数 据 库 系 统 基 础

王珊 著　　罗晓沛 审

中 国 铁 道 出 版 社

1998年·北京

需要全本请访问 www.ertongbook.com

(京)新登字 063 号

图书在版编目(CIP)数据

数据库系统基础/王珊著.- 北京:中国铁道出版社,1998.1
ISBN 7-113-02912-4

Ⅰ.数… Ⅱ.王… Ⅲ.数据库系统-理论 Ⅳ.TP311.13

中国版本图书馆 CIP 数据核字(98)第 02681 号

书　　名:数据库系统基础
著作责任者:王珊
出版·发行:中国铁道出版社(100054,北京市宣武区右安门西街 8 号)
策划编辑:严晓舟
责任编辑:严晓舟
封面设计:陈东山
联系电话:63549447
印　　刷:北京市燕山联营印刷厂
开　　本:850× 1168 1/32 印张:6.875 字数:173 千
版　　本:1998 年 4 月第 1 版第 1 次印刷
印　　数:1—8000 册
书　　号:ISBN 7-113-02912-4/TP·288
定　　价:15.00 元

版权所有 盗印必究

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社发行部调换。

序　　言

数据库是数据管理的最新技术,用以存储、管理各单位宝贵的信息资源。自60年代第一个数据库管理系统问世,至今虽然只有30多年,但数据库技术不论是在理论上、实现技术上还是应用上都取得了巨大的发展。现代数据库技术已成为信息系统和计算机应用系统的核心。无论是一个单位的财务系统、一个工厂的库存管理系统、一个学校的学籍管理系统,还是一个部门的综合信息管理系统、期货交易系统乃至全国范围内的行业信息系统(如全国税务征管系统),都离不开数据库系统的支持。特别是在信息化时代来临,Internet、Intranet高速发展的今天,信息资源具有的经济价值和社会价值越来越明显,越重要。正因为数据库技术和经济、社会的发展,与信息化建设有着密切的关系,使得这门技术、这个学科获得了巨大的原动力和深厚的应用基础。

纵观我国数据库应用的发展过程,我们发现,绝大多数用户是从微机数据库系统的应用和开发起步的。这些微机数据库管理系统,如dBase、Foxbase、Foxpro等等,易学易用、对硬件环境要求低,被广大用户誉为"大众数据库系统"。随着数据库应用范围的扩大,广大用户不仅需要这些简单灵活的微机数据库系统来管理、存储本地的数据,而且需要能支持众多用户、共享大规模数据(例如在分布环境下,通过网络)、更加安全可靠(具有安全性、完整性控制和故障恢复)的大型数据库管理系统,如Oracle,Sybase、Informix和DB2等。相应的也迫切需要更新知识,掌握数据库系统的新概念和新技术。

目前,有关数据库系统的著作、教材很多,大部分内容全面,篇

幅较大,比较适合大学本科生、研究生或高级系统开发人员使用。本书则面向广大用户和应用系统开发、管理人员,内容深入浅出、文字通俗易懂;具有理论联系实际、侧重实用等特点,是一本既讲解基础知识又介绍新技术的好书,也是一本难得的普及新型、大型数据库知识的好教材。

萨师煊

1998.2

前　言

信息技术的发展和应用已经影响着社会和经济生活的广泛领域。信息技术中计算机技术是其重要组成之一,而数据库技术又是计算机技术的重要组成之一。当前计算机的应用,不论是开发计算机信息系统,还是计算机应用系统,几乎都离不开数据库技术的支持,它已成为系统的核心和基础。可见当前普及数据库组织和技术有其重要意义和作用。

我国数据库技术的普及和应用起始于80年代初期,并是以微型计算机的数据库应用开始的。它的应用为我国的数据库的普及和计算机应用起到了积极的促进作用,为国民经济和社会发展起到了有益的作用。如我国的税务信息系统就是以微机数据库的应用开始并得到发展的。伴随着应用的发展也培养了一支掌握现代计算机应用技术的人员队伍。

