

SHUJUKU XITONG LILUN
JIQI XINJISHU YANJIU

数据库系统理论 及其新技术研究

刘月兰 杨秀荣 韩丽娜 编著



中国水利水电出版社
www.waterpub.com.cn

数据库系统理论 及其新技术研究

刘月兰 杨秀荣 韩丽娜 编著



中国水利水电出版社
www.waterpub.com.cn

· 北京 ·

内 容 提 要

本书系统全面地介绍了数据库系统概述、数据模型、关系数据库理论、关系数据库标准语言 SQL、关系数据库规范化理论、数据库系统的设计与实施、数据库的安全性、数据库事务管理与实现、数据库访问标准、现代数据库新技术。本书取材广泛,内容丰富,解析清楚,讲述明确,通俗易懂,可供从事数据库开发应用的研究人员和工程技术人员参考。

图书在版编目(CIP)数据

数据库系统理论及其新技术研究 / 刘月兰, 杨秀荣,
韩丽娜编著. —北京: 中国水利水电出版社, 2016. 12
ISBN 978-7-5170-4943-2

I. ①数… II. ①刘… ②杨… ③韩… III. ①数据库
系统 IV. ①TP311.13
中国版本图书馆 CIP 数据核字(2016)第 296750 号

责任编辑:杨庆川 陈 洁 封面设计:崔 蕾

书 名	数据库系统理论及其新技术研究 SHUJUKU XITONG LILUN JIQI XIN JISHU YANJIU
作 者	刘月兰 杨秀荣 韩丽娜 编著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址:www. waterpub. com. cn E-mail:mchannel@263. net(万水) sales@waterpub. com. cn 电话:(010)68367658(营销中心)、82562819(万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京鑫海胜蓝数码科技有限公司
印 刷	三河市佳星印装有限公司
规 格	184mm×260mm 16 开本 17.25 印张 420 千字
版 次	2017 年 1 月第 1 版 2017 年 1 月第 1 次印刷
印 数	0001—2000 册
定 价	59.00 元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换
版权所有·侵权必究

前 言

在社会信息化的今天,信息已经成为全社会宝贵的资源。就信息本身而言,大致可将其分为结构化信息、半结构化信息和非结构化信息三类。作为管理结构化信息的有效手段,数据库系统对于当今的研发部门、政府机关、企事业单位等都至关重要;作为数据库系统的核心,数据库管理系统 DBMS——特别是关系型数据库管理系统用于高效创建数据库和存储大量数据并对其进行有效管理和维护,其本身具有很大的工程实用价值。

虽然数据库技术产生于 20 世纪 60 年代末,但经过 40 多年的迅猛发展,已经是当前发展最快、最受人关注、应用最广泛的科学技术之一。数据库系统已经成为现代信息系统不可或缺的核心组成部分。在当今社会,不仅是传统的商业、管理和行政事务型应用离不开数据库,那些实时、过程 and 控制的工程型应用领域也要求并且开始使用现代(非传统)数据库。因此,人们越来越普遍地要求全面学习和掌握数据库的理论知识、系统技术、应用方法及其最新发展情况。作者根据多年研究,以培养技术型人才为目标编撰了本书。

本书共分为 9 章,分别是数据库系统概述、数据模型、关系数据库、关系数据库标准语言 SQL、关系数据库规范化理论、数据库系统的设计与实施、数据库的安全性 with 完整性、数据库事务管理与实现和现代数据库新技术。

本书具有如下特点:

- 既注重有重点地介绍数据库的基本原理和方法,又补充现代数据库系统的新技术、新知识、新水平和新趋势。
- 缩减传统数据库系统的部分内容,突出数据库理论与实践紧密结合的特征,结合应用实例讲解,突出能力训练。
- 本书根据知识点、要点及层次,结合实践的特点来组织内容,对部分难点配以直观的图示和具体的示例。
- 在内容选取、章节安排、难易程度、例子选取等方面,充分考虑到理论与实践相结合。力求既讲述知识,又介绍探讨问题的思路。

本书在编撰过程中得到了单位领导和同仁的热情帮助和支持,在此表示衷心的感谢!

本书的编撰参考了大量的著作和文献,在此向相关作者表示感谢!

在编撰的过程中,我们力求精益求精,但难免存在不足或疏漏之处,真诚希望同行专家和读者对本书提出宝贵的意见和建议。

作 者

2016 年 8 月

目 录

前言	
第 1 章 数据库系统概述	1
1.1 数据、信息及其管理	1
1.2 数据库与数据库管理系统	2
1.3 数据库体系结构	9
1.4 数据库技术的应用与发展	12
第 2 章 数据模型	17
2.1 数据模型概述	17
2.2 数据模型的组成要素	20
2.3 E-R 数据模型	20
2.4 关系数据模型	29
2.5 其他数据模型	33
第 3 章 关系数据库	42
3.1 关系数据库概述	42
3.2 关系数据结构及形式化定义	44
3.3 关系的完整性	50
3.4 关系代数	51
3.5 关系演算	65
第 4 章 关系数据库标准语言 SQL	69
4.1 SQL 语言概述	69
4.2 数据定义	71
4.3 数据查询	74
4.4 数据更新	89
4.5 数据控制	91
4.6 视图管理	96
4.7 嵌入式 SQL	99
第 5 章 关系数据库规范化理论	105
5.1 关系模式的非形式化设计规则	105
5.2 函数依赖	108
5.3 关系模式的规范化	111

