

数据库 管理系统 技术评述

〔英〕A·梅尼著

科学出版社



数据库管理系统技术评述

〔英〕A. 梅尼 著

朱致远 译

张作民 校

2P71/22

内 容 简 介

本书通过比较五个商品数据库管理系统(DBMS),对DBMS的主要技术进行了评述。这些技术是:数据模型和数据独立性;数据的安全性和完整性;DBMS的语言接口;DBMS的内部结构以及系统的安装和运行。

本书的读者对象是计算机软件人员,使用DBMS的技术人员,正在挑选DBMS的计算机用户,高等院校有关专业的师生和研究生以及关心计算机应用的管理人员。

Alan Mayne

DATABASE MANAGEMENT SYSTEMS

——A TECHNICAL REVIEW

The National Computing Centre Limited. 1981

数据库管理系统技术评述

〔英〕A. 梅尼 著

朱致远 译

张作民 校

责任编辑 宋义荣

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3月第一版

开本: 787×1092 1/32

1986年3月第一次印刷

印张: 6 7/8

印数: 0001—5,000

字数: 148,000

统一书号: 15031·701

本社书号: 4380·15-6

定价: 1.65元

前 言

本书将对某些重要的数据库管理系统 (DBMS) 进行技术评述。我们假定读者已具备较好的数据处理知识并阅读过有关数据库的一些入门性教材。

本书的目的是描述若干 DBMS 的主要功能, 同时解释它们是怎样进行工作的。我们将通过对下述 DBMS 的各方面功能进行比较来达到上述目的。这些 DBMS 是:

ADABAS;

DMS 1100;

IDMS;

IMS-DL/I;

TOTAL;

由于这些系统本身还在不断地改进, 因此本书将贯注于那些不随时间的推移而发生显著改变的主要特色和工作原理。

本书对以下的单位和个人将会有用:

- (1) 正在着手挑选一个 DBMS 的单位;
- (2) 进行数据库课程教学的单位;
- (3) 拥有上述 DBMS 之一(可作为培训辅导教材)的用户;
- (4) 对数据库和 DBMS 实现技术感兴趣的大学计算机科学系的学生。

目 录

第一章 引言	1
1.1 目的和背景	1
1.2 本书的结构	3
1.3 怎样使用本书	4
1.4 专用词汇	5
第二章 数据的逻辑结构和数据独立性	6
2.1 引言	6
2.2 总的体系结构	10
2.3 数据的组合	13
2.4 逻辑关系	17
2.5 数据独立性和连接	34
第三章 安全性和完整性	43
3.1 引言	43
3.2 后援, 恢复和重新启动	46
3.3 存取控制	58
3.4 数据的验证	63
3.5 多用户存取	65
第四章 语言接口	72
4.1 引言	72
4.2 DBMS 与应用程序的接口	75
4.3 数据定义语言 (DDL)	78
4.4 数据操纵语言 (DML)	84
第五章 数据库管理系统内部结构	105
5.1 引言	105

5.2	数据库管理系统与操作系统的接口	107
5.3	数据结构的物理实现	121
5.4	存取数据	162
5.5	缓冲技术	168
第六章	安装和运行	172
6.1	引言	172
6.2	安装和系统生成	175
6.3	建立数据库	179
6.4	运行支持	182
6.5	数据库重新构造和重新组织	185
6.6	监视和调谐	191
6.7	存贮优化	194
6.8	存取优化	197
	参考文献	202
	缩写词汇表	203
	索引	206

第一章 引言

1.1 目的和背景

首先需要解释一下数据库管理系统(DBMS)的含义。在这之前,先要给出数据库的定义:

“一个数据库是以某种方式组织的一个存贮的数据的集合,它能满足各种用户的数据要求。在这个集合中,尽管有些数据可能是有意识地重复的,但是,一般来说,每个数据项在数据库中只出现一次”(见参考文献[6])。

从以上定义出发,现在我们可以陈述 DBMS 的正式定义:

