

SHUJUKU SHEJI · SHUJUKU SHEJI ·

数据库设计

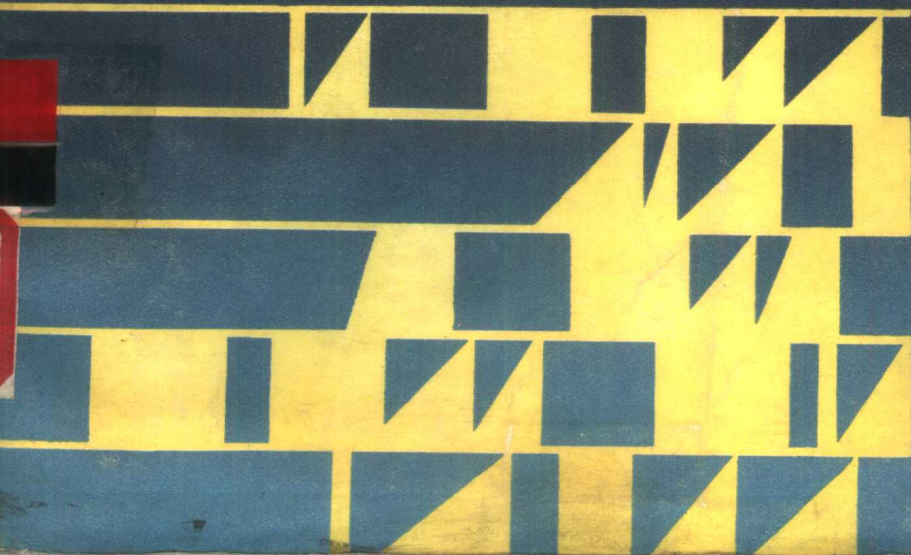
施伯乐

何守才 编著

潘锦平

上海交通大学出版社

计算机软件与应用丛书



计算机软件与应用丛书

数据库设计

施伯乐
潘锦平 编著
何守才

上海交通大学出版社

内 容 提 要

本书讨论如何用网型、层次型和关系型数据库管理系统解决企事业组织中的大型数据处理问题。重点介绍为实际问题建立数据模型的方法和工具。本书供从事数据库实际工作的专业人员阅读，也可作为高等院校有关专业数据库课程的教学参考书。

数 据 库 设 计

上海交通大学出版社出版

(淮海中路1984弄19号)

浙江上虞汤浦印刷厂排版

新华书店上海发行所发行

江苏太仓印刷厂印装

开本787×1092毫米 1/32 印张10.375 字数230000

1989年1月第1版 1989年2月第1次印刷

印数：1-3100

ISBN7-313-00306-4/TP·31 科技书目：182-249

定价：2.10元

计算机软件与应用丛书

(第一辑)

计算机数据安全原理	周锡龄	著
情报检索	王和珍	编著
数据库设计	施伯乐等	编著
程序设计方法学	朱三元	编著
微型计算机应用例解	白英彩等	编著

(第二辑)

计算机仿真	邱百光	编著
-------	-----	----

《计算机软件与应用丛书》编委会名单

顾问：刘振元 陈祥禄 朱雅轩

名誉主编：徐家福 杨芙清

主编：朱三元

编委：(按姓氏笔划排列)

白英彩 李玉茜 刘建明 何守才

陈涵生 陈娜芬 周锡龄 施伯乐

徐民祥 曹东启

序 言

当代社会是一个面向信息的社会,信息(数据)管理已逐步成为许多组织中最重要的活动之一。如何组织和管理数据,使它们最大限度地发挥应有的作用,是一个非常重要的问题。

数据库技术在计算机和信息科学领域中发展极快,借助于这门技术,各大企事业组织和政府部门中原来由人工进行的数据处理工作得以自动完成。

尽管数据库技术已发展到广泛使用的地步,但是如何有效地使用数据库系统,以便充分利用它的潜在能力,仍然是用户面临的一个重要问题。许多综述性文章和一些专著讨论了数据库技术理论,但是关于实际使用方法的讨论还很少。

