

数据库

——设计、性能与管理结构方法

知识出版社

数 据 库

——设计、性能与管理的结构方法

〔美〕S. 埃特里 著
康金玉 张承庆 译
许成钢 校

知 识 出 版 社

JS/30/25

内 容 简 介

本书系统地阐明了数据库管理系统,明确地规定了数据库管理与使用人员的任务,详细讲述了概念数据库、逻辑数据库和物理数据库的设计过程,并结合具体实例详细讨论了这些方法。

本书立论清晰、通俗易懂,可供工业、商业、院校和研究部门以及医疗、财政、政府等使用计算机的专业人员和管理人员阅读,也可作教科书或参考书。

本书曾得到王慧敏、梁俊华同志的帮助。

数 据 库

——设计、性能与管理的结构方法

〔美〕S·埃特里 著

康金玉 张承庆 译

许成钢 校

知 识 出 版 社 出 版

(北京安定门外外馆东街甲1号)

新华书店北京发行所发行 北京景山学校印刷厂印刷

开本787×1092 1/32 印张14.75 字数250千字

1987年6月第1版 1987年6月第1次印刷

印数: 1—6,300

统一书号: 15214·16 定价: 3.50元

ISBN: 7-5015-0046-0/TP.2

前 言

计算机革命看来可以用甚至连空想家都梦想不到的各种方式简单、轻巧地增添社会活力。在各种各样的部门中,用计算机存储和检索大量的信息。工业、医疗、财政和政府部门所处理的数据量正以几何级数急剧地增加,很难想象如果没有计算机,这些部门将如何发挥作用。这里我们将各类组织都称为“企业”。

称作“数据库”的信息仓库,是企业的整个计算机系统的基础。因此数据库的设计和性能大概是信息处理中最重要的两个方面。

虽然已经出版了一些对该领域进行学术论述以及对现有数据库管理系统和数据模型作全面评述的文章和书籍,但却几乎没有明确回答下列问题:

1. 数据库有关人员的职责是什么?
2. 怎样能够设计出符合性能标准的数据库?
3. 在不同的阶段会出现什么问题?

本书打算回答这些问题,并提供有关的信息。本书说明了设计满足信息处理环境所需要的功能和性能的数据库的原理。提出了数据库管理员、系统程序员、系统分析员、应用程序员、信息处理管理员、数据处理管理员和数据库用户的任务。本书内容简单但又不过于简单,它给出逐步设计数据库的方法。章节的组织方式使得数据库的设计者能按着它们的

顺序设计数据库。本书的方法与任何软件包无关，它不集中介绍任一厂家的数据库管理系统。

本书可供商业和政府中的实际工作者有效地使用，也可以作为商业或技术学校一个学期的课程，那里的学生对“为什么”和“怎么样”使用信息处理方法很感兴趣。

本书采用实例研究方法。在附录 A 中，对设计一个银行环境的数据库从头到尾进行了实例研究。附录 B 则完整地介绍了大学环境的实例研究。逐步设计数据库的方法以及将此方法应用于各个实例研究是本书的主要特色。

本书介绍的材料是从我在 IBM 系统研究所教授多年的课程中发展起来的。作为课程教材的手稿，其标题与本书相同，都是“数据库：设计、性能和管理的方法”。本书的部分内容曾在各种专业讨论会，包括计算机协会 (ACM) 的地方分会和数据处理管理协会 (DPMA) 上介绍过。来自学生和同事们的问题和建议使得本书内容更适时、更容易消化。而在讲课时出现的热烈反应使我决定将这部手稿作为一本书出版。

本书分为三篇，每篇包括若干章。第一篇：“数据库的管理”，论及怎样建立和维护数据库。第一章通过讨论的形式介绍了后面要探讨的论点。第二章叙述数据库管理员 (DBA) 的职责以及 DBA 和他 (或她) 的职员在信息处理系统中所起的作用。标题为“数据字典”的第三章讨论收集有关数据库环境信息的有用的方法。

第二篇“数据库设计 (概念模型和逻辑模型)”阐述了概念数据库和逻辑数据库的设计过程。第四章建立了数据库管理系统——数据模型的基础。对三种流行的方法——关系的、层次的和网状的数据模型作了简明的论述。第五章介绍概念模

型的设计过程。第六章介绍逻辑模型的设计过程。这两章都结合具体实例详细地说明了这些方法。

第三篇“数据库性能”讲述了采用数据库信息处理性能的几个同等重要方面。主要是强调如何设计出好的性能，也就是说如何使性能最佳化。第七章概述了数据库存取方法，这是性能的关键所在。第八章介绍了这些存取方法的一些实现方案。第九章用空间和时间的估计来讨论物理模型的数据库设计。第十章就设计、实现和运行阶段的具体性能问题进行了讨论。