5.4	关系模式的分解特性	120
5.5	关系模式的优化	127
5.6	关系查询优化	128
第 6 章	数据库系统的设计与实施	135
6.1	数据库设计概述	135
6.2	系统需求分析	139
6.3	概念结构设计	146
6.4	逻辑结构设计	160
6.5	数据库物理设计	164
6.6	数据库实施、运行及维护	166
第 7 章	数据库的安全性及完整性	170
7.1	数据库的安全性	170
7.2	数据库的完整性	190
第 8 章	数据库事务管理与实现	199
8.1	事务与事务管理	199
8.2	故障管理	202
8.3	并发控制与封锁机制	203
8.4	死锁和解决方案	215
8.5	数据库恢复技术	216
8.6	数据库复制与数据库镜像	219
8.7	数据库再设计的实现	223
第 9 章	现代数据库新技术	236
9.1	现代数据库系统概述	236
9.2	分布式数据库系统	238
9.3	对象关系数据库系统	249
9.4	数据仓库	257
9.5	数据挖掘	264
参考文献		269

第 1 章 数据库系统概述

1.1 数据、信息及其管理

1.1.1 数据与数据库

1. 数据

一提到数据(Data),人们在大脑中就会浮现像 3、4、3、-50 等数字,认为这些就是数据,其实不然,这只是最简单的数据。从一般意义上说,数据是描述客观事物的各种符号记录,可以是数字、文本、图形、图像、声音、视频、语言等。从计算机角度看,数据是经过数字化后由计算机进行处理的符号记录。例如,我们用汉语这样描述一位读者“王建立,男,年龄 18 岁,所学专业为计算机”,而在计算机中是这样表示的:(王建立,男,18,计算机)。将读者姓名、性别、年龄和专业组织在一起,形成一个记录,这个记录就是描述读者的数据。

数据本身的表现形式不一定能完全表达其内容,如 1 这个数据可以表示 1 门课,也可以表示逻辑真,还可以表示电路的通等。因此,需要对数据进行必要的解释和说明,以表达其语义。数据与其语义是不可分的。

2. 数据库

在日常工作中,人们会获取大量数据,并对这些数据进一步加工处理,从中获取有用的信息。在加工处理之前,一般都会将这些数据保存起来。以前人们利用文件柜、电影胶片、录音磁带等保存这些数据,但随着信息时代的来临,各种数据急剧膨胀,迫使人们寻求新的技术、新的方法来保存和管理这些数据,由此数据库(DataBase,DB)技术应运而生。

将你的家人、同事和朋友的姓名、工作单位和电话号码组织起来,就可以形成一个小型数据库,也可以将公司客户的这些数据组织起来,形成更大规模的数据库。数据库就像是粮仓,数据就是粮仓中的粮食。实际上,数据库是长期存储在计算机内的、有组织的、可共享的大量数据的集合。这里的长期存储是指数据是永久保存在数据库中,而不是临时存放;有组织是指数据是从全局观点出发建立的,按照一定的数据模型进行描述和存储,是面向整个组织,而不是某一应用,具有整体的结构化特征;可共享是指数据是为所有用户和所有应用而建立的,不

同的用户和不同的应用可以以自己的方式使用这些数据,多个用户和多个应用可以共享数据库中的同一数据。

1.1.2 数据处理和数据管理

在日常实际工作中,人们越来越清楚地认识到对数据的使用离不开对数据的有效管理,例如,企事业单位都离不开对人、财、物的管理,而人、财、物是以数据形式被记录和保存的,因此对人、财、物的管理就是对数据的管理。早期以手工方式对这些数据进行管理,现在大多以计算机对数据进行管理,使得数据管理成了计算机应用的一个重要分支。

数据处理是指对数据的收集、组织、存储、加工和传播等一系列操作。它是从已有数据出发,经过加工处理得到所需数据的过程。

数据管理是指对数据的分类、组织、编码、存储、检索和维护工作。数据管理是数据处理的核心和基础。

1.2 数据库与数据库管理系统

1.2.1 数据库系统

1. 数据库系统的组成

如图 1-1 所示,一个典型的数据库系统由四个部分组成:用户、数据库应用程序、数据库管理系统(DBMS)和数据库。然而,结构化查询语言(SQL)是一种国际公认的,被所有商业数据库管理系统产品所理解的标准语言。鉴于 SQL 在数据库处理中的重要性 and 数据库应用程序通常是用 SQL 语句来处理数据库管理系统的这一事实,我们可以用图 1-2 更加完善地描述数据库系统。