本书的目的是介绍一些比较实用的方法和工具,使得设计数据库的过程变得较易理解,这对广大用户用数据库技术来解决实际问题将是有所帮助的。

本书供从事数据库实际工作的专业人员自学用,也可供在校师生作为教学参考书。使用本书时,读者需具有“数据结构”、“数据库引论”、“操作系统”等方面的基础知识。

本书第一、二、三章由潘锦平副教授编写;第四、五章由施伯乐教授编写;第六章由何守才副教授编写。施伯乐教授仔细审阅了全书。

由于编者水平有限,书中错误、不妥之处,恳请读者不吝指教。

编 者

1987年1月

目 录

第一章 绪言	1
1.1 什么是数据库.....	1
1.2 什么是数据库设计.....	5
第二章 网型数据库	8
2.1 系统简介.....	9
2.1.1 总体结构.....	9
2.1.2 数据模式.....	13
2.1.3 数据子模式.....	29
2.1.4 数据操纵语言(DML)	31
2.2 设计方法.....	34
2.2.1 实体联系模型(ER)	34
2.2.2 逻辑记录存取法(LRA)	47
2.3 设计工具.....	63
2.3.1 设计ER模型的一个图形工具	63
2.3.2 Gerritsen 系统.....	72
2.3.3 Mitoma 系统.....	98
第三章 层次型数据库	120
3.1 系统简介.....	120
3.2 设计工具——DBDA系统.....	125
第四章 关系数据库	144
4.1 关系数据库概述.....	145

4.2 关系数据库的基本概念	147
4.2.1 笛卡尔积与关系	147
4.2.2 域和属性	149
4.2.3 规范化与关键字	150
4.2.4 关系模式与操纵语言	152
4.3 关系代数	156
4.3.1 传统的集合运算	156
4.3.2 关系运算(关系操作)	158
4.3.3 几个检索例子	162
4.3.4 小结	163
4.4 关系演算	164
4.4.1 公式 ϕ 的定义	165
4.4.2 几个检索例子	166
4.4.3 存储操作	169
4.5 两种常见的查询语言——SEQUEL与QBE	170
4.5.1 SEQUEL	170
4.5.2 QBE	175
4.6 System R	178
4.6.1 System R 体系结构	179
4.6.2 System R 的管理系统	181
4.6.3 System R 的数据结构	185
4.6.4 其他功能	191
4.7 关系规范化的意义	197
4.7.1 函数依赖	197
4.7.2 第一、第二、第三范式	200
4.7.3 BC 范式和第四范式	207

第五章 关系数据库的设计	216
5.1 函数依赖(续).....	216
5.1.1 函数依赖的定义与公理系统.....	216
5.1.2 闭包和最小覆盖的计算.....	223
5.1.3 函数依赖集的覆盖.....	226
5.2 关系模式的分解.....	229
5.2.1 关系模式分解的定义.....	230
5.2.2 无损连接的测试.....	233
5.2.3 保持依赖的分解.....	238
5.3 关系模式的规范化.....	242
5.3.1 规范化的定义.....	242
5.3.2 两种分解的算法.....	243
5.4 多值依赖和第四范式.....	251
5.4.1 多值依赖的定义.....	251
5.4.2 关于函数依赖和多值依赖的公理系统.....	253
5.4.3 多值依赖的测试.....	257
5.4.4 第四范式.....	260
5.5 关系数据库的逻辑设计.....	263
5.5.1 引言.....	263
5.5.2 需求分析阶段.....	266
5.5.3 概念设计.....	268
5.5.4 逻辑设计.....	270
第六章 微型机上的dBASE系统	271
6.1 dBASE II 概况	271
6.1.1 dBASE II 的功能.....	271
6.1.2 dBASE II 的特点.....	272

