道能力,以提高数据的传输速度。

(3)软件支持系统

这一部分包括DBMS、OS、各种宿主语言和实用程序等。DBMS是管理数据库的软件系统,而且还要在OS支持下才能工作。为了开发应用系统,还要有各种宿主语言,这些语言属于第三代语言范畴,并且与数据库系统有良好的接口。

应用开发支撑软件是为应用开发人员提供的高效率、多功能的交互式程序设计系统,一半属于第四代语言范畴,包括报表生成系统、表格系统、图形系统、具有数据库访问和表格输入/输出功能的软件、数据字典系统等。它们为应用系统的开发提供了良好的环境,提高效率20至100倍。这些软件系统均以DBMS为核心。

(4)用户

大型数据库的设计、加载、维护可能涉及许多人,这些人员可以分为四类,每类人员都从不

同的角度使用数据库系统,我们统称他们为数据库系统用户。下面分别介绍这四类用户如何使用数据库系统。

1)数据库管理员

在任何组织机构中,如果很多人共享相同资源,则需要一个特殊的人员来监督和管理这个共享资源。在数据库系统环境下,共享资源有两类:第一类是数据库,第二类是数据库管理系统软件和相关软件。这些资源的监督管理由数据库管理员完成。数据库管理员可以由一人担任,也可以由一组人担任。数据库管理员负责为存取数据库的用户授权,并协调和监督他们对数据库和数据库管理系统软件的使用。数据库管理员也负责安全性保护和系统的监督和改善。

2)数据库设计者

数据库设计者负责数据库总的数据库确定、数据库文件结构的设计、存取方法的选择和数据库的最后定义。这些工作完成后,所设计的数据库才能在数据库系统中实现。数据库设计者首先需要与所有用户接触,讨论研究用户的需要,为每个用户建立起一个适应于他的应用的数据库视图。然后,数据库设计者合并这些视图,形成完整的数据库定义。最后形成的数据库必须能够支持所有用户的应用需求。最后,数据库设计者与数据库管理员合作,在数据库系统中建立数据库并加载数据。在很多情况下,数据库设计者由数据库管理员担任。

3)最终用户

最终用户是数据库的主要用户,经常对数据库提出查询与更新要求等操作要求。数据库主要是为这类用户设计和存储的。最终用户可以分为以下三类:

①偶然用户 这类用户不经常访问数据库。他们每次访问数据库时往往需要不同的数据库信息。他们通常使用数据库查询语言表示查询要求。他们一般都是一个企业或一个组织机构的高中级管理人员。

②简单用户 数据库的多数最终用户都是简单用户。他们的主要工作是查询和更新数据库。他们一般都不直接使用数据库管理系统,而是通过应用程序员精心设计的并具有友好界面的应用程序存取数据库。银行的职员、航空公司的机票预定人员、旅馆总台服务员等都属于这类用户。

③复杂用户 复杂用户包括工程师、科学家、经济学家、科学技术工作者等具有较高科学技术背景的人员。这类用户一般都比较熟悉数据库管理系统的各种机制,能够直接通过数据库管理系统访问数据库。有些人甚至能够用数据库管理系统编制自己的应用程序。

针对上述多种最终用户,数据库管理系统需要提供多种界面和服务机制。简单用户不需要深入了解数据库管理系统的界面和服务机制,只需要熟悉他们所使用的应用程序的用户界面。偶然用户只需要熟悉他们经常使用的数据库管理系统的界面和服务机制。复杂用户则需要深入了解数据库管理系统的各种界面和服务机制。

4)系统分析员和应用程序员

系统分析员负责分析最终用户特别是简单用户的需求,给出适应这些用户需求的数据库事务的准确定义。

应用程序员负责把系统分析员提供的数据库事务的定义编制成计算机软件,并进行编码、调试、维护。

5)与数据库系统有关的其他人员

除了数据库的设计者、管理者、使用综合应用程序设计者以外,还有如下三类与数据库系

统有关的人员。

第一类人员是数据库管理系统的设计和实现人员。他们负责数据库管理系统软件本身的设计、编码、调试和维护。数据库管理系统是一个十分复杂的计算机软件系统,包括很多程序模块,如数据字典处理模块、查询语言处理模块、查询优化模块、数据存取模块、事务处理模块、安全与完整性维护模块等。所有这些模块的设计和实现都是数据库系统设计和实现人员的任务。

第二类人员是数据库系统工具开发者。数据库系统工具包括数据库设计软件包、性能监控软件包、自然语言查询处理软件包、图形界面软件包、应用程序辅助设计软件包等。数据库系统工具开发者负责这些工具软件包的开发。很多数据库系统工具的开发都是独立于数据库系统开发公司的软件开发者。

