

高等学校试用教材

数据库设计

姚卿达 编著

数据库设计

姚卿达 编著

高

TP-792

高等教育出版社

高等学校试用教材

数据库设计

姚卿达 编著

高等教育出版社

内 容 提 要

本书介绍数据库设计的一般概念和方法, 主要包括: (1) 数据库设计的预备知识; (2) 数据库设计的内容与方法综述; (3) 实用数据库的设计方法; (4) IMS 数据库设计; (5) 计算机辅助数据库设计; (6) 数据库设计实例研究; (7) 数据库设计工程的管理; (8) 附录。其中, 数据库设计的预备知识包括了数据库系统的基本概念、语义数据模型以及微型机数据库管理系统等内容。本书作者具有数据库教学经验与数据库设计实践经验, 本书取材比较丰富, 并具有较强的实用性。

本书可以作为高等院校计算机和管理有关专业“数据库设计”课程的教材或参考书, 也可以供从事计算机和信息管理工作的科研人员、工程技术人员以及其它有关人员参考。

责任编辑: 鲍 涌

高等院校用教材

数据库设计

姚卿达 编著

高等教育出版社出版

新华书店北京发行所发行

北京印刷一厂印装

开本 787 × 1092 1/16 印张 21 字数 494 000

1987年4月第1版 1987年4月第1次印刷

印数 00 001—12 330

书号 13010 · 01343 定价 3.25元

前 言

信息在经济建设中,在未来的社会发展中,显得越来越重要。信息准确灵通、传递及时,工作效率和生产效率就会大大提高,经济决策就能建筑在可靠的基础之上。信息已逐步成为决定生产力、竞争力和经济成就的关键资源。开发信息资源,服务于四化建设,这是我国信息工作者的重任。计算机是处理信息的现代化工具,发展现代信息系统,离不开计算机系统的支持,也离不开数据库技术。

数据库(也称为信息库)是信息系统的基础,它按照一定的结构方式将大宗的数据资料保存在存储器上,人们可以方便、及时、准确地从库中检索到所需的信息。数据库技术是信息系统的关键技术之一,它在六十年代中期萌芽,六十年代末和七十年代初开始发展,那时出现了DBTG数据库系统提案,有一些实用的数据库管理系统(如IMS、TOTAL)投入市场,此外,IBM公司研究员Codd又提出了关系数据库理论与方法,这些事实标志着数据库技术进入大发展阶段。经过十多年来的发展与应用,不论在数据库的管理软件DBMS(Data Base Management System)方面,还是在数据库设计与应用方面,都有着一系列成果,同时逐步发展和充实了数据库理论基础。数据库技术与数据通讯技术相结合,产生了各种联机信息系统和分布式数据库系统;数据库技术与人工智能相结合,产生了各种知识库与专家系统;硬件设计者为了适应数据库管理的需要,研制了各种数据库机器;……。所有这些成果,都在实际应用中显示出威力,在国际上,已出现了数以万计的大型数据库及其应用系统。

数据库设计是数据库技术的一个重要方面,目前已形成一门科学。数据库设计指的是对于一个给定的应用环境,构造(设计)最优的数据库模式,然后据此建立数据库及其应用系统,使得数据库既能有效地、安全地和完整地存储大宗数据,又能满足多个用户的信息要求与处理要求。数据库设计有不同的基础,一般来说是在一个通用的DBMS支持下,结合某个应用环境进行,但也有基于文件系统的数据库设计,甚至还有没有文件系统支持的专用数据库设计。

数据库设计属于软件工程学范畴,这门学科涉及到计算机科学的多个方面。对于从事数据库设计的人员来说,要求具备下述几方面知识:

- (1) 计算机科学基础:包括计算机原理、操作系统、数据结构与算法设计;
- (2) 数据处理基础:包括文件系统、数据处理语言(如COBOL)、数据处理系统;
- (3) 数据库系统基础:包括数据组织与存取方法、数据库模式、DBMS原理与结构、DBMS的运用、数据理论;
- (4) 软件工程基础:包括软件生命周期、软件工程原理与方法、系统分析与设计;
- (5) 数据库设计方法:包括设计步骤、方法、工具、规程与理论;
- (6) 应用环境方面的知识:包括应用环境的组织情况、地理分布、现行信息处理状况与流程、信息来源与使用、用户对信息系统的需求、发展预测等。

其中应用环境方面的知识是随着应用业务的变化而变化的,不同的数据库会有不同的应用环境。对于数据库设计人员来说,不是去死记某个特定环境的条件与参数,而是要有善于快速地熟悉和适应环境的能力,以及掌握参数、分析事物并建立系统模型的一套方法。

本书是高等学校计算机有关专业“数据库设计”课程的教材,考虑到不同类型的计算机有关专业的需要,以及在职科技人员和管理人员的需要,我们假定读者已具有上述之(1)、(2)和(3)的基础知识。但为了教学方便,在第一章里扼要地叙述了数据库系统的基础知识,还作了一些补充。对于已学过数据库这门课的学生来说,可以作为复习材料;对于未学过数据库的读者来说,可以作为入门材料或以第一章作为线索参考有关的书籍。

