

小型关系数据库管理系统 TDMS

瞿中伟 展大鹏 赵美锡

提 要 本文介绍在ALPHA微型计算机上实现的一个小型关系数据库管理系统TDMS(表格数据管理系统)。该系统目前已投入我校的学生档案和学籍管理的使用。

一、引 言

在任何管理部门中,大量杂乱文件和数据是没有什么用的,只有经过整理的数据,才能成为有用的信息。随着计算机应用范围的拓广,大量的数据单靠文件系统来管理已经远远不够了,因而在本世纪六十年代后期在文件系统的基础上产生了数据库这样一种处理信息的新工具。近年来,数据库管理系统作为信息系统的一个重要组成部分,愈来愈显示出它的重要性。

有关数据库的理论和方法,国内外都有很多文章很好地探讨过。然而,要真正建立一个各方面性能均佳的数据库管理系统恐怕是太难了。我们力图在微型机系统上建立一个实用的小型数据库管理系统,以解决诸如人事档案管理、财务管理、仓库管理一类的“管理信息系统”的问题。一般说来,微型机速度慢,内存外存容量小。但是注意到使用“管理信息系统”的管理者和职员总是把方便、有效的要求放在实时响应速度之上的,数据库管理员总是希望数据库容易建立、修改、维护和占用较少的存贮单元。例如,一个管理学生档案的干部,希望知道在600个新入校的学生中有哪些学生是来自北京而且年龄在18~23岁之间;他们高考的平均分数是多少。他关心的是计算机能否正确地找出来,操作是否简便。而计算机在2秒还是6秒作出回答,这是无关紧要的。若是制作一份600人按来源省市;相同省市的按考分高低;同考分的按年龄大小排序的花名册,3分钟乃至更长的时间也是可以容忍的(手工作需一个熟练的管理人员干半个月以上)。因而在ALPHA微型计算机上设计TDMS系统的目标是:1.有较好的数据独立性。数据与使用它们的应用程序相互独立,使得对于数据逻辑组织和物理存贮的修改不必修改应用程序,反之亦然。2.数据库容易修改和重构。因为数据库的重要特征之一乃是它们需要不断地修改和增生。3.设计一个方便的数据查询语言,允许用户以一种设计者无法准确预料的方法利用数据。4.有效使用存贮空间,减少数据冗余,较好地实现数据共享。5.有一定安全、保密措施,保证在用户越权使用和硬件故障情况下不破坏数据库。另外,设计时要尽可能利用微型机上原有的软件资源,使用简单的数据结构和程

序结构,以使得设计简单,维护方便。我们应该记住:“伟大的工程是简单的工程。”

二、数据库模型

在大多数管理部门中,人们表示数据的最自然的方法常常是使用如图 1 所示的二维表格。

学 号	姓 名	性 别	年 龄	籍 贯 ...
7912107	张 强	男	17	天津
7912108	王小兵	男	21	北京
7912109	王 红	女	21	北京
7912110	李 东	男	18	天津
7912111	张小山	男	20	上海
7912112	林 义	女	19	河北
7912113	邓 帆	女	18	上海
7912114	郝建民	男	23	江苏

图 1 学生档案中的部分数据

这样的表格具有下列性质:

- 1) 表格中的每个登记项代表一个数据项;
- 2) 每一列里的所有项目有同一类型(按列均匀);
- 3) 各列被指定一个相异的名字;
- 4) 各行相异;
- 5) 行和列的次序不影响其信息内容及使用它的任何功能的语义。

这样的表格在数学上称为关系。使用关系构造成的数据库就称为关系数据库。由于它具有简单灵活、独立性好、理论严格等优点,所以是颇有前途的一种数据库管理系统。

TDMS 系统就是为处理这样一类关系表格而设计的关系结构的数据库管理系统。TDMS 在操作系统 AMOS 支持下以 TQL 语言为数据操纵语言,采用人机对话方式工作,系统结构如图 2 所示。