6.2 dBASE II 基本概念	274
6.2.1 数据库文件	274
6.2.2 命令语句	276
6.2.3 文件类型	281
6.3 dBASE II 数据文件的创建	282
6.4 dBASE II 数据文件的操作	284
6.5 dBASE II 应用程序的编制	291
6.6 dBASE III 概述	295
6.7 dBASE III 的基本概念	297
6.8 dBASE III 数据文件的创建和修改	306
6.9 dBASE III 数据文件的操作	309
6.10 dBASE III 的其他特点及与 dBASE II 的转换	319

第一章 绪 言

1.1 什么是数据库

在现代社会中,数据是一种极为重要的资源。例如:生产计划、财务帐目、人事档案、车辆调度、仓库管理、旅馆房间和机票的预定、医院对病史资料的维护等等,无不需利用大量的数据。60年代出现的计算机文件系统,为人类有效地使用数据提供了自动化工具。但是随着数据量的增大,应用领域的推广,文件系统就显得难以适应了,数据处理人员逐步意识到,文件系统存在以下一些问题:

1. 数据冗余量很大

因为文件系统是根据每一应用程序的需要各自建立文件的,所以即使几个应用程序所需的数据有许多部分相同时,仍需建立各自的文件,这就造成大量的数据冗余,也就是说,数据不能共享。数据冗余既浪费存储空间,又使得数据的修改变得十分复杂,同时还易造成数据的不一致性,因此降低了系统的可靠性。

2. 数据和应用程序过分地相互依赖

因为文件是根据应用程序的具体要求设计的,所以数据结构和应用程序的编制密切相关,一旦数据的结构有改动,则必须相应地修改应用程序;反之,应用程序的扩充修改,也要求数据结构作出改动。这就使得系统的维护很不方便。

3. 缺乏对数据进行统一管理的方法

缺乏对数据的正确性、安全性、保密性等问题的控制手段。

针对文件系统的上述问题,在70年代,数据库技术迅速地发展起来,数据库技术的主要目标是:

1. 数据集中管理

即把某个企事业单位的数据资源集中起来统一管理,而让多个应用程序共享这些集中管理的数据,因此可以大大减少数据冗余的现象,并可统一考虑数据的合理性、安全性和可用性问题。

2. 数据独立性

数据独立性包括两方面:使应用程序和数据的具体存储或存取方式相对独立;使某个应用程序与其他应用程序所需的数据相对独立。前者称为物理数据独立性,后者称为逻辑数据独立性。显然,数据独立性越高,则系统越容易维护,因为数据存储方式的改变或个别应用程序的修改和扩充对已有的应用程序不会带来很大的影响。

当然,为了达到上述目标,数据库系统的开销要比文件系统大,但是由于计算机硬件价格的逐年下降,这一点已不成问题了。

图 1.1 是数据库系统的示意图。这里数据库(DB)是用户存储在计算机中的数据集合,而数据库管理系统(DBMS)是作为应用程序和数据库之间的界面的一个软件系统,它的作用是允许应用程序以抽象的方式来处理数据,即应用程序不必关心计算机是如何具体存储数据的。数据库管理系统是当今最复杂的软件系统之一。

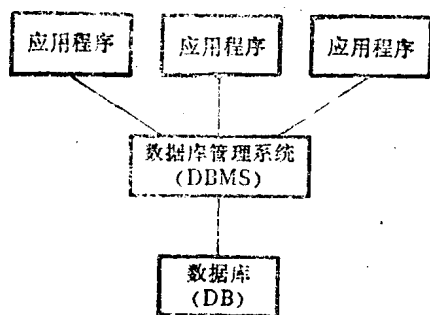


图1.1

为了对数据进行处理，数据库管理系统将数据库中的数据按一定的结构组织起来，数据结构的描述称为数据模型。现有的数据模型主要有层次型、网型和关系型三大类，所以数据库管理系统也按采用的数据模型相应地分三大类。

