

数据库系统导论

卷 I

〔美〕C. J. 戴特 著

科学出版社

数据库系统导论

卷 I

〔美〕C. J. 戴特 著

顾慧芳 董继润 译

罗晓沛 校

科学出版社

1986

内 容 简 介

本书系统地阐述了数据库的基本概念和一般技术,并反映了七十年代和八十年代初数据库技术的新方法和新成果。

本书以教科书形式编写,分两卷出版。本卷是数据库的基本部分,全书分五大部分,共二十八章。书中详细阐述了数据库技术的各个重要方面,数据库系统的结构、关系方法、层次方法、网状方法,以及三种数据库方法的回顾。各章附有习题,可使读者加深理解书中的内容。

本书可作为主修计算机科学和工程专业的学生和研究生教材,也可供其他学科从事计算机应用人员参考。

C. J. Date

AN INTRODUCTION TO DATABASE SYSTEMS VOLUME I

Addison-Wesley Publishing Company, 1981

2P50/09

数据库系统导论

卷 I

〔美〕C. J. 戴特 著

顾慧芳 董继润 译

罗晓沛 校

责任编辑 范铁夫 乐嘉敏

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1986年12月第一版 开本: 850×1168 1/32

1986年12月第一次印刷 印张: 19 3/8

印数: 0001-4,400 字数: 501,000

统一书号: 15031·758

本社书号: 4472·15-8

定价: 5.40 元

译者前言

随着社会主义现代化建设的进展,电子计算机在国民经济各个领域得到越来越广泛的应用。数据库是重要的计算机技术之一,它为计算机得到新的应用提供了可能性。

本书是数据库方面一本颇有影响的书,原书一九七五年出版第一版,一九七八年出版第二版,一九八一年出版的第三版(卷 I),本书是根据一九八一年第三版(卷 I)译出的。一九八二年出版的卷 II 是数据库的高级部分,专门讲授数据库的安全性和完整性。该书由科学出版社另行出版。可见,本书深受广大读者的欢迎。本书既介绍了成熟的传统方法,又及时反映了新颖的研究成果,不但使读者受到系统的学科训练,而且也能跟上数据库技术的迅猛发展。本书以教科书的形式编写,主要讲授数据库的关系,层次和网状三种方法,内容由浅入深,各部分又自成系统,以满足不同读者的需要。各章还附有思考题,可用来检验读者对各章内容的理解程度,引导读者开阔思路,了解基本原理的进一步应用。本书是数据库的主要部分,完全可以代替它原来的版本。

顾慧芳等同志参加了本书一九七八年版的翻译工作,并由仲萃豪、曹东启和施伯乐同志进行了校对。由于本书一九八一年版的出版,译者又根据新版重新进行翻译,并由罗晓沛同志审校。在此一并对审校者表示感谢。由于我们水平有限,译文难免有错,望读者批评指正。

第一版序言

计算机对我们社会的许多方面已经有了很大的影响。医药、法律、政府、银行、教育、交通、计划——这些只不过是计算机已经在起着重大作用的若干领域。可以预计，今后几年内计算机的应用范围还将大大增加；而且计算机对我们日常生活的影响也将相应扩大。为新的应用提供可能性的两个计算机技术领域（事实上，在多数情况下，它们是相当主要的）是远程通信和综合数据库。

在今后几年内，数据库系统将日益普及、并将变得越来越重要。目前，虽然已安装或正在研制的数据库系统的数目正在迅速增加，但它们仍是一个新兴的、尚需进一步探索的领域。因此，确实需要一本简明扼要论述这种系统基本概念的好基础教科书。本书就是试图满足这种需要而编写的。

假定本书的读者是对数据处理中的某些课题感兴趣的人。例如，系统分析员或设计员，应用程序设计人员，系统程序设计人员，听大学计算机科学课程或类似这种课程的学生，讲授这类课程的教员（实际上，本书就是按照作者长时期内在向IBM公司的职员讲授该学科的一门精读课程的教材编写而成的）。一般来说，本书要求读者对现代计算机系统的能力有相当的了解，特别是对于这种系统的文件处理特征有所了解。他们至少应该有一种高级程序设计语言的知识。然而，这些先决条件对他们来说并非是特殊要求。因此，对于使用或实现数据库系统的人，或对于只想扩大点计算机科学领域的一般知识的人，我希望本书可作为一本入门的教科书。