附录 A 是银行环境的实例研究。附录 B 则为大学环境的实例研究。最后附录 C 给出了第九章所用的概率论和数理统计的基本概念。

目 录

第一章 数据.....	(1)
1.1 数据源.....	(1)
1.1.1 企业.....	(2)
1.1.2 实体.....	(3)
1.1.3 属性.....	(3)
1.1.4 数据值.....	(3)
1.1.5 关键数据元.....	(5)
1.1.6 数据记录.....	(5)
1.1.7 数据文件.....	(5)
1.2 数据文件环境.....	(7)
1.2.1 通用例行程序的优点.....	(8)
1.2.2 常规数据文件环境的缺点.....	(10)
1.3 数据库环境.....	(12)
1.3.1 什么叫数据库?	(12)
1.3.2 数据库管理系统.....	(13)
1.3.3 数据集中的缺点.....	(13)
1.4 数据库管理员.....	(15)
1.5 数据独立性.....	(17)
1.5.1 什么是数据独立性?	(17)
1.5.2 数据独立性的两个阶段.....	(19)
1.5.3 怎样实现数据独立性.....	(21)

1.6 数据字典.....	(22)
1.7 数据库设计和性能预检.....	(23)

第一篇 数据库管理

第二章 数据库的管理.....	(28)
2.1 数据库管理的职能.....	(28)
2.1.1 数据库管理员(DBA)的责任.....	(29)
2.1.2 DBA和管理员.....	(32)
2.1.3 DBA和用户界.....	(33)
2.1.4 DBA和应用开发人员.....	(35)
2.1.5 DBA和系统人员.....	(37)
2.1.6 DBA和操作人员.....	(38)
2.1.7 DBA和软件卖主.....	(39)
2.1.8 DBA和硬件卖主.....	(41)
2.2 数据库系统的生命周期.....	(42)
2.2.1 数据库的设计.....	(45)
2.2.2 数据库的实际建立.....	(47)
2.2.3 现有数据集的转换与适应新建的 数据库的应用.....	(49)
2.2.4 把转换的应用和新的应用集中到 新的数据库.....	(49)
2.2.5 操作阶段.....	(50)
2.2.6 发展、变化和维护阶段.....	(50)
2.3 DBA、DBA工作人员及其责任.....	(51)
2.3.1 DBA在组织中所处的位置.....	(51)
2.3.2 DBA的工作人员.....	(54)

2.3.3 DBA各成员的责任.....	(55)
第三章 数据字典.....	(63)
3.1 什么是数据字典.....	(63)
3.1.1 数据字典的基本目的.....	(64)
3.1.2 数据库管理系统和数据字典.....	(67)
3.1.3 数据字典及其接口.....	(69)
3.1.4 理想数据字典: 要求和组织.....	(71)
3.2 实现数据字典的策略.....	(75)
3.2.1 成本-效益问题.....	(75)
3.2.2 何时实现数据字典.....	(79)
3.2.3 数据定义准则.....	(81)

第二篇 数据库的设计 (概念模型 和逻辑模型)

第四章 数据模型.....	(86)
4.1 什么是数据模型.....	(86)
4.2 数据模型内的关系.....	(88)
4.2.1 “一对一”关系(两个实体之间).....	(88)
4.2.2 “一对多”关系(两个实体之间).....	(88)
4.2.3 “多对多”关系(两个实体之间).....	(89)
4.2.4 “一对一”关系(两个属性之间).....	(90)
4.2.5 “一对多”关系(两个属性之间).....	(90)
4.2.6 “多对多”关系(两个属性之间).....	(90)
4.2.7 数据模型的概况.....	(91)
4.3 关系数据模型.....	(94)
4.3.1 关系数据模型的优点.....	(98)

4.3.2 关系数据模型的缺点.....	(99)
4.4 层次数据模型.....	(100)
4.4.1 层次树结构.....	(101)
4.4.2 层次数据模型的存储操作.....	(109)
4.4.3 层次数据模型的优点.....	(110)
4.4.4 层次数据模型的缺点.....	(112)
4.5 网状数据模型.....	(112)
4.5.1 一对多关系的表示.....	(116)
4.5.2 另外三种系型.....	(121)
4.5.3 用网状数据模型存储操作.....	(124)
4.5.4 网状数据模型的优点.....	(125)
4.5.5 网状数据模型的缺点.....	(125)
第五章 数据库设计(概念模型).....	(129)
设计数据库的概念模型.....	(129)
5.1 数据分析.....	(130)
5.1.1 收集有关现有应用的数据信息.....	(130)
5.1.2 收集有关未来应用的数据信息.....	(134)
5.2 设计过程的关系概念.....	(135)
5.3 图形表示法.....	(156)
5.4 设计过程.....	(157)
第六章 数据库设计(逻辑模型).....	(165)
数据库逻辑模型的设计.....	(165)
6.1 映射到关系数据模型.....	(166)
6.2 映射到层次数据模型.....	(172)
6.3 映射到网状数据模型.....	(189)