随着社会和经济的发展,各行业的业务量在增加,对计算机处理能力的要求也在增加,因此数据库的应用仅仅停留在微机数据库已经不能满足发展的要求,大家迫切需要了解和掌握管理能力更强、适应面更广、运行更安全的大型数据库技术组织,并从微机数据库的应用过渡到对大型数据库的开发和应用。本书正是为适应这一要求而编写的一本教材,起初是为了内部培训编写的。它从应用要求出发,简明地介绍了关系数据库的基本概念、数据库设计方法,并重点介绍了当前数据库技术发展中的一些新技术:客户/服务器、并行和分布数据库的有关概念,还对数据库管理系统内部实现技术,如进程结构与多线索技术做了简要介绍。它将有利于对现代数据库技术的了解和应用,本书较现行数据库教材更加简明,更侧重于应用。

需要全本请访问 www.ertongbook.com

本书由中国人民大学王珊教授编写。中科院研究生院罗晓沛教授审阅和修改了全书。本书的出版还得到国家税务总局信息中心的支持,廖朝晖、赵璇等同志针对税务信息系统,提供了许多实例,本书责任编辑严晓舟先生为本书的出版付出了辛勤劳动,在此一并向他们表示衷心的感谢。

本书编写时间仓促,定有不足和谬误之处,敬请读者批评和指正。

王珊

1998.2

目　　录

1 绪　　论 …… 1
1.1 概　　述 …… 1
1.1.1 数据、数据库、数据库管理系统、数据库系统 …… 1
1.1.2 数据管理技术的进展 …… 6
1.1.3 数据库系统的优点 …… 7
1.2 数据模型 …… 12
1.2.1 数据模型的基本概念 …… 12
1.2.2 三种主要数据模型 …… 14
小　　结 …… 21
2 关系数据库 …… 22
2.1 关系模型 …… 22
2.1.1 关系操作 …… 22
2.1.2 关系模型的完整性约束条件 …… 25
2.2 关系数据语言 …… 28
2.2.1 概　　述 …… 28
2.2.2 SQL 定义功能 …… 30
2.2.3 SQL 数据操作功能 …… 34
2.2.4 视　　图 …… 44
2.2.5 嵌入式 SQL(Embedded SQL) …… 50
2.3 数据库系统的三级模式与数据独立性 …… 56
小　　结 …… 59
3 数据库保护 …… 60
3.1 安全性 …… 60
3.1.1 安全性控制的一般方法 …… 61

3.1.2 数据库的安全性措施举例 …… 65
3.2 完 整 性 …… 68
3.2.1 关系系统的实体完整性和参照完整性 …… 69
3.2.2 用户定义的完整性约束 …… 74
3.3 并发控制 …… 75
3.3.1 基本概念 …… 75
3.3.2 封　　锁 …… 78
3.3.3 死　　锁 …… 81
3.4 恢　　复 …… 82
3.4.1 数据库的一致状态 …… 82
3.4.2 故障的种类与相应的恢复操作 …… 83
3.4.3 后援副本、日志文件和恢复 …… 85
小　　结 …… 89
4　数据库设计 …… 90
4.1 概　　述 …… 90
4.1.1 数据库设计方法 …… 90
4.1.2 数据库设计步骤 …… 91
4.2 需求分析 …… 94
4.3 概念结构设计 …… 100
4.3.1 概念模型 …… 102
4.3.2 局部视图设计 …… 108
4.3.3 视图的集成 …… 113
4.4 逻辑结构设计 …… 120
4.4.1 E - R 图向关系模型的转换 …… 121
4.4.2 数据模型的调整和完善 …… 122
4.5 数据库的物理设计 …… 124
4.6 数据库的实施和维护 …… 126
4.6.1 数据的载入和应用程序的调试 …… 126
4.6.2 数据库的试运行 …… 127

4.6.3 数据库的运行和维护 …… 128
小　　结 …… 130
5　数据库系统的客户/服务器体系结构 …… 131
5.1 概　　述 …… 131
5.2 客户/服务器一般概念 …… 132
5.3 客户/服务器结构的数据库管理系统 …… 136
5.4 客户/服务器模型与 Intranet 模型 …… 137
小　　结 …… 141
6　DBMS 进程结构及多线索的概念 …… 142
6.1 概　　述 …… 142
6.2 DBMS 进程结构研究 …… 143
6.3 多线索(Multi_Threaded)DBMS 的概念 …… 146
小　　结 …… 150
7　并行数据库系统(Parallel Database Systems) …… 152
7.1 概　　述 …… 152
7.2 并行数据库系统的目标及功能结构 …… 153
7.2.1 并行数据库的目标 …… 153
7.2.2 并行数据库的功能结构 …… 154
7.3 并行数据库的体系结构 …… 155
7.4 并行数据库的并行处理技术 …… 157
7.4.1 并行查询处理概述 …… 157
7.4.2 并行操作算法 …… 160
7.4.3 并行查询优化 …… 160
7.5 并行数据库的物理组织 …… 161
小　　结 …… 163
8　分布式数据库系统 …… 165
8.1 概　　述 …… 165
8.1.1 分布式数据库系统 …… 165
8.1.2 分布式数据库系统的特点 …… 168