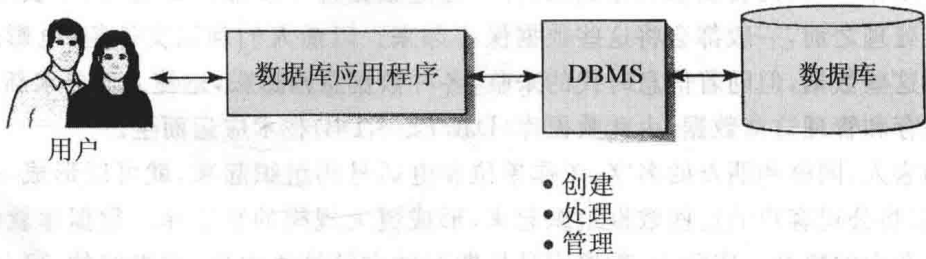


图 1-1 数据库系统的组成部分

从图 1-2 的右边开始,数据库是相关表和其他结构的集合。数据库管理系统(DBMS)是一个用来创建、处理和管理数据库的计算机程序。DBMS 接受 SQL 请求,然后把这些请求转

换成数据库上的操作。DBMS 是需要被软件供应商许可的一个又大又复杂的程序。几乎没有企业编写自己的数据库管理系统程序。

数据库应用程序(Database Application)是作为用户和 DBMS 中介的一个或多个计算机程序的集合。应用程序通过提交 SQL 语句给 DBMS,从而读取或修改数据库数据。应用程序同时又以表单或报表的方式返回数据给用户。应用程序可以从软件供应商那里获得,而且它们经常都是被写在内部。

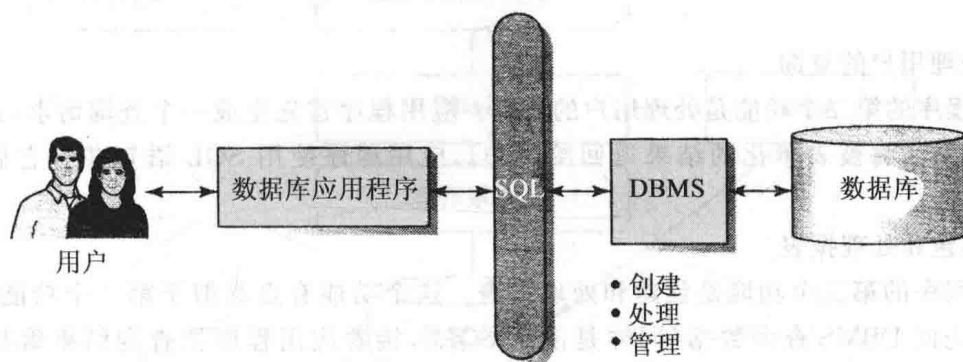


图 1-2 带有 SQL 的数据库系统组成部分

数据库系统的第四个组成部分是用户(User)。用户通过数据库应用程序明了事情,他们使用表单去读取、输入和查询数据,并且生成表达信息的报表。

图 1-3 所示为数据库系统的结构。

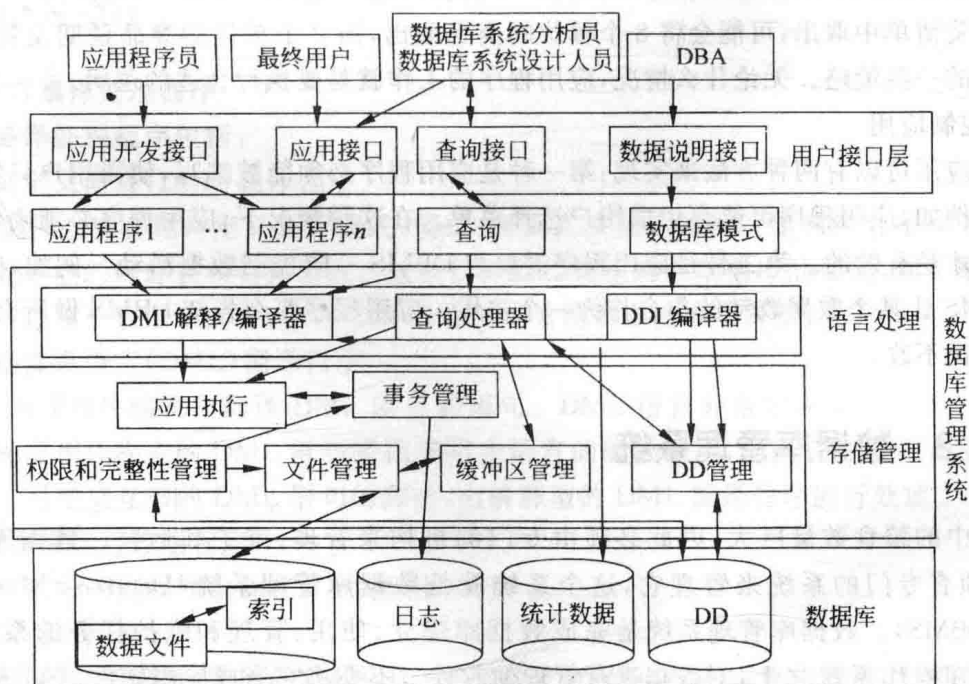


图 1-3 数据库系统结构

2. 数据库应用程序的基本功能

(1) 创建和处理表单

首先,应用程序创建和处理表单。像所有的数据输入表单一样,这个表单的目的是以实用的方式把数据呈现给用户,而不用考虑基础表结构。除了屏幕显示的表单之外,应用程序根据用户的操作处理数据,生成 SQL 语句,用于插入、更新或修改任何一个构成这个表单的表中的数据。