由于数据库设计跟计算机推广应用紧密相关,所以本书在讲解数据库设计基础与方法时,给出一些实例和实习题,力求做到设计方法与实际应用相结合。本书所讨论的方法是基于DBMS的设计方法,至于文件系统支持下数据库功能的实现或是专用数据库的设计问题,同样可以参考基于DBMS的设计方法,本书将不予赘述。

近几年来,微型机技术迅速发展,应用极为广泛,在微型机系统中出现了不少有效的、实用的和使用方便的DBMS,受到广大用户欢迎。例如,被誉为“大众数据库”的dBASE II、UNIX支持下的INFORMIX和INGRES,可以在多种环境下运行的多用户数据库管理系统DATAFLEX等,都是很好的、适于中小规模数据管理的DBMS。为了结合微型机的推广应用,本书有一定篇幅介绍了微型机上适用的DBMS和数据库设计问题。

为了提供计算机辅助数据库设计的实习程序,作者专门设计了一个计算机辅助数据库设计系统,并指导学生进行了具体的实现。这个系统叫CADBDS,它在操作系统MS-DOS支持下运行,适用于IBM-PC、MBE-16000、CORONA、MONROE 2000等机种。CADBDS是一个交互设计系统,读者可以用它作实习,也可以在此基础上修改算法和扩充功能。书中第五章专门讨论了CADBDS以及另一个辅助设计系统CALDBDE。

本书在选题与写作过程中,得到了中国人民大学萨师煊教授和复旦大学潘锦平同志关心与协助;得到了我校黄思曾、容蕙、陈敏、苏小燕、邵军、谢雯等同志的协助;暨南大学黄小蓓同志为本书提供了分布式数据库的有关材料;本书由山东大学董继润副教授和中国人民大学萨师煊教授审阅,提出了许多宝贵意见。作者在此一并表示衷心感谢!

作 者

1986年5月于中山大学

目 录

第一章 数据库设计的预备知识	1	3.2 需求分析与系统提案书	124
1.1 数据库含义与作用	1	3.3 数据分析与抽象	135
1.2 数据库系统	4	3.4 逻辑设计	145
1.3 数据库应用系统	9	3.5 物理设计	158
1.4 实体与实体联系图	11	3.6 实现、操作与维护.....	161
1.5 数据单位	16	3.7 安全性和完整性设计	167
1.6 数据库组织	21	习题.....	172
1.7 数据模型	22	第四章 IMS数据库设计	174
1.8 DBTG 网状数据模型	29	4.1 IMS 数据模型的特点	174
1.9 IMS 层次数据模型	40	4.2 基于 ERM的设计方法.....	175
1.10 关系数据模型.....	54	4.3 扩充 ERM方法.....	177
1.11 语义数据模型与知识库.....	66	4.4 扩充 ERM方法的一个实例.....	178
1.12 微型机数据库管理系统.....	81	实习三 用基于 ERM 和扩充ERM	
习题.....	93	方法进行数据库设计.....	186
实习一 微型机关系数据库的建立.....	95	第五章 计算机辅助数据库设计系统	188
实习二 关系数据库操作.....	95	5.1 ERM 的描述语言ERL.....	188
第二章 数据库设计概论	97	5.2 ERM 的输入、显示、插入与更新...	194
2.1 数据库设计的陈述	97	5.3 ERM 转换为 DBM	202
2.2 数据库设计的目标	99	5.4 CADBDS 操作流程与使用例.....	209
2.3 数据库设计方法综述	101	5.5 CADBDS 的不足之处.....	213
2.4 New Orleans DBD 框架	104	5.6 CALDBDE 简介	214
2.5 基于 3 NF的数据库设计.....	105	实习四 计算机辅助数据库设计.....	222
2.6 析取法	106	第六章 数据库设计实例研究	224
2.7 E-R方法.....	108	6.1 大学环境的数据库设计	224
2.8 多级方法	109	6.1.1 基于 3 NF的设计步骤.....	224
2.9 计算机辅助数据库设计	111	6.1.2 大学数据库的企业模式	225
2.10 数据库设计理论.....	114	6.1.3 大学数据库的逻辑模式	238
2.11 数据库设计的问题与发展情况...	118	6.1.4 大学数据库的物理模式与评价...	246
习题.....	121	6.2 银行数据库设计	246
第三章 数据库设计与开发方法	122	6.2.1 银行数据库的企业模式	247
3.1 数据库应用系统的生命周期	122	6.2.2 银行数据库的逻辑模式	263

6.2.3 银行数据库的物理模式	270	7.4 数据库设计工程的管理	297
6.2.4 银行数据库的物理模式与评价...	275	习题.....	300
实习五 基于 3 NF 的数据库设计	280	附录 1 dBASE-II 命令与信息表	302
第七章 数据库设计工程的管理.....	281	附录 2 DATAFLEX——多用户关系数 据库管理系统.....	309
7.1 工程组织与规划管理	281	附录 3 TOTAL 二级网状数据模型.....	323
7.2 数据字典与数据字典系统	286	参考文献.....	328
7.3 数据库环境	290		

第一章 数据库设计的预备知识

为了学习与讨论数据库设计方法和理论,读者必须具有数据库系统基础知识,包括数据库系统的基本概念、数据库系统结构、数据模型、数据库存储组织、数据描述语言、数据操作语言、数据库的安全性、数据的完整性和一致性、关系数据理论,以及典型的数据库系统(如 DBTG、IMS、SYSTEMR)分析。这些内容可以在萨师煊、王珊编写的《数据库系统概论》或其它数据库系统专门书籍中找到,但考虑到教学上的方便,以及后面各章的需要,在这一章里,将简要地叙述数据库设计中用到的概念,并补充“语义数据模型”、“知识库”和“微型机数据库管理系统”等内容,以供读者参考。