第三类人员是操作人员和系统维护人员。他们负责数据库系统赖以运行的硬件和软件环境的操作、维护和管理。

1.7 使用数据库管理系统的原因

为什么应用系统要选用数据库和数据库管理系统呢?因为数据库为共享的稳定数据提供了安全和有保障的存储介质。如果用户不需要数据在其作用范围外持久存储,数据库就不能为此类用户提供增值服务。反之,如果用户要求数据在某个程序或应用的作用范围之外存在,则数据库无疑是维护这些持久数据的合适之地。但是仅要维护持久数据并不意味着非用数据库管理系统不可。为判断数据库是否合适,我们还必须看被管理的数据是否有结构、联系和约束。如果数据没有这些特性,可能用文件系统更为合适。例如,如果我们希望知道一个汽车工厂的零件清单,用数据库就非常合适。决定用文件系统还是用数据库,必须考虑所存储的数据以及这些数据是如何使用的。数据库管理系统提供查询数据库的功能,从而找到相关联的数据项。例如,可使用汽车工厂的零件清单数据库,判断我们是否已具备了装备某个汽车所需的全部零件。查询的简单性取决于可用的数据库操纵语言、使用的逻辑数据库结构和存储数据项的粒度(*granularity*)。

1.8 数据库管理系统的应用领域

在当今信息时代,我们的生活越来越多地依赖信息的存取和使用,数据库管理系统正日益广泛地运用于最终用户之中。我们使用数据库访问银行账户信息,从而使存取钱更为快捷方便。我们购买股票时,使用数据库将一笔钱从银行户头转移到股票户头。股票交易商使用数据库为我们买股票,卖股票,或获取有关某种股票的信息。数据库设法完成众多日常信息的存取和操作动作。

教育系统中所使用的数据库的用途有很大不同。教育工作者使用数据库存储信息资料,查询信息,并检索需要的信息。数据库正进入早期儿童教育系统,该系统可访问和存储学习工具以及驾驭这些工具的信息。数据不必直接存储在使用它的地方,但在需要时即可通过数据库管理系统获得。

制造业依靠数据库来管理不断增加的信息。厂商在最初研究或设计阶段使用数据库维护

产品信息(例如计算机辅助设计系统);生成一个完整的产品设计,直至最小的部件;维护工厂使用的部件清单信息;收集有关产品制造、存储、编号、装卸的数据;以及监控消费者的满意程度(见图1.8)。

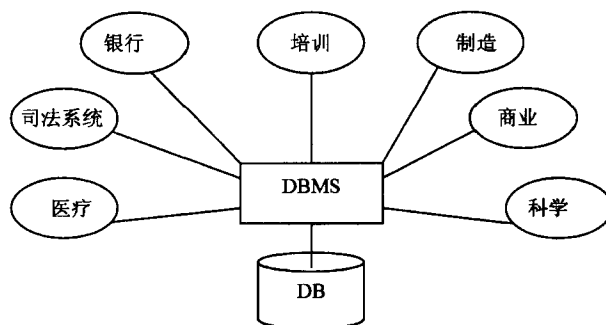


图1.8 数据库管理系统的用途

司法系统开始使用数据库存储案例,并使用户可迅速找到和访问大量丰富的信息。数据库有助于定义数据之间的相互联系并维护这些联系,这样就允许必要时查询这些信息。查询数据库从过去的案例中找出与指定关键字或联系相匹配的案例。

医学界使用数据库存储有关疾病的信息,科学界使用数据库存储科学发现的信息,以便现在和将来的研究人员可快速访问这些信息。这将有助于我们更好、更全面地研究疾病和科学发现。

1.9 数据库技术的新进展

数据库系统并非只以一种形式出现,随着数据库技术的发展,目前已开发出多种类型的数据库,如并行数据库、分布式数据库、多媒体数据库、实时数据库、容错数据库、安全数据库和混合(federated)数据库。这些种类数据库都是针对不同目的而设计的,每一类都支持其设计所针对的应用环境。

下面分别介绍这几种系统。

1.9.1 分布式数据库管理系统

分布式数据库系统并非单纯意味着数据是分布的。分布式数据库的定义也意味着它结合了知识、动作,以及对组成分布式计算机系统的分布式部件的控制(见图1.9)。要理解分布式数据库管理系统是什么,必须首先研究应用开发人员为何需要使这个系统分布化并加以控制。大多数分布式计算机系统被用来减轻和分配企业的工作负担,或者使数据处理功能有更靠近完成该功能的物理网点。这两种情况中,不是单纯转移数据处理功能或分布其计算,而是为了让用户看不到“分布”,目的是为了用户只看到一个单一的计算机系统。这意味着所有的资源管理功能必须在多个网点之间巧妙地加以协调,不留痕迹。