在层次型数据模型中，数据记录被组织成一个层次或树的结构，如图1.2所示。在这类模型中，记录间存在着“父子”关系。图1.2中的“部门”与“雇员”之间就为“父子”关系。在层次型模

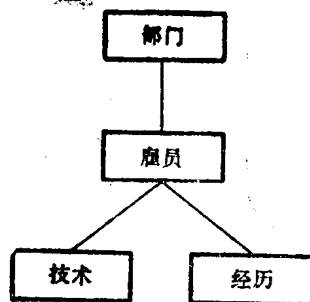


图1.2

型中,每个记录只能直接从属于另一个记录,而网型数据模型去除了这一限制。在网型数据模型中,数据记录被组织成一个网络,如图 1.3 所示,其中每个记录可从属于多个其他记录。由于任何层次结构或树都是网络的特例,所以层次型数据模型可以认为是网型数据模型的特殊情况。

在关系型模型中,数据被组织成表格形式,任何记录都不从属于其他记录(如图 1.4 所示),一些专门的关系操作可用来组合或选择记录,以达到修改和检索的目的。

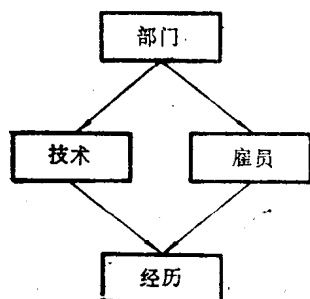


图1.3

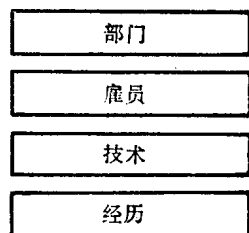


图1.4

已经证明,能被某类模型描述的数据也可用任一类其他模型来描述,所以某个具体问题究竟选用哪一类模型可根据用户的具体要求来决定。一般说来,层次型数据库的特点是简单,它特别适合于处理现实世界中层次状的信息结构。与关系型系统相比,网型系统具有较高的性能,但它的数据独立性较差;关系型系统的优点是数据独立性高,但在传统的计算机中,这类系统的速度较慢。近年来,联想存储器的出现将弥补关系型数据库

的这个缺陷。从数据操纵语言的角度来看,网型系统通常需由熟悉计算机程序设计的人员使用,而关系型系统则对不熟悉计算机技术的一般用户来说也是容易掌握的,所以很有发展前途。目前,关系型系统已是数据库领域中最活跃的一个分支。

1.2 什么是数据库设计

为一个特定应用环境建立数据库系统(如为民航局建立售票系统等)的过程称为数据库设计。

在1.1节中我们已介绍了数据库系统的基本组成,由于数据库管理系统是极其复杂的软件系统,而且目前软件市场上已有多种作为产品的DBMS出售,用户一般不必自行开发,所以数据库设计实际上是根据用户的具体要求选购合适的DBMS产品,并建立数据库(DB)和开发应用程序的过程。

数据库系统中,应用程序的开发技术同一般软件系统的开发技术是相同的,这是软件工程领域所关心的问题,所以数据库设计这一学术领域通常只着重讨论如何为特定的应用环境建立数据库的问题。

为用户进行数据库设计的目标是建立满足下列要求的一个数据模型:

(1) 能正确地反映特定的现实环境,这又包括两方面,一方面要能正确地反映用户的数据需求,另一方面又能支持用户对这些数据的加工需求。

(2) 能被某个现有的DBMS产品所接受。

(3) 能使系统具有较好的性能,如占用较少的空间,响应时间较快等等。

使用数据库系统的实践表明数据库设计是个很复杂的过程,设计人员既要详尽地理解用户的应用环境,又要了解选用的DBMS的具体特点和限制,还要考虑检索和修改的效率等等。特别是对一些大型的问题,名目繁多的数据项可达成百上千个,面对这样的问题,设计人员往往感到束手无策。显然,用临时凑合的方式进行设计不可能获得满意的结果,所以,如何为数据库设计建立科学的基础,并提供工程化的规范和技术就成为人们关心的课题了。在这样的背景下,到70年代后期数据库设计成为数据库领域中的一个热门课题。