数据库的组织采用外模型—概念模型—内模型三级映射,如图 3 所示。

概念模型是关系表格的全局逻辑描述。外模型是关系表格的用户视图的描述。一个关系表格可以有多个视图,它们是从概念模型中定义的关系

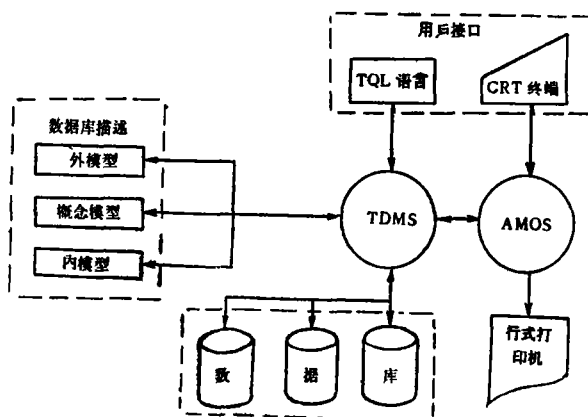


图 2 系统结构示意图

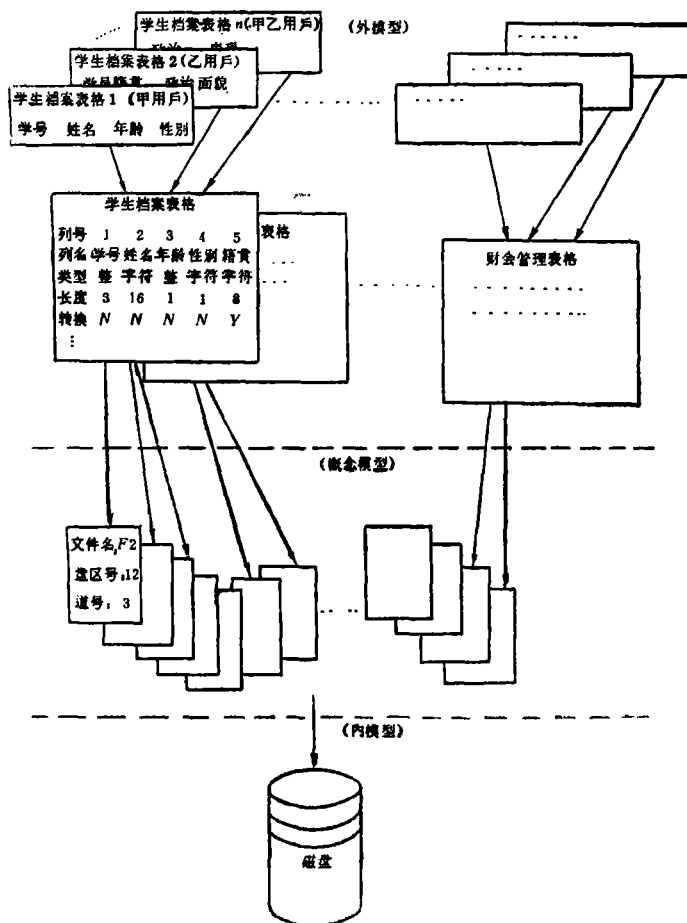


图3 外模型、概念模型、内模型三级映射示意图

表格里抽出的若干列，按照使用的方便组成子表。内模型是关系表格物理存储的描述。TDMS系统管理的表格数据以及主要由外模型、概念模型和内模型组成的数据库字典均驻留在磁盘存储器上，由TDMS系统对其构造并解释。用户通过表格名调用某个关系表格的某一视图，系统通过外模型—概念模型的映射得到所涉及的数据的逻辑定义细节，又通过概念模型—内模型的映射得到所需访问数据的物理存储细节。从而数据对用户是“透明”的。他们直接看到并操作的是二维表格中的数据项，而不必知道它们的逻辑组织和物理存储定位方法。

三、数据操纵语言

一个“管理信息系统”是否具有生命力，恐怕更多地取决于它提供的数据操纵手段是否简便的、强有力的、受到大多数非程序员用户的欢迎。我们为TDMS系统配制了一种表格数据查询语言TQL(Table-Query-Language)作为数据操纵语言。TQL语言在格式上和语法上都类似于M.Zloof研制的Query-by-Example语言。用户通过填写终端屏