1.1 数据库含义与作用

数据库一词由英语 database 译得,而 database 是由 data 和 base 两个词合并而成的,有时也写成 data-base,简称为 DB。如果直译,那就是“数据的补充基地”,就是说,当需要某种数据的时候,数据库都能随时提供。但是这个源语的解释没有定论,自六十年代末以来,有许多从事数据库研究的专家都给出了数据库的定义,从不同的角度去描绘数据库,在这里我们打算进行专门的讨论,只是引用两种比较普遍的数据库定义,第一种:所谓数据库是被存储起来的运用数据的汇集,其数据被特定的组织(如公司、银行、医院、大学、政府机关等大企业或事业单位)的应用系统所利用。第二种:可以把数据库定义为存储在一起的相关数据的集合,它以最优的方式为一个或多个应用服务;数据的存储独立于使用它的程序;对数据库插入新的数据,检索和修改原有数据均能按一种公用的和可控制的方法进行;数据被结构化,为今后的应用研究提供基础。当某个系统存在结构上完全分开的若干数据库时,则说该系统包含一个数据库集合。

第一种定义主要是从应用的角度去描绘;第二种定义则比较全面。从上述两种定义出发,可将数据库形象描述如下:

(1) 有大量的数据被集中存储在计算机的存储器中,汇集成“数据的补充基地”。这大量的数据汇集是包含多种多类的,而且不同种类数据之间是相互有关联的,而不是孤立的。

(2) 汇集起来的数据为特定组织内各个业务部门(计算机用户)所利用,他们可以使用类似英语或汉语的会话语言从数据库中得到所要的信息或报表,简单而且直观。

例如,在人事数据库的人员表中,可以在终端上打入命令

```
SELECT  姓名  年龄  性别  工资  奖金
FROM    人员表
WHERE   职称='程序员' AND 工资<60
```

而找出所有工资 60 元以下的程序员名单,并列出这些人的年龄、性别、工资和奖金情况。也可以通过专门的业务处理程序来运用数据库中存储的数据,即采用批处理方式或联机方式对存储的

数据进行检索、更新、追加、删除以及加工处理,从而及时地得到所需要的信息,并以报表的形式或其他形式提供出来。

(3) 数据库的数据是综合化和通用化的,它不是为个别用户,而是为多个用户存储数据。一个用户所关心的一般不是数据库的全部数据,只是其中的一部分,而不同的用户共享库中某部分数据,以不同方式重迭使用。

(4) 数据集中存储之后,信息(数据)资源的管理工作已从各个业务部门分离出来,整个组织需要有专门的人员(或小组)——数据库管理员 DBA (Data Base Administrator) 去做这项工作,这也是数据库的一个重要特点。

从上面的描述可以看到,数据库的作用是合理地、有效地存储数据,为有关的人员(应用)准确、快速地提供有用的信息。正因为数据库有这么重要的作用,才使它成为当今信息管理的重要工具。而且随着科学技术的发展与社会的进步,数据库技术亦显得越来越重要。这是因为社会上产生的信息量急剧增加,每一个人都通过各种渠道得到许许多多的信息,例如听报告、参加会议、听收音机、看电视、阅读文件、读报刊、看书、交谈、处理各项事务……,都会接收到各种各样的信息。一个办公室秘书,每天要处理的公函文件可能数以百计;全世界每年发表的科技论文达几百万篇以上;一种报纸所刊登的新闻每天可达十几个甚至几十个版面;一台快速打印机每分钟可以印出几万行数据;在太空技术、地下资源探测、海洋开发等方面所获得的信息更是惊人,而且形式多样、相互联系、错综复杂。有人用“信息海洋”、“信息爆炸”、“信息社会”来形容这种情景。

对于这急剧增长的信息,如何去收集和管理呢?如何从这大量的、杂乱无章的原始信息(数据)中及时而又准确地萃取对于某项活动有用的信息呢?又如何从已得到的信息中进一步推导出新的信息呢?这些就是信息处理这门科学技术所要研究的问题。

所谓信息处理或数据处理,就是对信息(各种形式的数据)进行加工处理,它包括对于数据进行操作的一系列活动,如数据的收集、存储、传播、检索、分类、加工或计算,打印各种报表或输出各种所需要的图形。信息处理的基本目的是从大量事实中萃取对用户有价值的要素,并以某种形式提供出来。

在信息处理的一系列活动中,其中的数据收集、存储、传播、检索、分类等操作是基本环节,这些基本环节统称为信息管理或数据管理,而数据的处理(加工、计算、打印报表等)是千变万化的,不同的业务就有不同的处理。在计算机信息处理系统里,数据管理技术所解决的是上述基本环节,而数据的处理则是靠事务处理程序或应用程序来实现的。数据库技术是数据管理的最先进技术,它所解决的正是信息处理的基本问题——信息资源管理问题。