8.1.3 分布式数据库系统的目标 …… 171
8.1.4 分布式数据库系统与并行数据库系统的区别… 172
8.2 分布式数据库系统的体系结构 …… 174
8.2.1 分布式数据库系统的模式结构 …… 174
8.2.2 数据分片 …… 176
8.2.3 分布透明性 …… 177
8.2.4 分布式数据库管理系统(D - DBMS) …… 180
8.3 查询处理和优化 …… 183
8.3.1 一个实例 …… 184
8.3.2 查询处理和优化要解决的问题 …… 186
8.3.3 查询优化的目标 …… 188
8.3.4 连接查询的优化 …… 189
8.4 分布事务管理 …… 191
8.4.1 分布事务的恢复 …… 191
8.4.2 并发控制 …… 193
8.5 分布式数据库的发展前景和应用趋势 …… 195
小　　结…… 197
参考文献…… 198
附录　中英文词汇对照及索引…… 200

1 绪　论

数据库是数据管理的最新技术,是计算机科学的重要分支。进入信息时代的今天,信息资源已成为企业和部门的重要资源财富,建立一个满足各级组织信息处理要求的行之有效的信息系统也成为一个企业或组织生存和发展的重要条件。因此做为信息系统核心和基础的数据库技术就得到越来越广泛的应用。从小型单项事务处理系统到大型共享系统,从一般商用事务处理到计算机辅助设计与制造(CAD/CAM)、计算机集成制造系统(CIMS)、办公信息系统(OIS)、地理信息系统(GIS)等,越来越多的新的应用领域采用数据库进行信息资源的存储和管理。当前对于一个国家来说,数据库的建设规模、数据库信息量的大小和使用频度已成为衡量这个国家信息化程度的重要标志。因此,进行信息系统设计与开发的计算机人员和专业人员都必须具备数据库基础知识和数据库应用能力。

1.1

1.1.1 数据、数据库、数据库管理系统、数据库系统

在系统地介绍数据库知识之前,首先介绍一些最基本的概念。

1. 数据(DATA)

数据在大多数人的头脑中第一个反应就是数字。广义地理解,其实数字只是数据的一种形式,而且不是数据库技术最主要的研究对象。数据的种类很多,文字、图形、图象、声音、学生的档案记录、货物的运输情况、纳税人的登记资料和申报资料……,这些都

需要全本请访问 www.ertongbook.com

是数据。为了了解世界,交流信息,人们需要描述这些事物。在日常生活中人们直接用自然语言(如汉语)描述。在计算机中,为了存储和处理这些事物,人们就要抽出对这些事物感兴趣的特征组成一个记录来描述。例如:在纳税人基本信息档案中,如果税务登记机关最感兴趣的是纳税人的姓名、电话、法人代表、行业、经济类型等信息,则可以这样描述:

(北京手表厂,62517888,张丰,钟表制造业,国有企业)

可以对数据做如下定义:描述事物的符号记录称为数据。因此上面的纳税人记录就是一个数据。对于上条纳税人记录知情的人会得到下面的信息:北京手表厂的电话号码为 62517888,法人代表为张丰,行业属于钟表制造业,经济类型为国有企业;而不知情的人要说明以后才明白。可见,数据的形式还不能完全表达其内容,需要经过解释。所以数据和关于数据的解释是不可分的,数据的解释是对数据含义的解释,数据的含义称为数据的语义,数据与其语义是不可分的。

2. 数据库(DATABASE—DB)

数据库,顾名思义,是存放数据的仓库。只不过这个仓库是在硬盘上,而且数据是按一定的格式存放的。

数据库是长期存储在计算机内、有组织的、可共享的数据集合。数据库中的数据按一定的数据模型组织、描述和存储,具有较小的冗余度,较高的数据独立性和易扩展性,并可为各种用户共享。