幕上的一个二维表格来描述对系统所管理的表格数据的检索和处理,象翻阅各种图表一样的清晰、简单。对用户的数学和程序设计方面的知识要求不高,它适合于一般管理人员的工作习惯和心理。

二维表格框架具有如图4的形式。

当系统显示 BEGIN 引导符后,用户即可开始填表。先在表名项填入需要进行操作的表格名,系统便将对应的视图中的各个列名在表格中列名项的位置上显示,用户再在相应的列和行上填上一些条件和命令,最后在行属性项上用 END 结束提问,系统

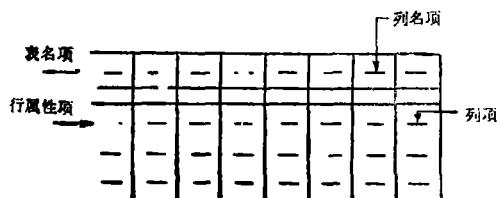


图4

就将该次查询的结果输出。例如,对于图1中的表格,提问“找出所有年龄在18~23岁之间的男学生并统计总数”可以如表1那样填写。

其中,“P.”表示“印出”(print)。

表1

‘男’是常量,表示需要找出的学生其性别是‘男’。“ ≥ 18 ”和“ < 23 ”也是常量,它们限定了要找的学生的年龄范围。而“ x' ”是一变量,以右上角带一撇“'”为标识,变量代表要找出的学生

表1	学号	姓名	性别	年龄	籍贯
	$P.x'$	$P.$	男	≥ 18	
	CNT.			< 23	
END					

在这一列(学号)的所有可能值,作为绘制统计曲线、设计报表格式以及多个关系表格的链接变量。“CNT.”是一库函数,表示对符合条件的学生人数计数。这一查询写成谓词演算式为: $\{(x,y): \exists Z \exists W(x,y,男,W) \in \text{图1中的表} \wedge W \geq 18 \wedge W < 23\}$ 。查询的结果是:

7912108	王小兵
7912110	李东
7912111	张小山

如果是插入、删除或更新操作,则需在行属性项填上I., D.或U., 例如要求将北京来的学生的年龄增加一岁,可如表2那样填写。

第一行填写的是条件,行属性项空

表2

表示检索(同前例),第二行的行属性项是U.表示更新,在年龄项下, $x' + 1$ 表示更新的结果是原有的值加1,操作的结果是所有其籍贯是“北京”的学生的年龄的值均增大了1岁。

表1	学号	姓名	性别	年龄	籍贯	...
					北京	
U.				$x' + 1$		
END						

实际上, TQL 查询语言就是用户填写一张表格为计算机作一个示范。各行的次序是无关紧要的。所有的条件均是“与”的关系。可见 TQL 语言具有二维语法和高度的非过程性。即用户只需要告诉系统要找怎样的数据,而不必告诉它怎样去找。语法简单,可以应付非预期查询,适合非程序员用户使用。

TQL 允许的条件运算符和函数有:

- 数值比较: =、 $\neq$ 、>、<、 $\geq$ 、 $\leq$
- 逻辑非: $\neg$
- 前缀比较: —
- 库函数: P. 打印;
 C N T. 求总数;
 % 求百分比;
 M A X. 求最大值;
 M I N. 求最小值;
 S T A. 求五项统计值 (包括和、平均、平方和、方差与标准偏差);
 A O. 按升序排列;
 D O. 按降序排列;
 U N. 删去重复。

库函数的用法都是指对列的,如同前例中的 P. 和 C N T. 一样。

TDMS 系统为用户提供了报表设计的功能。用户可以定义输出报表的表头、每页最大行数、行间隔、列间隔、花边等,使得查询结果按照用户的意图打印成美观的、符合要求的报表格式。TDMS 系统还有输出“虚拟”数据的能力,这些“虚拟”数据是库中真实数据的函数。例如,图 1 的表中只有学生的年龄而无出生年份,设查询时指定年龄项的变量为 AGE',则可以在设计报表格式时填写 1983—AGE' 就会得到出生年份。当某学生的年龄值为 21 时,输出结果便是 1962。又如,设学生数、理、化考试成绩项的变量分别指定为 X'、Y' 和 Z',则可以输出 $(X'+Y'+Z')/3$,就会得到数、理、化平均成绩。这一特性使得凡是由数据库中现有数据计算得到的数据都不必真实地存贮了,从而节省了存贮空间,减少了数据冗余度。“虚拟”数据的计算表达式可以使用 +、-、 $\times$ 、 $\div$ 、指数、括号以及开方($\sqrt{\quad}$)、阶乘($n!$)、对数($\lg, \ln$)、三角函数等十多种函数调用。