由于数据库技术在各方面的广泛应用,而且取得了显著的效益,所以有人用“七十年代是数据库时代”、“八十年代是分布式数据库时代”等提法来描绘数据库技术的发展与应用。事实证明,数据库给用户部门的信息处理工作带来许多好处。

数据库给组织所带来的好处有下面几点:

(1) 可以全面、准确和及时地从大量数据中提取所需要的信息。

由于计算机存储容量很大,数据库可以存储几十兆(Mega)甚至几百万兆字节(Byte)的数据,与某方面应用有关的全部数据都可以长期保存在一起,而且能反映数据之间关系,加上计算机具

有高速及自动化程度高等特点,所以,一旦需要,可以很快地提供出有关信息。数据库是调查研究的重要工具,可以为经营决策、编制计划、侦察破案、资料分析等工作提供最合适的参考信息。

(2) 可排除数据的冗余性。

冗余性指的是数据被重复存储。在文件为基础的数据管理系统中,各种业务都有自己专用的文件来保存数据。即使几个不同的应用程序所使用的数据是相同的或部分相同的,也必须建立各自的文件。因此,同一个数据不可避免地要重复地记录在好几个文件之中,都加以保存。这样既浪费了大量的存储空间,也使数据修改变得十分困难。而采用数据库技术,就能够把上述记录的重复和存储空间的浪费缩小到最低限度。

(3) 可避免数据的不一致性。

如果同一数据放在多个的文件中,各有不同的更新时期,那么在某个时期从不同文件中就会产生差异。这就难免造成用户(终端用户)对于输出结果的不信任。数据库采取集中管理与控制,数据是由统一的方式来提供,对所有的用户,数据库以相同的方式来回答询问,因此一般不会产生这种不一致性,或者说保证了数据间的相容。

(4) 数据共用。

不仅现行数据管理系统的文件可以综合化,而且新系统的建设也可以使用已经存在的数据库。

(5) 标准化的推进。

从数据的描述到数据的操作,都容易订出标准。这就使得系统的维护和系统之间的数据交换变得简单。而且对数据共享系统的建设、维护和扩充所需的工本也可以大幅度下降。

(6) 提供安全性保证。

安全性指的是采取各种措施保护数据库,以防止不合法使用数据库中数据,特别是防止恶意的破坏和非法的存取。

由于数据库规定了各种数据的检索、更新、追加和删除的手续,使用数据有一定的规程,还要核对口令、有的系统还采取多层安全措施(如用户标识、存取控制、操作系统安全保护、密码存储等),使安全性得到保证。

(7) 维持数据完整性。

完整性指的是数据的正确性与相容性,数据库能满足对某些数据值的制约,并且经过各种修改后仍保持这种性质。其目的是为了防止数据库中存在不符合语义的数据以及错误信息的输入输出。

即使回避了不一致性,也还可能包含有错误的数据。例如每周工作时间 200 小时,一个人的年龄超过 500 岁,或者不可能有的职工编号等等,由于数据库可以规定检查手续与完整性约束条件,而且又是集中管理,所以容易避免这些差错。数据库可以根据数据的性质和含义,排除有错的数据。

(8) 可调整互相争夺的要求。

对各部门之间矛盾的数据要求进行调整,能够从整个组织来考虑,提供最好的服务。

(9) 可以减少编制和维护应用程序的工作量。

由于数据的组织和管理已从程序中分离出来,数据具有独立性,程序员原来花费在设计数据结构、建立和维护文件上面的时间与精力可以节省下来,并且修改程序可以单独进行。另外,当改变数据库的数据结构时,可以不修改程序。

总之,数据库给组织带来不少好处,同时也产生了一些新的问题需要解决,这些我们将在后面谈及。

1.2 数据库系统

数据库存储的数据,只有在业务处理程序对其进行操作时,才能向人们或机器提供有价值的信息。这就意味着数据库的作用需要在计算机系统以及数据库管理软件的支持下,通过有关的业务处理程序的运行来体现,所以了解和研究数据库不能孤立地进行,而要运用系统的概念和方法,把有关联的成分综合在一起来考虑。把与数据库有关的各个成分有机地结合在一起,就构成了通常所说的数据库系统。现在让我们从数据库的作用及其组成部分的角度,给出数据库系统的轮廓。

数据库系统是用于组织和存取大宗数据的管理系统,它是有关的计算机系统(硬件与基本软件)、数据库及其描述机构、数据库管理系统(DBMS)、用户及其应用程序以及数据库管理员(DBA)的总和,是由这几方面组成的具有高度组织的总体。

图 1-1 给出了数据库系统的结构图。数据库系统是在计算机系统基础上建立起来的,下面对系统的成分做几点说明。

1.2.1 计算机系统

计算机系统指的是用于数据库管理的计算机硬件设备和基本软件。计算机硬件设备的组织方式有各种形式,如集中型计算机系统,分布式计算机系统。此外还有专门的数据库机器 DBM 与数据库计算机 DBC。基本软件指的是操作系统、服务程序、编译程序和通讯软件等等。

数据库系统需要很大的主存,以存放基本软件、数据库管理软件、各种应用程序以及表格缓冲区等,也需要海量磁盘存储器,存放数据信息。数据库系统所需的主存,一般在几百 KB 以上,磁盘容量一般在几百 MB 以上。