人们总是千方百计地收集各种各样的数据,然后处理它们。目的是从这些数据中得到有用的信息。在科学技术飞速发展的今天,人们的视野越来越广,数据量急剧增加。过去人们手工处理数据,现在人们借助计算机科学地保存和管理复杂的大量的数据,以便能方便而充分地利用这些宝贵的信息资源。

3. 数据库管理系统(DATABASE MANAGMENT SYSTEM —DBMS)

了解了数据和数据库的概念,就应该研究如何利用计算机科学地组织和存储数据、如何高效地获取和管理数据。完成这个任务的是一个软件系统——数据库管理系统。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。它的主要功能包括以下几个方面:

(1)数据定义功能

DBMS 提供数据定义语言(DATA DEFINITION LANGUAGE 简记为DDL),用户通过它可以方便地定义数据。

(2)数据操纵功能

DBMS 向用户提供数据操纵语言(DATA MANIPULATION LANGUAGE- DML)实现对数据库的基本操作,包括:查询、插入、删除和修改。

(3)数据库的运行管理

这是DBMS 运行时的核心部分,包括并发控制、安全性检查、完整性约束条件的检查和执行、数据库的内部维护(如索引、数据字典的自动维护)等。所有数据库的操作都要在这些控制程序的统一管理下进行,以保证数据的安全性、完整性以及多用户对数据库的并发使用。

(4)数据库的建立和维护功能

它包括数据库初始数据的输入、转换功能,数据库的转储、恢复功能,数据库的重组织功能和性能监视、分析功能等。这些功能通常是由一些实用程序完成的。数据库管理系统是数据库系统的一个重要组成部分。

4. 数据库系统(DATABASE SYSTEM —DBS)

数据库系统通常是指带有数据库的计算机系统。因此,广义地讲,数据库系统不仅包括数据库本身(指实际存储在计算机中的数据)还要包括相应的硬件、软件和各类相关的人员。下面分别介绍这几个部分的内容。

(1) 硬件及数据库

数据库系统包括数据库管理系统和数据库本身。而且由于一般数据库系统数据量都很大,加之DBMS丰富的强有力的功能使得自身的规模就很大,因此整个数据库系统对硬件资源提出了较高的要求,这些要求是:

·有足够大的内存以存放操作系统、DBMS的核心模块、数据缓冲区和应用程序。

·有足够的大的磁盘等直接存取设备存放数据,有足够的磁带(或微机软盘)作数 据备份。

·要求系统有较高的通道能力,以提高数据传送率。

(2) 软件

数据库系统的软件主要包括:

· DBMS,实现数据库的建立、使用和维护功能。

· 支持DBMS运行的操作系统。

· 与数据库接口的高级语言及其编译系统,便于开发应用程序。

· 以DBMS为核心的应用开发工具。应用开发工具是系统为应用开发人员和最终用户提供的高效率、多功能的应用生成器、第四代语言等各种软件工具。如:报表生成系统、表格软件、图形系统等。它们为数据库系统的开发和应用提供了有力的支持。

· 为特定应用开发的数据库应用系统。

(3)人员

开发、管理和使用数据库系统的人员主要是:数据库管理员、系统分析员、应用程序员和最终用户。他们各自的职责分别是:

· 数据库管理员(Data Base Adm intstrator—DBA)

数据库是整个企业或组织的数据资源,因此企业或组织设立了专门的数据资源管理机构管理数据库,DBA 则是这个机构的一组人员,总的来说,他(们)负责全面地管理和控制数据库系统。具体的职责包括:

①决定数据库中的信息内容和结构

数据库中要存放哪些信息,是由 DBA 决定的。因此 DBA 必须参加数据库设计的全过程,并与用户、应用程序员、系统分析员密切合作共同协商,搞好数据库设计。

②决定数据库的存储结构和存取策略

DBA 要综合各用户的应用要求,和数据库设计人员共同决定数据的存储结构和存取策略以求获得较高的存取效率和存储空间利用率。

③定义数据的安全性要求和完整性约束条件

DBA 的重要职责是保证数据库的安全性和完整性。因此 DBA 负责确定各个用户对数据库的存取权限,数据的保密级别和完整性约束条件。

④监控数据库的使用和运行

DBA 还有一个重要职责就是监视数据库系统的运行情况,及时处理运行过程中出现的问题。当系统发生各种故障时,数据库会因此遭到不同程度的破坏,DBA 必须在最短时间内将数据库恢复到某种一致状态,并尽可能不影响或少影响计算机系统其它部分的正常运行。为此,DBA 要定义和实施适当的后援和恢复策略。如周期性的转储数据,维护日志文件等。有关这方面的内容将在下面做进一步讨论。