第三篇 数据库性能

第七章 数据的存储和存取方法	(200)
7.1 用户和数据库之间的接口.....	(200)
7.2 内部模型(物理模型)存取方法.....	(202)
7.2.1 物理顺序法.....	(204)
7.2.2 索引顺序存取法.....	(205)
7.2.3 索引随机法.....	(210)
7.2.4 倒置法.....	(213)
7.2.5 直接法.....	(214)
7.2.6 散列法.....	(215)
7.3 外部模型(用户视图)存取方法.....	(218)
7.3.1 两个记录之间的关系.....	(219)
7.3.2 外部模型存取方法.....	(219)
第八章 存取方法的实现	(228)
8.1 层次数据库管理系统.....	(228)
8.1.1 IMS (信息管理系统).....	(228)
8.2 网状数据库管理系统.....	(249)
8.2.1 CODASYL —— 数据库任务组	(249)
8.3 倒置文件数据库管理系统.....	(271)
8.3.1 ADABAS(适用数据库系统).....	(272)
第九章 数据库设计(物理模型)	(284)
9.1 数据库设计的两个步骤(物理模型).....	(284)
9.2 设计数据库的物理模型.....	(286)
9.2.1 层次数据库管理系统.....	(287)
9.2.2 网状数据库管理系统.....	(291)

9.2.3 倒置文件数据库管理系统	(297)
9.3 数据库物理模型的评价	(306)
9.3.1 存储空间的估计	(306)
9.3.2 时间估计	(314)
9.3.3 数据库的应用	(320)
第十章 性能问题	(324)
10.1 实现	(324)
10.1.1 物理数据库的建立	(324)
10.1.2 转换与合并	(327)
10.2 操作	(329)
10.2.1 后援和恢复	(329)
10.2.2 重组	(332)
10.2.3 重构	(334)
10.2.4 性能监视和调整	(337)
10.2.5 数据库环境中的安全性	(344)
10.2.6 数据库环境中的保密性	(350)
附录 A 为银行环境设计一个数据库	
“大众银行”——实例研究 1	(360)
附录 B 为一个大学环境设计的数据库	
“大学”——实例研究 2	(412)
附录 C 基本概念：概率论和数理统计	(458)

第一章 数 据

一个企业赖以生存和发展所必须的各种信息是它的命根子。本章介绍这些信息的来源,以及数据库系统如何为了适应对信息进行数据处理的需要而发展起来的。一个成功的数据库所必备的一个特点是它的应变能力。高度的应变能力可以使数据库不太介意信息处理要求的变化。本章将讨论应变力的有关问题,并预先概述设计有适度性能的数据库的步骤。

1.1 数 据 源

在任何文明社会中,信息处理都是一项必不可少的工作。由于经济与人口的增长,越来越需要对彼此有关的大量商业数据和行政事务数据进行统一管理。相互有关的各种数据集中在一起就称为系统。任何一个系统如果要大大减轻人们的工作负担,它除了应当拥有一般的知识以外,大抵来说还需要有一个能反映真实世界的、十分复杂的模型。

任何信息系统的最关键部分都是它所存储的数据。对于一个企业来说,这些数据通常是跨越不同部门的。例如,管理决策所需要的信息经常来自整个企业。至于生产过程的决策,则需要有存货信息,订货信息和销售策略等。这就意味着

有关整个企业的那些数据，要以简单而容易存取的形式存储。

在电子数据处理初期，数据都存储在纸带或穿孔卡片上。由硬件设备来读这些纸带和卡片；而由编好的应用程序来处理数据，并把数据变成一份份的报告。应用程序的源语言是由编译程序翻译成机器指令的。为了协调各种电子数据处理活动，发展了操作系统。随着时间的推移，操作系统、编译程序、设备以及管理电子数据处理系统的各种规章制度变得日臻完善了。

然而，基本的目的并没有变。目前，数据处理系统可以是老式的，那只不过是办事员在活页夹里写入各种数据；也可以是新式的，使用了最快速的计算机和现有的各种最先进的硬件与软件。而基本目的却始终都是以最合算的成本适时地在规定的地点向需要信息的人提供准确的信息。企业一词可以是指银行、工厂、公用事业公司、大学、医院、超级市场、百货商店或农场等等。