1.2.2 数据库及其描述机构

数据库是存储器上存放的相互有关的数据集合,它具有最小冗余,共享性、独立性、统一管理与控制等特点,它通过能在多种业务上公用的数据管理软件,集中地进行存储、编辑、使用、管理与维护。一个数据库系统可以包括几个不同的数据库。

数据库与传统文件的根本区别在于数据间有了联系,具体地说有三点:i) 数据库包含了若干个文件,通过表格或连结路径等方式实现数据之间的联系;ii) 数据库能反映现实世界中信息间联系及含义(即数据语义);iii) 文件只包含数据本身(指数据记录),而数据库则包含了“基本数据+关于数据的数据”,如反映联系的、反映语义的、为了管理控制而附加的各种数据。数据库是

怎样反映现实世界的信息及信息间联系呢？又怎样把各种类型的数据组织在一起呢？这是通过数据模型来实现的。比如建设一所学校，首先要有个平面图以及各个建筑物的结构图和施工图，然后根据图纸搭架、堆砖砌瓦，筑起合乎需要的各幢楼房。建立一个数据库也要有个图，这个图是描绘数据库所包含的数据内容以及数据记录间联系方式(信息结构方式)的数据结构图，然后根据数据结构图装填数据和操作数据。数据库的数据模型(DM——Data Model)就是指能反映记录间联系的数据结构方式。对于一个数据库的描绘，有不同的角度和级别，按照 ANSI/X 3/SPARC 数据库管理系统研究组提出的建议，数据库的描绘分三级进行，即外部模型、概念模型和内部模型。

概念模型(conceptual model)是数据库数据内容和结构方式的完整表示，代表着整个组织中各部门的综合数据观点。

外部模型(external model)是数据库中某部分数据的描述，它是概念模型的逻辑子集，即它可以从概念模型经过转换而得到。代表着某些用户的数据观点。

内部模型(internal model)是数据库的物理描述，即描述数据库数据内容如何在存储介质上安排与存放，它与概念模型之间可以通过概念模型/内部模型映象而转换。

三级模型之间的关系及它们在数据库系统中的地位如图 1-1 所示。

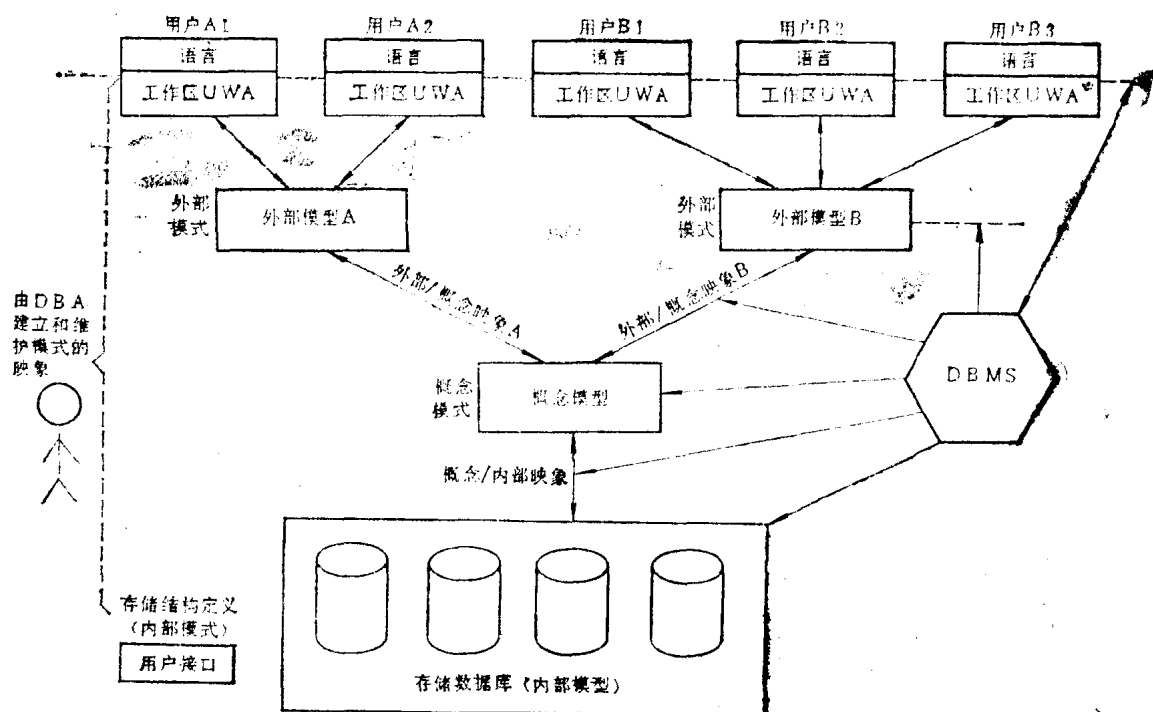


图 1-1 数据库系统结构

概念模型与外部模型属于数据逻辑结构方式，内部模型属于数据物理结构方式。在一个具体的数据库系统中，概念模型通过概念模式(conceptual schema)来定义(描述)，外部模型通过外部模式(external schema)来定义，而内部模型则通过内部模式(internal schema)来定义。可