(2) 处理用户的查询

应用程序的第二个功能是处理用户的查询。应用程序首先生成一个查询请求,并且发送给 DBMS,然后将被表单化的结果返回给用户。应用程序使用 SQL 语句并将它们传递给 DBMS 处理。

(3) 创建和处理报表

应用程序的第三个功能是创建和处理报表。这个功能有点类似于第二个功能,因为应用程序首先向 DBMS 查询数据(同样是使用 SQL),接着应用程序把查询结果编排成报表样式。

(4) 执行应用逻辑

除了生成表单、查询和报表,应用程序还会采取其他方式根据特定应用逻辑来更新数据库。例如,假设一个用户使用订单录入应用程序请求 10 个单位的某个商品,进一步假设当应用程序查询数据库后(通过 DBMS),发现只有 8 个单位的该商品在库存中,那么接下来应该怎么做呢?这就取决于具体应用程序的逻辑。可能会将这一结果告知用户,但任何一个商品都不会从存货清单中取出;可能会将 8 个单位的商品取出,而 2 个单位的商品延期交货;也可能采取其他的一些策略。无论什么情况,应用程序的工作就是要执行合适的逻辑。

(5) 控制应用

控制应用可以有两种方法来实现:第一种是应用程序必须能被编写,使得用户只能看到逻辑选择。例如,应用程序可能会生成用户选择菜单。在这种情况下,应用程序必须确保只有合适的选择才是有效的。第二种是应用程序需要与 DBMS 一同控制数据活动。例如,应用程序指导 DBMS 让某个数据改动的集合作为一个整体。应用程序要么告知 DBMS 做所有的改动,要么一个都不改。

1.2.2 数据库管理系统

粮仓中的粮食数量巨大,因此必须由专门的机构来管理、维护和运营。数据库也是一样,也必须有专门的系统来管理它,这个系统就是数据库管理系统(DataBase Management System,DBMS)。数据库管理系统是完成数据库建立、使用、管理和维护任务的系统软件。它是建立在操作系统之上,对数据进行管理的软件。不要将它当成应用软件,它不能直接用于诸如图书管理、课程管理、人事管理等事务管理工作,但它能够为事务管理提供技术、方法和工具,从而能够更好地设计和实现事务管理软件。数据库系统如图 1-4 所示。

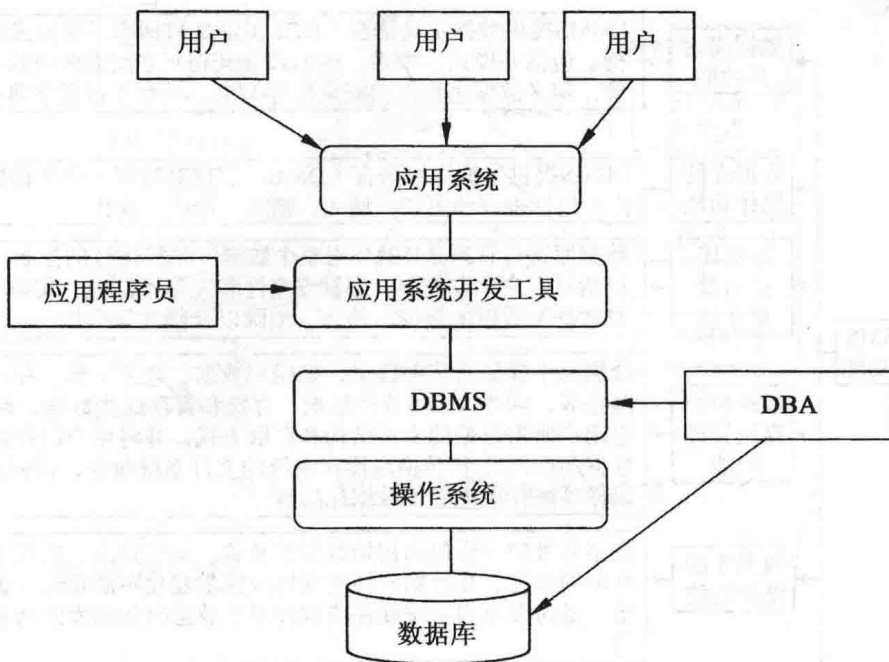


图 1-4 数据库系统

1. DBMS 的功能

DBMS 的具体功能如图 1-5 所示。

2. DBMS 组成

(1) 语言编译处理程序

语言编译处理程序包括：

1) 数据定义语言(DDL)翻译程序

DDL 翻译程序将用户定义的子模式、模式、内模式及其之间的映像和约束条件等这些源模式翻译成对应的内部表和目标模式。这些目标模式描述的是数据库的框架，而不是数据本身。它们被存放于数据字典中，作为 DBMS 存取和管理数据的基本依据。

2) 数据操纵语言(DML)翻译程序

DML 翻译程序编辑和翻译 DML 语言的语句。DML 语言分为宿主型和交互型。DML 翻译程序将应用程序中的 DML 语句转换成宿主语言的函数调用，以供宿主语言的编译程序统一处理。对于交互型的 DML 语句的翻译，由解释型的 DML 翻译程序进行处理。