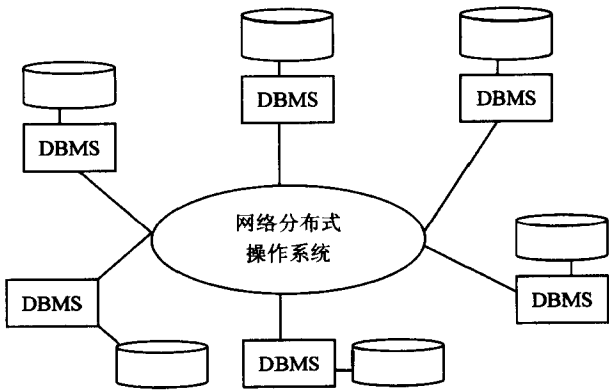


图19 分布式数据库系统

1.9.2 实时数据库管理系统

实时数据库系统并不单是一个速度快的系统。实时是指操作系统和与其交互的现实世界之间的时态(temporal)交互。实时控制系统可能是集中式、分布式或混合式的。实时系统的显著特征是它在操作的各方面都用到时间(见图1.10)。

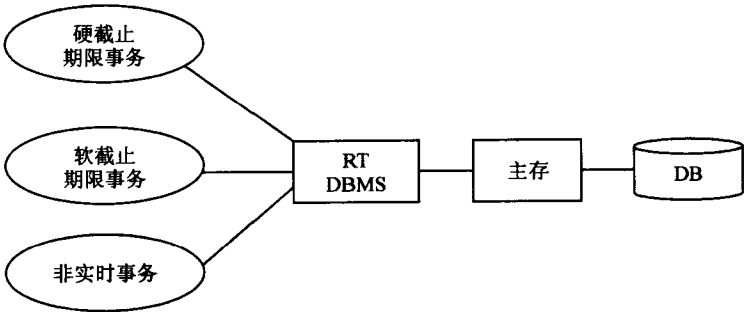


图110 实时数据库系统

实时计算系统过去已被应用于物理控制之中,在这种应用中,计算机提取物理系统的有关信息,在获取的信息基础上决定作何响应(控制动作),并完成控制动作。所有这些动作都按如下方式完成:输入和输出功能的响应都被预先确定,响应时间受到限定。这些响应还以这样的方法加以优化——每个输入条件都有可预先确定的输出条件,它们总是在同一时刻并以同种方式发生。

这并不是说实时计算只限于控制应用。实时系统的最恰当的定义是,能够在时间帧内对现实世界的活动作出可预先确定的响应,从而可以进行正确的控制。因此,实时系统的正确性不只是用计算机的响应来定义。

1.9.3 容错数据库管理系统

容错系统是指一个系统遇到硬件和软件部件故障,还能够保持某种设计好的服务等级。一个容错系统与实时系统有一些相似的特征。容错系统必须预先进行分析,描绘出系统中所有可能的出错点,并设计出系统检测、修正和从错误中恢复的方法。所有这些功能应尽量减少对运

行中的应用的影响(见图1.11)。

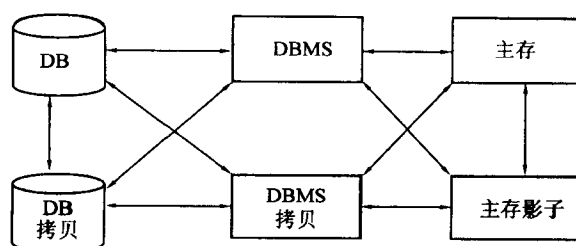


图1.11 容错数据库系统

容错系统必须注意到各种各样的故障来自众多系统元素。一个故障可能是由硬件、软件、两种部件的组合或作用于系统的外部条件产生的。故障可能是断断续续、瞬间发生的,或者可能持久存在。容错系统必须能监控系统、检测故障、隔离引发故障的条件,以及纠正故障或为更高层管理机构提出建议,从而使系统恢复到用户可接受的状态。

1.9.4 安全数据库管理系统

安全系统是指在一个系统中,用户在何时、在多大范围能完成何种操作都是可控制的。例如一个公司的人事系统,它可能希望允许所有用户能查看自己的档案,以及查看适合其工作的信息,但它可能不希望让所有用户查看其他雇员的档案,或者甚至不让他们看到自己的某一部分信息。为了提供这种服务,数据库系统必须定义访问权限,并检查试图访问数据的用户是否具有这些权限。这就是一个非常简单的安全系统的例子(见图1.12)。