“一个 DBMS 是一组通用的程序,它们帮助和控制每个用户对数据库的访问和使用,诸如,增加、修改和检索数据。这些程序还包括了提供数据独立性、完整性和安全性的功能。”(见参考文献[6])

我们假定读者已经熟悉数据库的一般概念。因而,本书不包括有关什么是数据库的描述,也不讨论从数据库可以得到的好处以及数据库的缺点。本书的目的在于描述几个数据库管理系统所提供的功能,并且用通俗的语言来解释这些功能是怎样工作的。

为了达到上述目的,我们对五个广泛使用的 DBMS 的逐个功能进行比较。这五个 DBMS 是有代表性的一些数据库管理软件。这种方法将突出各种 DBMS 之间的差别,并且说明实现某种特定功能可能的不同途径。本书描述的五个 DBMS 和研制它们的组织是:

ADABAS	Adabas 软件有限公司
DMS 1100	Sperry Univac 公司
IDMS	Cullinet 公司和 ICL 公司
IMS-DL/I	IBM 公司
TOTAL	Cincom 系统有限公司

DMS 1100 和 IDMS 被称为 ‘CODASYL’ 系统。也就是说,它们的基本设计思想以及它们所提供的功能都是以 CODASYL 认可的推荐文本为基础的。CODASYL 通常负责发展 COBOL 程序设计语言。然而,数据库的功能也适用于其它的程序设计语言。

尽管上述两个 DBMS 都是基于同一个推荐文本,但是他们各自实际支持功能及其实现方式仍然是有差别的。当然,同时描述两个类似的系统势必导致某些重复(在许多情况下,两个系统之间的差别是细微的)。

我们通过研究卖主提供的资料来编写对这些 DBMS 所进行的描述。这些文献包括:基本的产品介绍小册子,程序员参考手册以及包含有关系统内部结构信息的资料。为了澄清各种疑点和得到进一步的信息,作者访问了除 IBM 公司以外的所有卖主。在此基础上,作者先写出每个系统的草稿,然后请相应的卖主进行审定。所有的卖主都提出了建设性意见,纠正了若干错误,并对原文做了修改。如果本书还存在错误,那应该由作者负责。

DBMS 的卖主们总是在不断地改进他们的产品。因而,一本对 DBMS 提供非常详细描述的书将会很快过时。本书并不打算提供定型的描述。我们把主要目标集中在用通俗的语言来说明这些系统是怎样工作的。某些 DBMS 功能(例如,某些并非广泛使用的恢复功能)并未收集进来。出于这一原

因,作者建议那些利用本书帮助他们挑选 DBMS 的单位直接与卖主联系,以便保证他们的决定以最新的信息为基础。本书未提到由某 DBMS 所支持的某种已知的特色,并不一定意味着该特色不存在或者没有其他方式能满足要求。作为一本集中于主要特色(改变不甚频繁的)的书,应当在相当长的一段时期内具有参考价值。

1.2 本书的结构

本书分成以下几章:

第一章 引言

第二章 数据的逻辑结构与数据的独立性

第三章 安全性和完整性

第四章 语言接口

第五章 数据库管理系统内部结构

第六章 安装和运行

第二、三、四章涉及到作者认为是一个 DBMS 的最重要的三个方面。本书的细目结构(如目录所示),可以用作比较任何一个 DBMS 的参考框架。另外,读者还可以从参考文献[2]和[3]中找到有关选择准则的详细清单和讨论。

逻辑结构化的功能很重要,因为它将影响到设计和实现应用系统的方式。它们决定了一个组织的数据之间的自然关系可以怎样地被模型化,以便在计算机上实现。第二章还包括有关数据独立性的讨论。

第三章中描述的安全性和完整性功能显然是至关重要的。当以任何方式破坏数据库的事件发生时,重要的是能够修复它。由于数据库中的数据为若干个用户所共享,规定一

些存取控制限制,以便阻止非授权的用户存取数据库某些部分是必要的。如果数据库允许更新,则数据的共享也可能会引起某些问题。当某些事件发生时,DBMS 必须能够保护数据的完整性。例如,在两个不同的程序试图同时更新同样的数据时,就有这样的需要。