(2) 数据库运行控制程序

数据库运行控制程序主要有 6 种，如图 1-6 所示。

(3) 实用程序

实用程序主要有初始数据的装载程序、数据库重组程序、数据库重构程序、数据库恢复程序、日志管理程序、统计分析程序、信息格式维护程序以及数据转储、编辑等实用程序。数据库用户可以利用这些实用程序完成对数据库的重建、维护等各项工作。

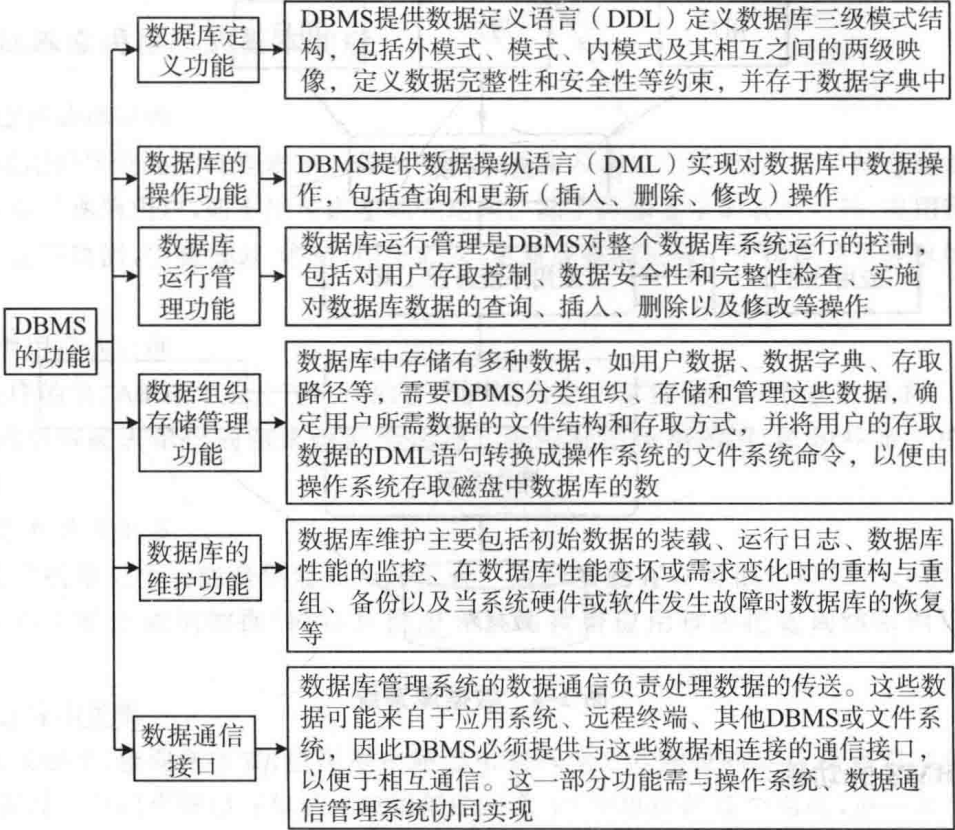


图 1-5 DBMS 的功能

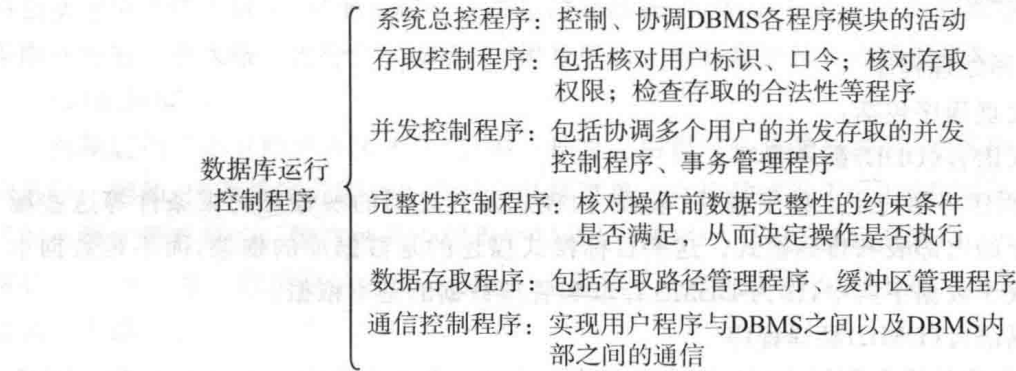


图 1-6 数据库运行控制程序

3. DBMS 的工作过程

在数据库系统中,当用户或一个应用程序需要存取数据库中的数据时,应用程序、DBMS、操作系统、硬件等几个方面必须协同工作,共同完成用户的请求。下面以一个程序 A 通过 DBMS 读取数据库中的记录为例来说明 DBMS 的工作过程,如图 1-7 所示。