TDMS 系统还提供了一种绘制统计图表的功能。这是针对很多管理人员喜欢直观的曲线表示法而设计的。例如绘制学生的数学课成绩的分布曲线,看一看是否符合正态分布,统计成绩在 100~90, 89~80, 79~70, 69~60 和 60 以下的学生各有多少人?各占总数的百分之几?用户只需要用变量指明要绘制曲线的那个项目,定出上限、下限与步长,系统便给出各数据值段的记录个数、占总数的百分比,并在适当的坐标位置上用离散点构成数据分布曲线。

TDMS 系统使用口令和权限字加强保密性和安全性。没有口令,连表格的视图都看不到。权限分为可删、可插、可改。任何越权使用均会被指出。

在使用 TQ L 语言进行查询时,如果用户不知下一步如何做或者忘记了某些命令或库函数,只须在行属性项填入 H E L P 便可得到提示。

总之, TQ L 语言是一种语法简单而功能较强的数据操纵语言。实践证明,从未学过程序设计的用户也能在几天之内掌握它。

四、数据存贮组织

T D M S 系统采用按列分离存贮的方法, 利用操作系统中原有的随机文件管理进行物理存贮。大量的实际数据是存贮在磁盘上的。如果独立地设计一个包括盘面分配、分页机制的物理存贮管理就要花很大气力。我们在设计中经常考虑的是: 怎样充分利用原有的软件资源? 这点做得好的话, 可以为设计带来很大便利。按列分离存贮方法的好处在于: 1. 和关系表格按列均匀的性质是一致的, 可以想象将这些文件平面排列即可得到一个二维表格。2. 索引和数据项本身一致, 无冗余, 任何一列均可作为关键字检索。3. 二维表格可以任意添加新列而不影响其它列。4. 查询中不涉及的列可以不调入内存, 从而大大减少换页次数。5. 数据项可以比较长。由于逻辑记录的大小实际要受到磁盘物理记录块大小的限制, 使用通常的存放方法, 数据记录(表格中的一行)长度有限, 而使用按列分离存贮, 每一数据项可以大到占一个物理记录块, 表格列数不受限制, 从而表格中的一行数据记录的长度也就没有限制。总之, 采用按列分离存贮使得数据库设计中许多问题, 特别是数据独立性问题能够比较容易地得到较好的解决。

数据记录的存放是不排序的。库函数 A O. 和 D O. 可以使查询的结果在输出时依任意指定的若干列的值排序。数据记录之间不是稠密存放而是适当地留有小部分自由空间。删去一行的操作使得自由空间加大而插入一行的操作将有关数据项插在自由空间中(在没有自由空间时插在最后), 这样避免了修改操作时移动大量数据而影响查询速度。系统在一定时间后对自由空间进行一次整理。用户可以决定自由空间的大小。例如在学生档案数据库中, 根据学生的变动情况, 自由空间保持在5%左右。

请注意这样一类数据项, 例如学生的“籍贯”, 三十个省、市、自治区名, 每个用字符表示约占8字节。若每人存贮一次则存贮量很大。我们对这一类数据项采用换码技术。先定义一张对照表, 在输入时由系统将字符转换为只占1个字节的指针, 内部存贮及比较均用此