DBMS 的用户与系统本身之间的接口也是很重要的。此处的用户是指直接与 DBMS 本身接触的任何人。这包括应用程序员和负责对系统描述所要求的数据库的人。DBMS 的语言接口决定了系统操纵数据的能力。语言是否易于使用将在很大程度上影响培训的开销,同时也将在程序设计期间的出错率和调试时间上反映出来。

第五章评述了 DBMS 更为技术的方面。它描述了各种 DBMS 与计算机操作系统的接口。它还包括了怎样将数据存放在物理存贮器上以及怎样实现逻辑数据结构的描述。本章的信息对于数据库系统(软件)程序员是特别感兴趣的。了解 DBMS 内部结构对于那些设计(物理)数据库以产生一个有效的系统的人员来说,也是十分有用的。

最后一章考虑 DBMS 的安装以及运行时的维护。它描述了当数据库增长和改变时,对于重新构造和重新组织数据库所提供的功能。一个运行着的系统的性能也会引起一些非常令人头痛的问题。因此本章还讨论了有关系统监视和调谐的功能。优化存贮空型和存取时间等特殊问题也很重要,所以我们把帮助解决这方面问题的 DBMS 功能也编进了本章。

1.3 怎样使用本书

在编写本书时,已考虑到,读者可以按下述几种方式使用

它:

按物理顺序 读者可以从头至尾地通读全书,以这种方式使用本书时,读者将会发现某些重复信息,特别在第二章以及其它章节 DMS 1100 和 IDMS 的描述中。第四章对某些读者可能是难读的,然而可以放心地翻阅过去,因为以后各章并不依赖于它。尽管第六章有一小部分内容与第五章有关,但是对于技术细节不感兴趣的读者仍然可以跳过第五章。

按管理人员关心的顺序 一位不关心技术讨论而忙忙碌碌的经理可以跳过有关五个 DBMS 的具体描述。

按 DBMS 的顺序 主要对某一 DBMS 感兴趣的读者可以跳过其他 DBMS 的描述。在任何一节里,对每一 DBMS 的描述都是独立于其它 DBMS 的;为了能够按照这种方式使用本书,在必要的地方我们有意识地重复一些信息。

查目阅读 利用本书前面的目录和后面的索引,读者可以象使用一般参考书那样来使用本书。

直接查阅 如果知道页号,则可以直接查找有关信息。

1.4 专用词汇

在此类书中,专用名词带来一些特殊问题。每一个 DBMS 都有它自己的一套行话。作者曾经考虑过在本书中使用一套标准的术语,但是又考虑到读者将它们与卖主的文献相对应时会产生更多的困难(特别是将本书用作培训教材的一部分时)。因此,我们还是使用各卖主的专用词汇。希望读者能顺利地从一个专用词汇转到另一种而不致混淆。对于这一点唯一的例外是术语“字段”(field)用于本书的始终。因为它比“数据项”(data-item)或“数据元”(data-element)更易书写、打字和阅读。

第二章 数据的逻辑结构 和数据独立性

本章的目的是:

解释逻辑结构和数据独立性功能的重要性;同时,对五个 DBMS 的每一个;

描述系统的总的体系结构;

解释数据怎样被分成字段,组合项,记录和文件;

描述怎样将不同的记录型联系在一起,以便将它们之间的自然关系模型化;

考虑系统支持哪一级的数据独立性。

2.1 引言

第一章中有关数据库的定义陈述了:数据库是按某种方式组织的存贮数据的集合,而如此组织的数据库可以满足各种用户的数据要求。一个数据库所存贮的各数据项并非彼此孤立地存在,而是以某种方式彼此相关。例如,表示人名、身高、体重和年龄的这些数据项就是彼此相关的,并且可以方便地“组合”在一起。

数据间的关系还可能是更为复杂的。如果一个人拥有几辆小汽车,那么这个人与其拥有的小汽车的细目间就有一对多的关系。如果几个人共同拥有一个汽车队,那么在人的细目与车的细目间就有多对多的关系。每个人是若干辆汽车的