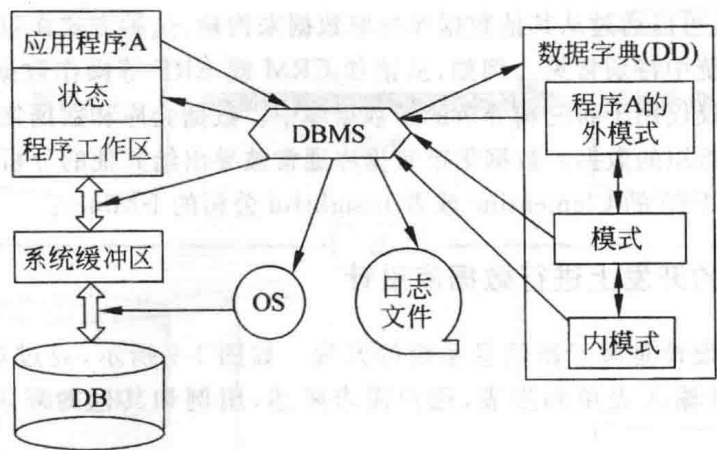


图 1-7 DBMS 存取数据操作过程

1. 2. 3 数据库设计

1. 在已有的数据上进行数据库设计

如图 1-8 所示,第一种数据库设计是从已有的数据中构建数据库。在某些情况下,给一个开发队伍提供一组电子数据表或是包含数据表的文本文件,要求他们开发一个数据库,并将来源于电子数据表和其他表的数据导入到新的数据库中。

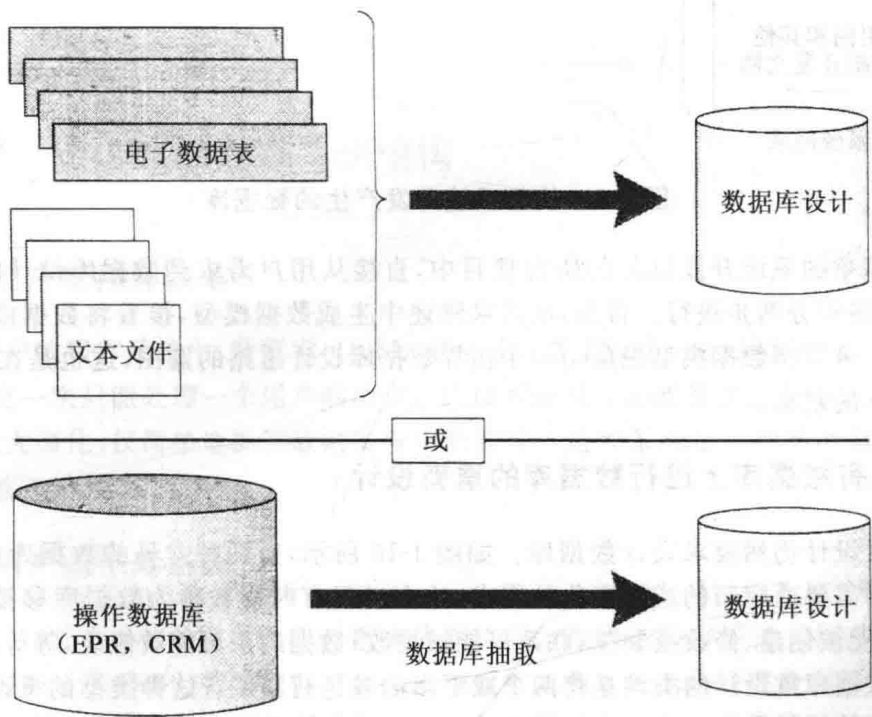


图 1-8 从已有数据中构建数据库

此外,数据库也可以通过从其他数据库抽取数据来构建,这种方式在包含报表和数据挖掘应用的商务智能系统中特别常见。例如,从诸如 CRM 或 ERP 等操作数据库中抽取的数据,可能被复制到一个仅仅用于研究和分析的新数据库中。数据仓库和数据集市数据库用来存储特意为研究报表而组织的数据。数据集市数据库通常被导出给其他的分析工具,比如 SAS 的 Enterprise Miner,SPSS 的 Clementine 或者 Insightful 公司的 I-Miner。

2. 在新系统的开发上进行数据库设计

另一种数据库设计是源于新信息系统的开发。如图 1-9 所示,通过对新系统需求的分析,例如需要的设计输入表单和报表,用户需求陈述,用例和其他的需求等来创建数据库设计。