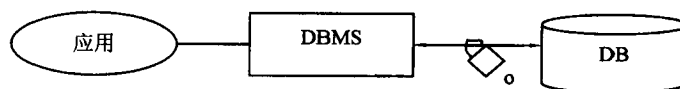


图1.12 安全数据库系统

安全不只限于这种简单的访问授权,它还包括更详细的安全检查。我们可能希望限制用户只经查找一些无关联的数据就能进一步推断出其他信息的能力。在此形势下,我们可能需要提供对数据各个部分或数据间联系的详细控制。

1.9.5 异构数据库管理系统

异构系统是指由多个各不相同的子系统组成的系统(见图1.13)。例如,一个公司的每个部门都有自己的数据处理需求。人事部门有自己的一系列需求,工程部门有另一些需求,而制造部门的需求又不同。每个部门的计算机系统的硬件和软件可能也不同。如果要求这些不同的系统进行交互,我们就必须用一些通信媒质将它们连接起来,这样就有了一个异构系统。这种系统需要解决一些特殊问题,基本问题涉及到最底层的硬件数据表示。如果一台计算机使用8位字表示,另一台使用16位字,而另一台使用32位字,我们应如何处理这些差异才能使系统间能够交互有用的数据呢?

由于以上这个例子所展示的需求情况,异构数据库系统已得以发展。公司的某个部门购买数据库后,将本部门的数据录入数据库,并编写多个使用这些共享数据的应用程序。一旦要求这些数据库的数据被其他部门共享时,我们就会发现问题。异构数据库系统的开发正是为了解

决这些遗留的数据库问题。

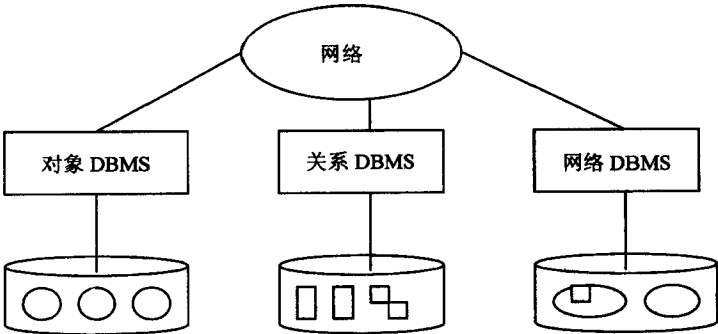


图 1.13 异构数据库管理系统

1.9.6 多媒体数据库管理系统

高性能计算、大容量数据存储能力的发展,以及国家信息基础设计(信息高速公司)的规划,将促进多媒体计算系统的使用和发展。而多媒体计算系统需要数据库管理系统。多媒体计算系统使用各种数据源,例如图像、视频、语音、声音和文本,并将这些数据结合到应用和产品中。这些复杂的数据源应便于被计算系统访问,并且数据表示应便于被交叉使用。为了支持交互的用户使用,多媒体计算系统要求信息存取和表示是同步进行和实时存取的(见图1.14)。

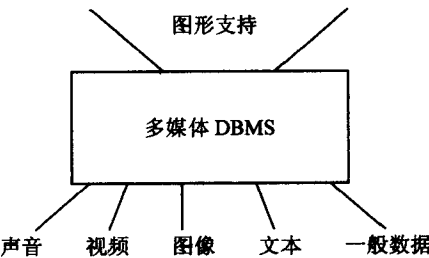


图1.14 多媒体数据库系统

多媒体数据库管理系统将实时数据库的需求与交互式图形系统的需求结合起来。

1.9.7 数据库管理的未来

信息需求的增长使数据库系统的地位渐涨。数据库系统的发展是为了满足共享信息用户的需求。将来的数据库系统也是如此。数据库和数据库管理系统正在探寻前所未有的应用领域。数据库系统正应用于医学监控、医学诊断、医学研究、计算机辅助设计、计算机辅助工程、计算机辅助制造、能源管理、图书馆管理、信息宣传、航空系统、太空系统、防御系统、汽车导航、天气预报、交通预计、旅馆预订和许多其他领域。

随着在线信息的增加而且越来越多的用户希望访问在线信息,今后将开发和生产更多面向应用支持的新数据库管理系统。第7章将研究数据库系统研究和发展的趋势,展望这一技术的未来。