见,数据模式是用来定义数据模型的工具,通常是通程序语言来编写数据模式,不同的级别有不同的数据描述语言,而不同的数据库系统,其数据描述语言和编写方式也都各不相同。

请注意,在一些介绍数据库系统的资料或书籍里,对数据模型与模式又有不同的说法。例如,有的书把概念模型与内部模型分别叫逻辑模型与物理模型;在 DBTG 系统中,概念模式、外部模式与内部模式分别相当于模式(schema)、子模式(subschema)和存储模式(storage schema)。

在讨论数据库设计方法时,较多场合使用“逻辑模型”、“逻辑模式”、“物理模型”与“物理模式”等叫法,而且有时把“概念模型”分为高级概念模型(只与现实世界有关,不涉及计算机世界)与低级概念模型(与现实世界和计算机世界均有关),按这种观点,本节所说的概念模型应是低级概念模型。另外,在讨论数据库设计方法时,往往不去严格区分“模型”与“模式”,而只关心描绘的级别。

由于概念模式与内部模式之间有概念/内部映象,所以内部模式改变时,有可能通过修改此映象而使概念模式保持不变,从而不必修改程序,这称为程序与数据的物理独立性。类似地,由于外部模式与概念模式之间有外部/概念映象,即使模式改变,也有可能通过修改映象使外部模式保持不变,而不必修改程序,这就是程序与数据的逻辑独立性。从而可见,三级模式结构对于提高程序与数据的独立性是有很大作用的。

1.2.3 数据库管理系统 DBMS

数据库管理系统(DBMS——Data Base Management System)是处理数据库存取和各种管理控制的软件。它可以说是数据库系统的中心枢纽,与各个部分都有密切的联系,应用程序对数据库的操作全都通过 DBMS 进行。

为了说明 DBMS 的作用,让我们来看看怎样通过 DBMS 进行存取。在数据库里,以记录为存取数据的基本单位,一个记录可以是表格的一行,可以是一个人的有关材料的数据描述集合,可以是一种产品的数据描述集合等等。当应用程序需从数据库取得一个记录时,要做下列十一件事(参见图 1-2):

- (1) 应用程序 A 向 DBMS 发出读一个记录的调用命令,并给出调用记录所需要的信息。
- (2) DBMS 找出程序 A 启用的子模式(或程序数据描述),为命令补充信息。
- (3) DBMS 找到模式(或全局逻辑数据描述)并为调用记录进一步补充信息。
- (4) DBMS 探测物理数据库描述,确定要读的记录在存储器上的位置。
- (5) DBMS 向操作系统发出读取所需记录的命令。
- (6) 操作系统与保存数据的物理存储器相互作用,进行读数据的有关操作。
- (7) 所需的数据从存储器传送到系统缓冲区。
- (8) DBMS 从系统缓冲区数据中导出应用程序所需的记录,并进行某些转换,变成应用程序要求的格式。
- (9) DBMS 将数据从系统缓冲区传送到程序 A 的工作区,或叫用户工作区 UWA。
- (10) DBMS 向程序提供它从数据库读取记录的状态,包括读取记录成功的信息或不

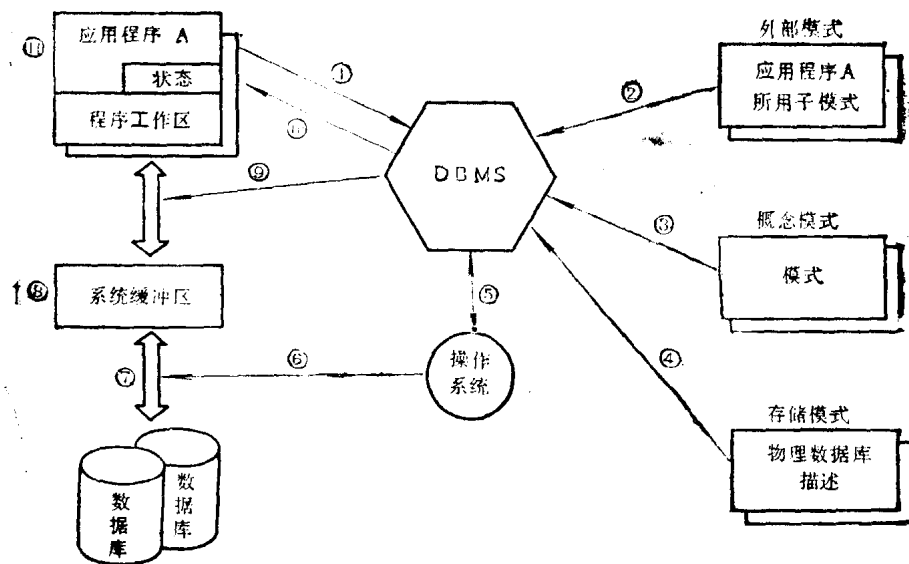


图 1-2 使用 DBMS 为应用程序求取一个记录的操作顺序

成功的错误指示。

(11) 然后,应用程序在它的工作区中,对数据进行操作。

如果要修改一个记录,其所做的事情之顺序也类似。通常,首先是读,然后在工作区里进行修改,最后向 DBMS 发布写回修改数据之命令。写的过程中, DBMS 将在系统缓冲区进行所需的一切转换,即与读数据时相反的转换,然后才向操作系统发出适当的写(write)命令。