为了探讨数据库设计的技术,1978年,数据库工作者在美国新奥尔良市召开了数据库设计会议,与会者比较一致地认为数据库设计过程(或称数据库设计生命期)应该分成四个阶段:1)分析用户需求;2)建立概念性数据模型;3)实现设计;4)物理设计。可以看出,这里前面两个阶段只是理解并确定问题,后面两个阶段才是寻找问题的解答。

分析用户需求主要是通过调查研究、阅读有关材料、面对面座谈等方式从用户那里获得数据库需要的信息,即数据需求、数据加工需求和各种限制条件。

在第一阶段的基础上,建立一个“概念性的”数据模型。概念性数据模型是将用户的数据需求精确地用书面形式表达出来,它描写的是从用户角度看到的数据库,而与数据库内部的具体存放等实现方式无关。

实现设计阶段首先要根据所收集的用户需求选购合适的数据库管理系统,然后将概念性数据模型转换成能被这个数据库管理系统接受的一个逻辑数据结构。所以也常把数据库设计的前三个阶段总称为“逻辑设计”。

物理设计是进一步设计数据模式的一些物理细节，如文件的基本结构、存取方式、索引的建立等。物理设计方案的选择不会影响系统产生所需输出信息的能力，而只影响处理的速度和所需的开销，所以物理设计本质上是优化问题。

一般数据库设计问题都可参照新奥尔良会议提出的框架进行，当然也要考虑到每个具体问题的特点来灵活地处理。由于人们对一个问题的认识总要经过去粗取精、去伪存真的过程，所以上述四个阶段的反复是不可避免的。

应该指出的是：数据库设计与应用程序的开发是紧密相关的。可以看出，数据库设计生命期的四个阶段同软件生命期的需求分析、建立系统说明书、总体设计和详细设计等阶段是相对应的，所以数据库设计必须与应用程序的开发并行地进行。

为了建立正确、高效、可靠、易于修改和易于扩充的数据库系统，数据库设计过程需要有设计方法的指导和设计工具的辅助。设计方法给出了规范的设计工序，使设计工作可以有计划、有步骤、纪律化地进行；而“设计工具”可以自动地完成设计过程中一些冗长繁琐的工作，因而能大大减少所需的人力和时间，并提高设计的质量。

70年代以来，数据库工作者陆续研究总结出多种设计方法，如 Chen 的ER方法、Smith、Kahn、Bubenko 等方法；也开发了许多工具系统，如DBDA系统、Gerritsen、Mitoma 系统等。本书后面各章将选择一些有实用价值、有代表性的方法和工具进行讨论。

第二章 网型数据库

网型数据库系统的代表是 DBTG 型数据库系统, 这是在 COBOL 文件系统基础上发展起来的一类系统。

COBOL 语言很适用于企事业管理中的一些数据处理问题, 在国外应用广泛。在文件系统逐步向数据库系统发展的时期, 负责为 COBOL 语言提供标准文本的组织 (CODASYL) 专门成立了数据库工作组 (DBTG), 其任务是为采用 COBOL 语言的数据库系统提出标准文本。这个工作组在 1971 年 4 月发表了一个标准文本, 它给出了下列标准:

模式数据描述语言 (schema data description language);

子模式数据描述语言 (subschema data description language);

数据操纵语言 (data manipulation language)。

由于有了标准文本, 许多计算机厂家先后推出了一系列以这一标准文本为基础的数据库系统, 如 UNIVAC 公司的 DMS-1100、Cullinane 公司的 IDMS、Honeywell 公司的 IDS 等。这些系统都部分地实现了 DBTG 的标准, 故称为是 DBTG 型的。

本章将讨论 DBTG 型数据库系统的数据模式、设计方法及其辅助工具。