本书分为六大部分¹⁾：

1) 在第三版，关于安全性和完整性的材料已放到补充卷中。

1. 数据库系统的总体结构

2. 关系方法

3. 层次方法

4. 网状方法

5. 安全性和完整性

6. 评论、分析和比较

每一部分又分成若干章节。第一部分是数据库系统概念的一概介绍，并且特别概述了设计这种系统的三种不同的方法，即关系方法、层次方法和网状方法。接着，第二部分相当详细地阐述关系方法，第三部分阐述层次方法，第四部分阐述网状方法。第五部分讨论数据库系统的安全性和完整性问题。第六部分将书中前面介绍过的一些重要问题归并在一起，再稍作深入探讨。

对前面提出的结构需要作一些说明。例如，论述关系方法的第二部分，实际上主要是致力于介绍 Codd 的思想，E. F. Codd 博士是公认的关系数据库权威。但应该公正地指出，大多数现有（1974）的商用系统是采用另外两种方法中的一种而设计的。那么，为什么还要强调关系方法呢？对这个问题至少有两种回答。

（1）关系方法可看作数据理论的开端，它提供了理解和比较其它两种方法的极好基础，以及一种评价任何现有系统的合适的标准或尺度。即使关系方法不具有其他的优点，仅就它的理论上的正确性和永久性也会使它成为一个教学上的理想的方法。

（2）大部分现有系统不是采用关系方法这一事实，可以看作是计算机技术本身发展过程的自然结果。由于早期的直接存取设备容量较小，存取时间较长，传统上着重磁带和卡片这类顺序存储介质，计算机本身可用的存储容量的限制，以及类似的各种考虑，对大部分早期系统的最初设计有着重要的影响。而从现代的硬件和技术来看，似乎有可能设计和实现一个没有早期设计中那些缺点的系统。具体地说，很多专家现在都确信将来一定会看到一个或多个用关系方法实现的大规模系统（事实上，在本书出版之前，很多结合进关系概念的商用系统已开始出现）。

根据以上论述，也许读者很快就会得出结论：本书始终有点倾向于支持关系方法。作者认为这种倾向是完全有道理的；但是不向读者说明本书的这种倾向，那也是不诚实的。

尽管有刚才所表明的观点，但层次方法和网状方法显然还是很重要的，它们已经积累了多年的经验。原来的第三和第四两部分比较详细地论述了这些方法（我希望公平，不顾偏见）。第三部分完全是根据 IBM 公司现有的信息管理系统（IMS）阐述的，该系统已成功地在很多计算机设备上运行。把这个系统选为第三部分的基础是因为它是层次方法的一个极好实例，当然还因为它本身就是一个重要的系统。由于同样的理由，第四部分是根据 CODASYL COBOL 委员会数据库任务组（DBTG）的报告来写的。因此，我希望第三和第四两部分除了能作为层次和网状方法的一般内容介绍之外，特别还可作为了解 IMS 和 DBTG 系统的一个指南。但是，讲授具体系统并不是本书的主要目的，恰当地说，本书的目的是着重于介绍一般概念，其中涉及到一些专门系统主要也是用来说明某些概念（由于这种原因，对很多别的重要系统都没有花费过多的笔墨去介绍）。即便如此，对 IMS 和 DBTG 的描述还是特地深入到了大量的专门细节¹⁾。对这些系统的细节不太感兴趣的读者，如果愿意的话，可以略去该教材的某些部分——主要是第十九到第二十二章，以及第二十四章中（适当指出的）某几节。

对有关术语的说明：

数据库系统领域同其它很多新兴学科一样，还没有公认的术语表。特别是，IMS 的术语很多方面与 DBTG 的术语是不同的。对于这些差别，本书企图在第一和第二两部分中根据 IMS 术语和 DBTG 术语定义出一种“中间”术语去加以消除（但是一旦这样做了之后，那么在以后的讨论中，对每个系统就应普遍使用这种“矫正”的术语）。反过来讲，第一和第二两部分中的术语也是从多种来源中综合出来的。

1) 在第三版中，对系统 R 也是如此。

关于本书结构的几点说明:

(1) 我一直想写一本教科书而不是一本参考书。当然这两个目标并非是完全不相容的。的确,我希望在很大程度上能兼顾两者,但一旦两者之间发生冲突,那末我们的目标仍是写一本教科书而不是参考书。为此目的,我断然删去了某些次要部分以使主要部分更为清晰,根据同样的理由,也对另外一些部分进行了简化,尽管如此,我仍尽可能把材料写得详细些(读者如要更详细地了解,可适当地参阅其它有关书刊文献)。

(2) 因为这是一本教科书,所以大部分章的后面都附有一组习题,我热切地希望读者尽量去做这些习题中的几个。答案可在书末找到,这些答案有时对所提问题能提供更多的知识。

(3) 每章之后都附有参考文献,其中许多文献都加了注解。正文中用方括号括起来的数字表示所要参考的文献。例如,[1.3]指第一章末尾参考文献表中的第三条,即 CODASYL 系统委员会发表在 BCS Computer Bulletin, Vol.15, No.4 和 Communications of ACM Vol.14, No.5 上的那篇论文。

我还要对书写过程中所得到的帮助表示感谢。首先感谢 Codd 博士对我的鼓励,感谢他允许我使用他发表过的很多资料(特别是本书的第二部分),并感谢他对初稿提出的宝贵意见。下面这些人也非常热情地阅读了草稿并提出了许多宝贵的批评和建议: Joel. Aron, Jan. Hazelzet, Roger. Holliday, Paul. Hopewell, Larry. Lewis, Salah. Mandil, Bill. McGee, Herb. Meltzer, John. Nicholls, Terry. Rogers, 和 Tom. Work。我还应该感谢 Julius. T. Tou 教授,他是第四届计算机和信息科学国际讨论会(Miami Beach, Florida, 1972年12月14—16日)的组织者,还要感谢 Plenum 出版公司(本次会议录的出版者),他们同意我使用那次讨论会上我提出的一篇论文作为第三章的基础。也应感谢 IBM 公司的很多学者,他们对本书原始教材所提的意见,对促成本书是很有帮助的。最后,我要感谢 IBM 公司,准许我利用该公司的时间和人力、物力以完成撰写本书的大量工作。

但我应该强调，本书的内容由我负全部责任；书中的观点只表明我自己的看法，并不代表 IBM 的正式意见。

C. J. 戴特 1974年11月

(1981年3月修改)

第二版序言

自第一版发表以来,数据库研究领域已经发生了许多变化。DBTG 数据操作语言和子模式数据描述语言已被 CODASYL COBOL 委员会接受而合并进 COBOL 中,而且一些以 DBTG 为基础的系统现已商品化。IMS 已加上了辅助索引和各种其它特征。以关系概念为基础的商品系统已开始投入使用。各种标准化活动正在进行。意义深远的也许是世界上很多大学和类似的机构对该学科正在显露出一种前所未有的兴趣。本版就打算反映一些这方面的活动。其中包括了大量的新材料,这是本书再版的主要原因;当然,也趁此机会校正了第一版中的某些错误,并改进了若干段落的描述。本版还列入了许多新的参考文献,并多半都加了注释。

这一版与前一版的重要区别可概括如下:

第一部分:为了引进 ANSI/SPARC 术语,对整个系统结构作了修改.采用对比叙述方式把三种方法统一起来,并作了补充。

第二部分:把关系数据结构的讨论扩展为单独一章;另外还增加了两章,分别讨论 SEQUEL 和 Query By Example;关于进一步规范化的一章已全部重写,并包括了第三范式的改进处理方法和新的第四范式。所有其它章节也都作了较大的修改。