DBMS 通常并行地处理多个应用程序对数据库的存取,系统缓冲区为 DBMS 所服务的全体应用程序所共享,应用程序完全是通过 DBMS 同系统缓冲区相互作用的。

有的 DBMS 还处理在读写记录前的寻址与搜索(寻址与搜索也可能放在操作系统里)。有的 DBMS 也处理终端传输线路上的数据传输。这一部分程序叫做数据库数据通讯(DBDC)系统。

DBMS 的处理还包括:① 安全性控制——防止非授权的用户或应用程序去存取或修改数据;② 监督数据库活动或其它形式的系统测试设备;③ 压缩数据处理;④ 根据模式中其它数据项计算一个子模式中某数据项之值,或根据不同物理表示计算模式数据项的程序;⑤ 数据有效性与完整性检查程序;⑥ 错误判断(指 DBMS 执行应用程序存取命令时可能发生的各种错误)。

1.2.4 用户及其应用程序

用户(user)指的是运用数据库进行各种业务处理工作的人或部门,用户的业务处理是通过专门的应用程序来实现的。数据库用户主要有三类:

终端用户——通过设置在各部门办公室或工作室里的计算机终端,利用数据库提供的信息

去处理各种业务的人,例如仓库管理员、售货员、航线预订员、图书管理员、海关的征税人员和检查员等,当他们有了计算机终端之后,就是典型的终端用户。终端用户使用简易直观的终端操作语言或与业务结合紧密的专用操作命令作为人与计算机之间的接口。

应用程序员(application programmer)——负责设计、编制和维护应用程序的人员。随着一个组织的发展,用计算机处理的业务往往会不断增加,或处理方式有所改变,这就需要由应用程序员来进行程序设计和维护工作。应用程序员通常使用某种高级语言(如 COBOL 语言)以及数据库的数据操作语言(DML)来编写程序。

系统分析员(system analyst)——在一个大的组织里,由于数据库系统较庞大和复杂,对于整个组织内各部门的数据要求和处理要求,往往要有专门的人员去分析,并提出规范说明,提出系统发展的提案书和设计书,这些工作单靠程序员是完成不了的。因此,在使用大型计算机系统的单位里,还应有系统分析员专门从事系统分析与设计工作。系统分析员需要有比较全面的计算机系统的软件、硬件基础,需要掌握数据库技术,善于调查研究和分析,对现行业务流程应有较深入的了解,能及时提出计算机化的方案。

当一个单位缺乏应用程序员和系统分析员时,也可以采取委托计算机系统服务部门或其它单位承担的方式或聘请的方式来解决。

1.2.5 数据库管理员 DBA

数据库汇集了整个组织的数据,成为多数用户共享的资源,数据管理工作又从各个部门分离出来,作为整个组织的综合管理功能的一个环节。数据(信息)资源和人力、物力、资金、设备等一样,作为组织的基本资源之一。数据资源的管理工作,不是要不要的问题,而是什么时候加以设置、由什么样的机构去承担的问题。

数据(库)管理员(DBA——Data Base Administrator)正是这样一种从事数据资源管理的机构。它是数据库系统功能的组成部分,它由一组人员组成并担任数据库的管理工作。DBA 是对整个数据库进行控制和管理的人,他(们)的主要职责是:① 利用系统提供的数据描述机构来模拟实际的业务问题,决定数据库信息内容及其关系的恰当描述,组织相应的数据库;② 决定数据库的存储结构和存取方法;③ 调解用户间在数据要求方面的冲突;④ 监督数据库的使用(即数据存取活动)、响应以及保密性;⑤ 数据库再组织;⑥ 测定性能,以便进一步适应用户的数据要求。

DBA 只是起控制与管理的作用,而不是数据的支配者,也不是数据的所有者。数据资源完全归部门或个人所有。DBA 为了进行管理工作,需使用“数据字典/数据地址表”来掌握关于数据的信息。所谓“数据字典/数据地址表”或叫“数据字典/数据目录(DD/D——Data Dictionary/Directory)”,是用来描述数据库中存储数据的属性的,也就是关于数据属性的数据,亦称为元数据(meta data)。具体说来,数据字典/数据目录中所收藏的,是数据库中各种数据的名称、形式、意义、结构、类型、取值的范围、来源、用途、收集负责人、使用权、保密性、有效期间等。数据字典/数据目录存储在计算机存储器上,除了供 DBA 进行控制与管理数据库使用外,也可以供系统分析人员和程序员在设计程序与操作数据库时参照。

1.2.5 数据库系统的其它成分

前面谈了数据库系统的四个基本组成部分和相互关系。数据库系统往往还需要其它辅助成分,例如使用数据库系统需要的各种语言、用户接口(包括人-机交互作用系统)、文字与图形信息处理系统。

语言——为了使用数据库系统并达到数据库管理之目的,需要有各种语言,包括:数据描述语言 DDL,数据操作语言 DML 或数据(子)语言 DSL,物理数据描述语言 PDDL,应用程序员使用的程序设计语言(如 COBOL、FORTRAN、PL/1 或汇编)及嵌入其中的 DSL,终端参数用户使用的询问语言,也许还有系统程序设计语言和允许修改数据模式、子模式、内部模式的维护操作语言。