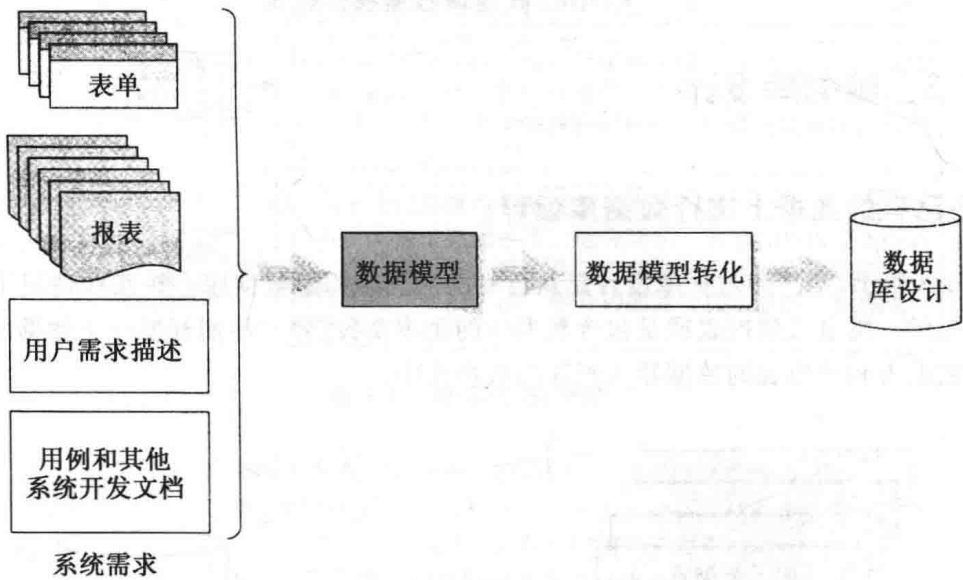


图 1-9 从新系统开发产生的数据库

在除最简单的系统开发以外的所有项目中,直接从用户需求到数据库设计的步伐太大。相应地,开发进程分两步进行。首先,从需求陈述中生成数据模型,接着将数据模型再转化为数据库设计。可以将数据模型想象为一个指导数据库设计道路的蓝图,这也是在 DBMS 中构建实际数据库的基础。

3. 在已有数据库上进行数据库的重新设计

数据库重设计仍然要求设计数据库。如图 1-10 所示,有两种常见的数据库重设计方式。一种是数据库需要适应新的或是变化的需求,这个过程有时候被称为数据库移植。在移植的过程中,表可能被创建、修改或删除;联系可能被更改;数据约束可能被修改,等等。

另一种数据库重设计的类型是将两个或更多的表进行集成。这种类型的重设计在更改或是消除遗留系统时很常见。在企业应用集成中,当两个或更多的本来独立的信息系统被修改为一起工作时,这种重设计也很常见。

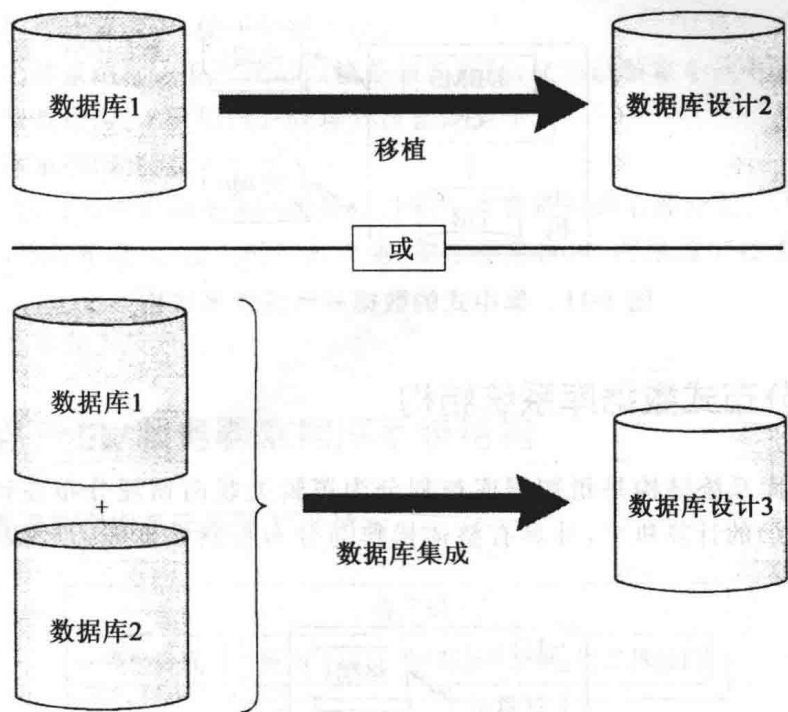


图 1-10 从数据库重设计起源的数据库

1.3 数据库体系结构

1.3.1 集中式数据库系统结构

1. 单用户数据库系统

在单用户数据库系统中,数据库、DBMS 和应用程序都装在一台计算机上,由一个用户独占,并且系统一次只能处理一个用户的请求。因而系统没有必要设置并发控制机制;故障恢复设施可以大大简化,仅简单地提供数据备份功能即可。这种系统是一种早期最简单的数据库系统,现在越来越少见了。

2. 多用户数据库系统

图 1-11 示出了一种多用户数据库系统体系结构。数据的集中管理并服务于多个任务减少了数据冗余;应用程序与数据之间有较高的独立性。但对数据库的安全和保密、事务的并发控制、处理机的分时响应等问题都要进行处理。使得数据库的操作与设计比较复杂,系统显得不够灵活,且安全性也较差。

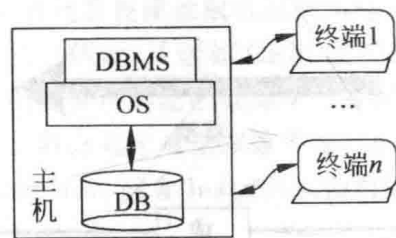


图 1-11 集中式的数据库系统体系结构

1.3.2 分布式数据库系统结构

分布式数据库系统结构是指数据库被划分为逻辑关联而物理分布在计算机网络中不同场地(又称结点)的计算机中,并具有整体操作与分布控制数据能力的数据库系统,如图 1-12 所示。