第三部分:增加了关于辅助索引的一章,并根据 IMS 最近的版本对其它各章作了相应的修改。

第四部分:为了体现 CODASYL 数据描述语言和 COBOL 委员会所作的改变,对所有的章节都作了修改。

第五部分¹⁾:为了体现 IMS 和 DBTG 的变化,对这两章都作了修改。除此之外,还收入了另外一些关系系统。并且,对有

1) 在第三版中,这些材料已放到补充卷中。

关完整性约束和并发性的论述作了很多补充。

第六部分：这一部分完全是新编入的。

再次为编写本书所得到的帮助表示衷心的感谢。特此感谢对第一版作出好评，并鼓励我扩大、改写第二版的那些人。在这一点上，应该特别感谢 Ted. Codd, Frank. King, Ben. Shneiderman, 和 Mike. Stonebraker. 在一些技术问题以及对本版原稿的许多部分的审阅及评论工作上，我还得到了下面这些人的大力帮助：David. Beech, Don. Chamberlin, Rod. Cuff, Bob. Engles, Ron. Fagin, Peter. Hitchcock, Roger. Holli-day, Bill. Kent, Bill. Lockhart, Ron. Obermarck, Vern. Watts, 以及 Moshe. Zloof, 象前一版一样，我要特别感谢 IBM 公司对我工作的支持。还要感谢出版“Datamation”杂志的“技术发行公司”准许我引用1976年4月该刊的一篇文章作为改写第三章的基础，感谢 ACM 准许我在改写第六部分时引用 ACM 拥有版权的三篇文章（参考文献〔28.1〕，〔27.1〕和〔28.3〕）中的材料。最后但却是非常重要的，我要感谢 Addison-Wesley 公司在本书（第一版和第二版）的全部出版过程中所表现出的巨大热情，鼓励和耐心。

C. J. 戴特 1977年6月

（1981年3月修订）

第三版序言

数据库技术领域以不断增长的速度在继续发展中。事实上，已经到了这样的程度，其中几个不同的子领域正开始作为独立学科而出现。同时，有迹象表明，一些早期建立起来的领域正在以一种比较稳定的状态逐渐被确定下来。尽管本书新版的时机已经成熟，但遗憾的是想在一卷内既要保持原书的篇幅又要充分论述这一课题已是不可能了，因此，现将材料大致分为“基本的”和“高级的”两大部分；本卷包括“基本的”材料，而“高级的”材料挪到预定的补充卷中。即使这样，本卷还是形成了一本完整和自成体系的书。实际上，它在很多方面完全可以取代以前的版本；只有安全性和完整性方面的论述例外(第二版的第五部分)，在某种意义上讲，这部分内容与前版的其余部分比较起来已是不太“基本的”了，所以已把它挪到补充卷中，而在本卷中通过引进一些更新的材料，特别是大大扩充关系方法的论述，以补充被删去了的这部分内容。

这一版和它前一版的主要差别综述如下。

可以看到：就第一，二，三和第五部分中变化的内容来说，大多数都是对前一版书中内容的澄清或补充（有的是非常重要的补充）。相反，第四部分的变化则是很多实在的变化（这反映了相应原始文献的变化）。换句话说，第二版第四部分现在已被认为是过时了的；而第一，二，三，和第五部分的内容，就现在的情况来说仍是相当正确的，但也不能认为是最新的。

第一部分 第一章已作了修订，提供了对主题的更完美和全面的介绍，并给出了 ANSI/SPARC 结构的更好描述。第二章的内容已扩充为包括 B 树的讨论，并扼要地介绍了可扩充散列。

第二部分 这部分几乎完全是重写的。第四章经过扩充更好

地解释了关系结构，尤其是加进了两种基本的完整性规则。第五一十章(取代第二版的第七和第十章)全面描述了关系系统R，用它作为说明各种关系系统概念的工具¹⁾。第十一章 Query By Example (第二版第八章)已修订，使它与IBM公司发行的QBE产品相一致。第十二章(第二版第六章)比以前更详尽地论述了关系代数。十三章(第二版第五章)比前版更为形式地介绍了基于元组和域的关系演算，并在它的几个例子中采用了QUEL (INGRES系统的语言)，而不再采用DSL ALPHA，因为DSL ALPHA从未实现过。十四章(第二版的第九章)扩充了内容，包括：好的和不好的分解的概念，“最终的”第五范式，以及第四范式的更好处理。最后，删除了第二版中的第十一章(“某些关系系统”)，但该章的大部分内容仍以参考文献注解的形式保留了下来。

第三部分 这部分的内容扩充为包括片段级可感知的材料(第十七章)，多PCB的使用(第十八章)，以及有关快速路径数据库新的一章(第二十二章)。

第四部分 这部分材料根据最近的CODASYL规范，以及ANS COBOL委员会(X3J4)和ANS数据描述语言委员会(X3H2)的工作文件进行了全面修订。另外，对习题解答也作了较大的改进。

第五部分(第二版第六部分) 第二十七章详尽地论述了统一数据库语言UDL，大大扩充了前版的25.2节。第二十八章是对前版25.3节的扩充。

还有一个变化，即在前两版中常用的一个术语“数据模型”，本书中基本上不用了。遗憾的是，该领域的工作人员赋予该术语以各种不同的解释，因此已成为某些概念混乱的根源。在这方面，我不认为自己是无可责怪的。在十二章中仍出现了这个术语，其