要掌握电子数据处理技术，必须弄懂用来描述数据的一些术语以及数据表示方法。

1.1.1 企业

企业是任何一种组织，例如银行、大学、工厂或医院。

问题在于如何区分整个企业（例如象汽车厂、计算机厂、化工企业或钢铁厂那样的大公司）和大公司中的组织机构。这些组织机构可以代表一个企业（如一个大汽车厂的车身分厂、计算机厂的数据处理分厂）。在这种情况下，可以把这些分厂本身称为并看作是企业。

许多企业会需要下列一种或几种信息：人员、工资单、存货单、采购定货单、应收帐款、销售报告、实验室测试数据、课程登记、金融交易、病历表。这些项目的信息关系到人员、地点、事物、事件或概念。

1.1.2 实体

实体是记录了有关信息的人物、地点、事物、事件或概念。

银行环境中实体的例子是 CUSTOMERS (顾客)、BANK ACCOUNTS (银行帐户) 和 MORTGAGE LOANS (抵押贷款)。仓库中的实体是 SUPPLIERS (供应厂商)、PARTS (零件)、SHIPMENTS (装货) 等等。

1.1.3 属性(或数据元)

每个实体都有一些用来描述其特征的基本属性。一座房子可以通过大小、颜色、年代及周围情况来描述。银行的顾客也可以用象姓名、地址、还可能有顾客识别号等属性来描述。银行帐户可以用帐目类别、帐号和帐户结余等来说明。顾客识别号、顾客姓名和顾客地址是用来描述 银行 “CUSTOMER” (顾客) 这个实体的三个属性 (见图1.1)。通常把属性称为数据元、数据域、域、数据项或基本项。

1.1.4 数据值

数据值是每个数据元中包含的实际数据或信息。数据元“顾客姓名”可以取“希金斯教授”和“马克·吐温”这样的值。数据元所取的值可以是定量的, 定性的或描述性的, 这取决于数据元描述实体的方式。图1.1表示实体、数据元和数据元的

取值之间的差别。

实体	属性 (或数据元)	值 (数据)
顾客	顾客识别号	123456789
	顾客姓名	希金斯教授
	顾客地址	纽约斯诺街 55 号
帐户	帐目类别	支票
	帐号	534250
	帐户结余	20.53
供应厂商	供应厂商名	威廉姆·莎士比亚
	供应厂商号	S1
	供应厂商地址	纽约布克街 100 号

图 1.1 有关企业的信息可以表示为若干实体，而一个实体又可以表示为若干数据元。称数据元取的值为数据。

有关企业的信息可以表示为若干实体，而一个实体又可以表示为若干数据元。称数据元取的值为数据。一个实体的所有数据元所取的一组值称为一个实体事件。这些实体通过一定关系相互连接。把用数据元表示的实体和连接实体的环节共同构成的内在模型称为概念模型。这种概念模型给出了企业中数据流程的总视图。

常常由于不能区分实体、数据元和数据元的取值而使关于“数据”的讨论产生混乱。关于顾客，有实体 CUSTOMER，有一些表示顾客的数据元（例如，顾客标识号、顾客姓名和顾客地址），还有数据元的取值，如“123456789”，“希金斯教授”以及“纽约，斯诺街55号”。因此区别实体，数据元和数据元的取值是十分重要的。

1.1.5 关键数据元

某些数据元具有这样的特点，即已知实体的某一特定数据元的取值，便可以辨认出同一实体的其它数据元的取值。已知顾客的识别号为“123456789”，我们便可以确定该人是希金斯教授，他在银行有一支票帐户。我们把从它们可以推出其它数据元的那些数据元称为关键数据元，关键数据元也可以叫做实体标识符。

有可能存在着两个或更多个能唯一标识一个实体的数据元，这时我们称这些数据元为关键数据元的“候选者”。由人（用户或设计者）确定用哪个候选者来存取实体。关键数据元的设计应当十分仔细，因为适当的选择可以帮助人们正确地设计概念模型。

1.1.6 数据记录

数据记录是有关数据元的取值的集合。从图1.2 可以看到数据元包括顾客姓名、顾客识别号、帐目类别、…、备注。“Prof. Higgins 123456789 Checking...Notify Customer Tel 555-1234”就是其中一个数据记录。数据元为实体 CUSTOMER、ACCOUNT和 TRANSACTION 取值。这些值构成数据记录。数据记录存储在某种介质上。介质可以是人脑、纸张、计算机存储器，或是计算机的辅助存储设备如磁带、磁盘、或磁鼓，仅举这几个例子说明。

1.1.7 数据文件

一些数据记录构成数据文件，因此，数据文件是数据记录的有序集合。图1.2“分行经理报告（例外）”就是打印形式的