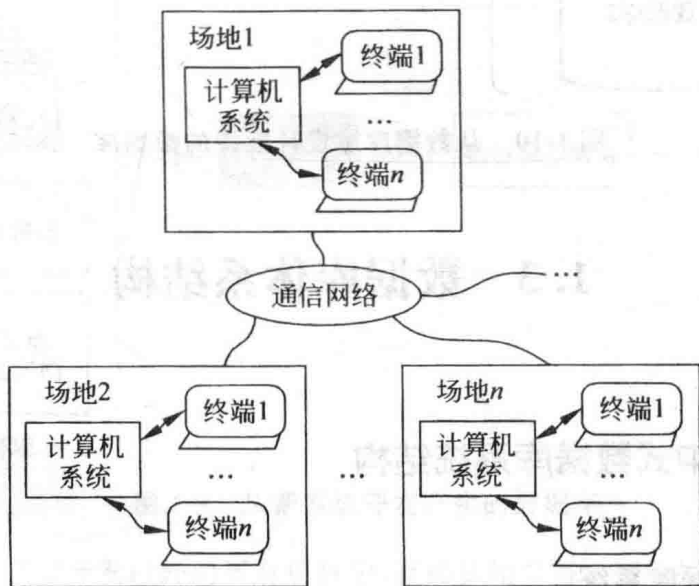


图 1-12 分布式数据库系统体系结构

例如,银行中的多个支行在不同的场地,一个支行的借贷业务可以通过访问本支行的账目数据库就可以处理,这种应用称为“局部应用”。如果在不同场地的支行之间进行通兑业务或转账业务,这样要同时更新相关支行中的数据库,这就是“全局应用”。

分布式数据库系统与集中式数据库系统相比有以下优点:

- ①可靠性高,可用性好。由于数据是复制在不同场地的计算机中,当某场地数据库系统的部件失效时,其他场地仍可以完成任务。
- ②适应地理上分散而在业务上需要统一管理和控制的公司或企业对数据库应用的需求。
- ③局部应用响应快、代价低。可以根据各类用户的需要来划分数据库,将所需要的数据分

布存放在他们的场地计算机中,便于快捷响应。

④具有灵活的体系结构。系统既可以被分布式控制,又可以被集中式处理;既可以统一管理同系统中同质型数据库,又可以统一管理异质型数据库。

- 分布式数据库系统的缺点:
- ①系统开销大,分布式系统中访问数据的开销主要花费在通信部分上。
 - ②结构复杂,设计难度大,涉及的技术面宽,如数据库技术、网络通信技术、分布技术和并发控制技术等。
 - ③数据的安全和保密较难处理。

1.3.3 客户机/服务器数据库系统结构

客户机/服务器数据库系统如图 1-13 所示。

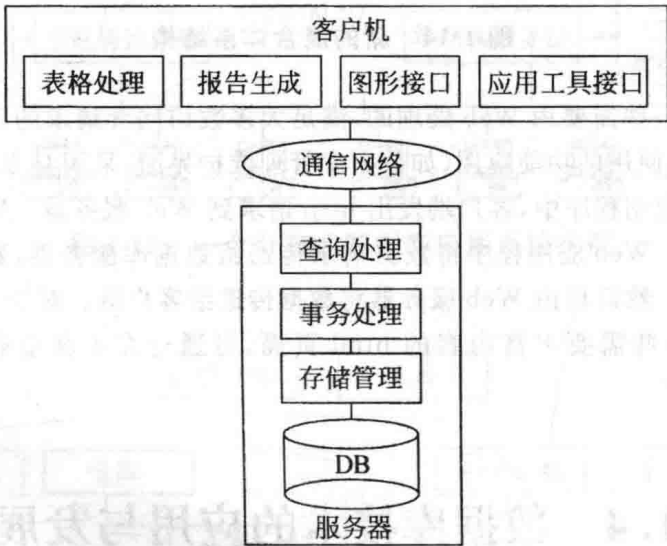


图 1-13 客户机/服务器数据库系统体系结构

客户机也称为系统前端,主要由一些应用程序构成,例如,图形接口、表格处理、报告生成、应用工具接口等,实现前端应用处理。数据库服务器可以同时服务于各个客户机对数据库的请求,包括存储结构与存取方法、事务管理与并发控制、恢复管理、查询处理与优化等数据库管理的系统程序,主要完成事务处理和数据访问控制。

客户机/服务器体系结构的好处是支持共享数据库数据资源,并且可以在多台设备之间平衡负载;允许容纳多个主机,充分利用已有的各种系统。

现代客户机和服务器之间的接口是标准化的,如 ODBC 或其他 API。这种标准化接口使客户机和服务器相对独立,从而保证多个客户机与多个服务器连接。

一个客户机/服务器系统可以有多个客户机与多个服务器。在客户机和服务器的连接上,如果是多个客户机对一个服务器,则称为集中式客户机/服务器数据库系统;如果是多个客户机对应多个服务器,则称为分布式客户机/服务器数据库系统。分布的服务器系统结构是客户