部分拥有者，而每辆汽车又为若干人所拥用。

以独立于任何计算机的实现方式标识基本数据组及其相互关系叫做建立数据模型。建立数据模型是设计数据库的第一步。如果在实现数据库之前没有做好这一步，那么将会失去采用数据库方法所带来的某些好处。建立数据模型和数据库设计已超出本书的范围，有兴趣的读者可以参看参考文献[6]和[7]。

总之，一个DBMS必须提供能够表示数据项之间自然关系的功能。换句话说，DBMS应该具备以某种方式模型化数据逻辑结构的能力。利用所提供的逻辑结构化功能，能否容易地实现数据模型，对于设计数据库和开发应用系统所要做的工作有明显的影 响。如果一个DBMS逻辑结构化的功能受到限制，那么就有可能出现过多的重复数据，而且为了避开它的短处，有可能需要对程序采取一些特殊的策略。如果一个单位正在选择一个DBMS，那么应该保证所挑选的DBMS的逻辑结构化的功能能够以一种清晰和自然的方式模型化他们的数据。

至此，有必要解释一下术语“记录”和“记录型”的含义。“记录型”是用来表示刻画所有类似客体的集合的专用术语。在说到单个记录时，使用“记录”这个术语，描述一个具体的客体。

数据项之间的关系是通过将数据项组合成组合项和记录来表示的。这些关系称为字段间的关系，有时也称为记录内部的关系。更为复杂的一对多关系是通过允许一组合项在一记录中重复多次，或者以某种方式建立记录型之间的联系来表示的。不同的记录之间的关系称为记录间的关系。本章后面将描述这五个DBMS所支持的各种类型的关系。

在第一章有关 DBMS 的定义指出：它应该“包含提供数据独立性的功能”。数据独立性允许改变数据的物理或逻辑结构而不引起改变应用程序所看到的数据。这个功能还为使用数据库软件带来了某些好处。所以，由一个 DBMS 所提供的数据库独立性的级别是非常重要的。在选择一个系统时，对此应该做认真的调查。

为了说明由数据独立性带来的某些好处，有必要考查一下在使用传统的文件时，程序是怎样依赖于数据的。在那种情况下，使用文件的每个程序都包含存放在该文件中的一种（也可能是几种）记录型的格式定义。程序逻辑还要考虑文件的结构。例如，它的处理方式。如果文件包括多个记录型，诸如批标题，批细目和批总计等记录型，则情况更是如此。

如果以任何方式改变文件的格式或结构，则必须修改所有使用该文件的程序。例如，如果在文件的记录中加入一个新的字段以支持一个新的程序，则将要求每一个使用该文件的程序都做相应的修改，尽管它们并不实际使用这些新字段。在这个简单的例子中，只需要重新编译现有的程序。然而，可能有上百个程序。这个例子说明一个看上去很简单的改变可能带来耗时费工的工作。

经常要求的另一种修改形式是更改一个字段的长度。显然，使用该字段的程序将直接受到影响，从而需要重新核对程序的逻辑。乍看起来，可能会认为，没有实际使用该字段的那些程序只须重新编译就行了。其实不然，因为如果用户程序将包含改变长度的那个字段的组合项或记录传送到工作存贮区，那么工作存贮区的记录描述也需要作相应的改变。例如，如果应用程序将一个记录传送到一个工作文件，那么该工作文件的记录描述也必须作相应的修改。因而必须检查所有使

用到该文件的程序逻辑，包括并不实际使用有关字段的那些程序逻辑。

上面列举的例子表明了与传统文件有关的某些维护问题。通过使应用程序独立于文件的组织方式可以克服这些问题。有三类不同的独立性：

设备独立性 程序独立于存贮数据的物理设备。所有的 DBMS 都提供这一级的独立性。事实上，对于传统文件，好的操作系统的数据管理功能也提供了这一级的独立性。

物理数据独立性 程序独立于数据的物理组织方式。这一点意味着可以改变记录在数据库中的物理位置，还可以改变数据的物理格式。例如，从二进制到浮点格式。至于记录结构或记录间关系的改变则被看成是逻辑改变。