⑤数据库的改进和重组

DBA 还负责在系统运行期间监视系统的空间利用率、处理效率等性能指标,对运行情况进行记录、统计分析、依靠工作实践并根据实际应用环境,不断改进数据库设计。不少数据库产品都提供了对数据库运行情况进行监视和分析的实用程序,DBA 可以方便地使用这些实用程序完成这项工作。

另外,在数据运行过程中,大量数据不断插入、删除、修改,时间一长,会影响系统的性能。因此,DBA 要定期对数据库进行重组织,以提高系统的性能。

当用户的需求增加和改变时,DBA 还要对数据库进行较大的

改造,包括修改部分设计,即数据库的重新组织。

· 系统分析员

负责应用系统的需求分析和规范说明,他们要和用户及 DBA 相结合,确定系统的硬软件配制并参与数据库的概要设计。

· 应用程序员

应用程序员负责设计和编写应用系统的程序模块。

· 用户

这里用户是指最终用户(End User),他们通过应用系统的用户接口使用数据库。常用的接口方式有菜单驱动、表格操作、图形显示、报表书写等。

在一般不引起混淆的情况下常常把数据库系统简称为数据库。

1.1.2 数据管理技术的进展

对数据的管理是指人们对数据进行收集、组织、存储、加工和传播的一系列活动的总和。但人们借助计算机进行数据管理却还是近二十年的事。研制计算机的初衷是利用它进行复杂的科学计算。随着计算机软件和硬件的发展,其应用远远地超出了这个范围。在计算机硬件、软件发展的基础上,在应用需求的推动下,数据管理技术经历了手工管理、文件系统、数据库系统三个阶段。

1. 手工管理阶段

50 年代以前,计算机主要用于科学计算。从硬件看,外存只有纸带、卡片、磁带,没有直接存取设备;从软件看,没有操作系统及管理数据的软件;从数据看,数据无结构,由用户手工管理,数据量小,且数据间缺乏良好的逻辑组织,数据针对某一应用程序,无独立性。

2. 文件系统阶段

50 年代后期到 60 年代中期,出现了磁鼓、磁盘等直接存取的存储设备。1954 年通用电器公司(GEC)推出的第一台商业数据处

理的电子计算机 UNIVACI,标志着计算机开始进行事务处理并存储信息文件,产生了电子信息系统。人们得益于计算机惊人的处理速度和大容量的存储器,从而解脱了在大堆文件中寻找数据之苦,这种基于计算机的信息处理系统也就从此迅速发展起来。

这种信息处理系统是把数据组织成相互独立的数据文件,利用"按文件名访问,按记录进行存取"的管理技术,并可以经常对文件进行修改、插入和删除的操作,这就是文件系统。文件系统实现了记录内的结构性,但整体无结构。数据面向应用,数据共享性、独立性差,冗余度大,管理和维护的代价大。

3. 数据库系统阶段

60 年代后期,出现了大容量磁盘,存储介质容量大大增加且价格下降。在此背景下,为了解决多用户、多应用共享数据的要求,使数据为尽可能多的应用服务,出现了数据库这样的数据管理技术。数据不再针对某一应用,而是面向全组织,具有整体的结构化,共享性高,冗余度小,且有一定的独立性,系统对数据进行统一的控制。数据库的数据量猛增,用户增加,数据库系统的出现使信息系统的研制从围绕加工数据的程序为中心转向围绕共享的数据库来进行。这样,既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,提高了决策的可靠性。

从文件系统到数据库系统,标志着数据管理技术的飞跃。80 年代后不仅在大型机上,在多数微机上也配置了 DBMS,数据库技术得到广泛的应用和普及。

图 1-1 列出了三个阶段数据管理技术的特点。

1.1.3 数据库系统的优点

与手工管理和文件系统相比,数据库系统的优点是很突出的:

1.查询迅速、准确,而且可以节省大量的纸面文件

以一个大型仓库管理为例,用手工管理,如要查找"上海生产的商品的名称、规格、数量"就可能要翻阅帐本,费时、费力。使用数

需要全本请访问 www.ertongbook.com