用户接口——用户所关心的是系统的功能、使用方式、准确性、有效性以及其它有关的数据逻辑组织等。用户与系统的接口是属于外部级的,对终端用户来说,人-机交互(问答)系统、询问语言以及使用系统的策略就是直接与用户联接的接口。例如用在情报检索的数据库系统中,往往是用户提出问题,系统加以审查,给出有关响应,并提出下一步操作或向用户提出询问,让用户给出更多的信息或者要求后再作分析,这样逐步深入,直到检索出所需情报(也许没有)为止。在问题过程中,系统可以帮助用户选择检索策略,纠正用户提问的错误。

文字与图形信息处理——在许多实际环境中,仅靠字符方式输入/输出是不够用的,往往需辅以文字与图形的输入/输出。例如,我们是使用汉字的国家,总是希望计算机能接受汉语并输出汉字,尤其在情报检索中更是如此。要使数据库系统具有文字与图形信息的处理能力,除了相应的设备(象汉字键盘、汉字显示器、图形显示器等等)之外,还需要有相应的处理软件和文字图形库。

1.3 数据库应用系统

数据库系统是用于信息资源管理的先进工具,也是数据处理的核心机构,它在整个计算机应用的领域里,往往起着决定经济效益的作用。

计算机应用几乎已深入到一切领域和部门,可以归纳为科学与工程计算、过程控制、信息处理和其它非数值应用等方面,而信息处理所占比重最大。前面讲过,信息处理就是对信息(各种形式的数据)进行加工处理。信息处理的基本目的即是从大量事实中萃取对用户有价值的要素,并以某种形式提供出来。用于信息处理的计算机系统称之为信息处理系统,早期的信息处理系统主要是采用文件技术,通过文件组织信息,通过文件系统管理信息,六十年代末以来,由于数据库技术和计算机数据通讯技术的发展,实现了计算机化的全面信息处理系统——管理信息系统 MIS(Management Information System)。当前发展的各种信息系统中,往往是以数据库系统作为核心机构,为了更确切地反映这一类信息系统的技术实质,又能与 DBMS 或通常讲的数据库系统有所区别,我们采用“数据库应用系统”来称呼它们。数据库应用系统是指以数据库系统作为核心机构的信息系统。由于数据库应用系统的复杂性及其在计算机应用中占有重要位置,所以在数据库应用系统的开发中,应注意下列问题:

(1) 采用软件工程原理与方法

由于数据库应用系统的大型性和复杂性,系统成分较多,参加系统建设的人员往往有十几个甚至几十个,采用什么方法来指导系统建设和软件开发工作呢?有两种不同选择:一是沿用传统的编制小程序的那种手工方式;二是采用软件工程的方法,也就是用一些计算机科学和其它科学的知识,以工程方法和规范来指导软件研制。实践证明:是否采用软件工程的方法,对于系统的成败、系统质量以及交付使用期的保证具有重要意义。软件工程正是为了组织大型软件生产、克服手工方式的某些缺点而提出来的一种工程科学——软件工程学。这门学科自1968年问世以来,不断地发展,已取得了很多成果,对于指导数据库应用系统建设有重要的作用。

在许多书中,对软件工程学的内容、方法与工具作了讨论,根据软件工程方法,数据库应用系统生命期包括六个阶段:

- 需求分析
- 说明
- 设计(包括数据库设计)
- 编码
- 调试(包括模块调试与综合调试)
- 运行与维护

其中,系统分析阶段的任务是理解和表达用户要求,说明是产生系统说明书或提案书,这些工作主要由系统分析员来做;设计阶段的任务是建立系统结构,对系统的输入、输出、文件、数据库、程序模块及处理算法等作出详细设计,产生各种设计说明书,这些工作主要由设计人员来做。设计阶段的重点是数据库设计;编码阶段的任务是编写程序,产生计算机能接受的代码,这主要由程序员来做;调试阶段的任务是对程序模块进行调试,产生可工作的系统;运行与维护阶段的任务是根据运行评价和发现的问题,从事维护和改进,产生改进的系统。

根据国际上一些大型软件开发的统计,六个阶段的工作量比例以及所花的成本比例大致如图1-3、图1-4所示。工作量比例如图1-3(不包括维护),成本比例如图1-4。

(2) 保证数据库设计质量

在数据库应用系统的设计中,数据库的设计是一个中心环节,数据库的结构方式以及效率问

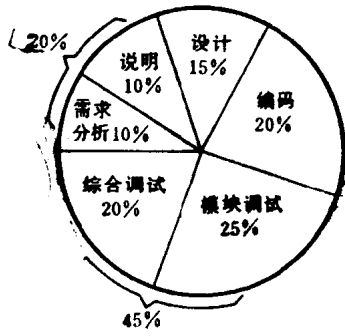


图 1-3 软件开发各阶段工作量比例图

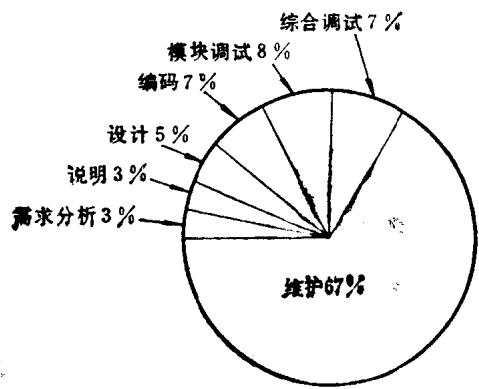


图 1-4 软件开发各阶段成本比例图