逻辑数据独立性 这一点允许改变数据的逻辑结构而不影响到应用程序的数据视图。这就意味着可以改变一个记录型的结构。例如，在记录型中加入一个新的字段，对现有字段重新排序或改变字段的长度以及删去未使用的字段等。这也意味着能够增加，改变或删除关系。

通过硬件和软件资源来提供完全的数据独立性其代价是非常高的。因此，还没听说有 DBMS 提供完全的数据独立性。DBMS 的设计者必须在完全独立性的想法和提供它所带来的实现上的困难之间进行权衡。

为了讨论五个 DBMS 所支持的逻辑结构和数据独立性功能，有必要介绍某些基本概念和专用名词。因此以下两节将描述各系统总的体系结构以及它们分别支持怎样的数据组合。

为了提供数据独立性，DBMS 必须在数据的物理表示和程序的逻辑视图之间映象数据。将程序的逻辑视图与物理存贮器中的数据联系在一起的处理称作为连接。

2.2 总的体系结构

本节简单地介绍五个 DBMS 的总的体系结构。目的是引入与每个系统有关的某些基本的专用名词。

2.2.1 ADABAS

ADABAS 是以部分倒排文件结构为基础的。ADABAS 的数据字典存在着数据的概念视图,物理视图和逻辑视图。概念视图由与组织功能有关的标准字段和文件的定义组成。物理视图是存放在 ADABAS 系统文件(称为 ASSOCIATOR)中的 ADABAS 数据库文件和字段的定义。逻辑视图(又称为用户视图)是在规定格式和顺序应用中个别应用的数据视图,它与数据如何存放在 ADABAS 数据库上无关。

2.2.2 DMS 1100

DMS 1100 是以 CODASYL 建议为基础的。它有两级数据定义。最高级是模式(SCHEMA),它定义数据库的所有数据、关系和设备介质控制参数。模式是用数据定义语言(DDL)来描述的。模式的信息存放在一个称为“元数据库”的数据库中,系统还将它转换为一个模块以便在运行时使用。

第二级定义是子模式(SUBSCHEMA),它定义了数据库的全局视图的一个子集。子模式用来描述可能被一个应用系统存取(包括如何使用它们的信息的部分数据库)的一个局部视图。子模式是用子模式数据定义语言(SDDL)定义的。通常,一个模式有若干个与之联系的子模式。每个子模式描述数据库的一个不同的应用视图。子模式的信息也存放在元数据库

中,同时也放在子模式模块中以备运行时使用。

2.2.3 IDMS

IDMS 也是以 CODASYL 建议为基础的。它有三级数据定义。最高级是模式,它规定了数据库的所有数据描述,数据关系和输入/输出参数。这是用模式数据描述语言来规定的。模式信息存放在一个称为 DIRECTORY 的专用的 IDMS 文件中。

第二级是设备介质控制块(DMCB)。它提供了在逻辑和物理数据描述之间的映象信息。DMCB 是用设备介质控制语言(DMCL)或服务描述语言(SDL)来规定的。可以有一个或多个 DMCB 来描述整个数据库或数据库的一个子集。DMCB 的信息存放在 DIRECTORY 文件和控制块模块中,当执行数据库应用时将后者装入并使用。

最后一级是子模式,它描述了数据库的一个子集。该子集确定了一个特定应用的数据库的局部视图。它规定了数据库中的哪些成份可能被存取以及怎样被存取。

子模式用子模式数据描述语言来描述。通常有若干个子模式来描述不同的数据视图和规定各种存取授权。子模式信息也存放在 DIRECTORY 中,同时还存放在子模式模块中,以备运行时使用。

2.2.4 IMS-DL/I

一个 DL/I 数据库上的数据用一个或多个数据库定义(DBD)模块来描述。每个 DBD 描述一个物理数据库。每个物理数据库定义了一个物理记录型。

一个逻辑记录用一个或多个节段(SEGMENT)来定义。