1) 正当本书将付印时(1981年1月)，IBM公司宣布了一种名为SQL/DS的新数据库程序产品。SQL/DS是在DOS/VSE操作系统支持下工作的，它几乎包括了系统R的全部特点。

含义与前二版中是不同的。遗憾的是，数据库系统的术语目前仍处于不断变化的状态中。

我非常感谢许多朋友和同行，他们通过对这一版草稿的各部分提供建议和审阅，并提出意见给了我很大的帮助。尤其是感谢 David. Beeth, Don Chamberin, Ted. Cood, Bob. Engles, Ron. Fagin, Bill. Kent, Pete. Lazarus, Jim. Panttaja, Franz. Remman, Reind. Van de Riet, Bob. Smead, 并特别感谢 Phil. Shaw, 他完成了任何一个作者都所希望的审阅手稿的全部任务。我也应感谢 Karen. Takle. Quinn 和 Per. Groth 协助我使用 IBM 公司 Santa Teresa 图书馆的 QBE 系统，并感谢 Paul. Pittman, Russ. Williams, 和系统 R 研制组的成员帮助我用系统 R 进行实验。如同前两版一样，我要感谢在写本书时 IBM 公司所给予的支持。我也应感谢 ACM 允许我引证 ACM 拥有版权的三篇论文（参考书目 [1.14], [4.6] 和 [5.1]）中的材料。最后，我再次愉快地感谢 Addison-Wesley 公司全体人员在本书出版过程中所表示的友谊、合作和严谨的工作作风。

C. J. 戴特

1981年3月

目 录

第一部分 数据库系统结构

第一章 基本概念	2
1.1 数据库系统是什么?	2
1.2 工作数据	5
1.3 为什么需要数据库?	8
1.4 数据独立性	11
1.5 数据库系统的总体结构	16
1.6 分布数据库	27
习题	28
参考文献和提要	29
第二章 存储结构	33
2.1 引言	33
2.2 一些样本数据的可能表示方法	36
2.3 物理记录接口: 索引技术	45
2.4 一般索引技术	51
习题	55
参考文献和提要	56
第三章 数据结构和相应的操作	62
3.1 引言	62
3.2 关系方法	63
3.3 层次方法	66
3.4 网状方法	70
3.5 高级操作	73
3.6 小结	78
习题	80
参考文献和提要	80

第二部分 关系方法

第四章 关系数据结构	82
4.1 关系	82
4.2 域和属性	84
4.3 关键字	87
4.4 外延和内涵	90
4.5 小结	91
习题	93
参考文献和提要	93
第五章 系统 R 的结构	95
5.1 背景	95
5.2 结构	97
参考文献和提要	103
第六章 系统 R 的数据结构	105
6.1 引言	105
6.2 基本表	105
6.3 索引	108
6.4 讨论	110
习题	112
参考文献和提要	113
第七章 系统 R 的数据操作	114
7.1 引言	114
7.2 检索操作	114
7.3 库函数	133
7.4 更新操作	138
7.5 系统 R 字典	141
7.6 讨论	143
习题	145
参考文献和提要	147
第八章 嵌入的 SQL 语言	149
8.1 引言	149

8.2	不涉及位置指针的操作	151
8.3	涉及位置指针的操作	153
8.4	动态语句	158
8.5	讨论	162
	习题	163
	参考文献和提要	163
第九章	系统 R 外层结构	165
9.1	引言	165
9.2	视图	165
9.3	对视图的DML操作	168
9.4	视图和数据独立性	173
9.5	小结	175
	习题	176
	参考文献和提要	176
第十章	系统 R 内层结构	178
10.1	关系存储系统	178
10.2	段和页	178
10.3	文件和记录	180
10.4	存取路径	181
10.5	例子	184
10.6	RSS号码表	187
	参考文献和提要	187
第十一章	QUERY BY EXAMPLE	189
11.1	引言	189
11.2	检索操作	191
11.3	树结构关系上的检索操作	196
11.4	库函数	200
11.5	更新操作	202
11.6	QBE字典	205
11.7	讨论	210
	习题	211
	参考文献和提要	211