指针。输出时由系统将指针转换为字符。这样, 可以节省大量存贮单元。转换是需要时间的, 但由于内部比较运算时使用较短的指针内容, 时间上也得到了补偿。

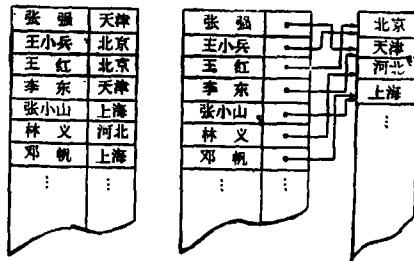


图 5

五、程序实施

T D M S 系统是用扩展 BASIC 和宏汇编语言写的。利用 BASIC 的会话功能, 对非过程性的数据查询语言进行预编译, 包括语法分析、报错纠错、优化和生成半目标

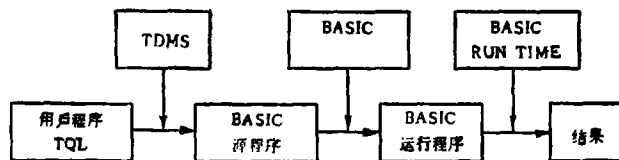


图 6

代码——扩展BASIC源程序，再利用已有的编译程序形成目标代码，交给运行时间包解释执行。这种预编译—编译—运行的方法是需要一定时间的。但由于有效地利用了原有系统的资源，使得实现工作简单，程序清晰。

ALPHA 机的用户内存是32K。我们采用内存复盖技术，将程序分成比较独立的几个模块，依次调入内存执行。

TDMS系统在分时操作系统的支持下可由多个用户同时使用对其管理的数据进行查询。但在使用删除、插入、更新命令时，系统将所涉及的文件封锁以禁止其他用户访问，操作完毕后解除封锁。

为保证数据的安全，所有数据都有1~2份备份。备份存贮在活动软盘上，用户在必要时可命令系统作一次转贮以保持原本与备份的一致。系统还将每两次转贮之间用户操作的细节记录下来，以保证数据在遭受破坏后能够正确地恢复，如图7所示。

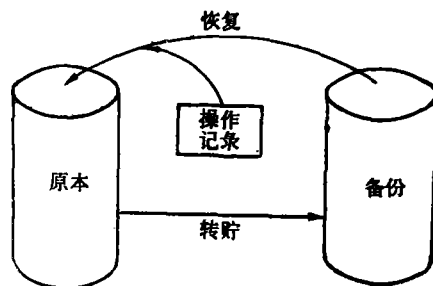


图 7

六、应 用

目前，TDMS系统已应用于我校的学生档案和学籍管理。每个学生的登记项有学号、姓名、性别、出生年月、籍贯、家庭住址、民族、政治面貌等二十几项和4个学年约50门课程的考试（查）成绩，根据用户需要建立了12个用户视图。对于目前已装入的三百多个学生的查询，一般提问的响应速度约为2~6秒（单用户情况下），但要求排序输出时响应较慢，这是由于内部存贮没排序的原因。

TDMS系统中可以定义多个（暂定255个）类似学生档案、学籍数据库这样的关系表格。每个表格最多可有255列，每一列数据最多可以占256000字节。对于每一个关系表格，用户还可以根据需要定义多个使用视图。数据库定义语言和数据库操纵语言TQL有着相同的形式，建立、更新一个数据库也是比较方便的。

七、结 束 语

TDMS系统是在微型机上建立关系数据库的尝试。由于我们理论和实践水平不高，很多东西是干中学，学中干，系统还是很不完美的，欢迎批评指正。

（本文由TDMS主要设计人瞿中伟执笔）

参 考 文 献

- [1] C.J.Date, An Introduction to Database Systems, Second Edition, Addison-Wesley(1975,1977).
- [2] J.Martin, Computer Database Organization, Prentice-Hall(1975),

- [3] M.M.Zloof, Query by Example-A Database Management Language, IBM Research Report(1975).
- [4] 吴国平、赵美锡, 一个实时图形显示软件系统, 国防科技大学学报, 1981年第2期。
- [5] 曾民族、高崇谦, 电子计算机在情报工作中的应用, 科学技术文献出版社(1979).

A Miniature Relation Database Management System TDMS

Qu Zhong-wei Zhan Da-peng Zhao Mei-xi

Abstract

A miniature relational database management system TDMS(Table-Data Management System)implemented on microcomputer ALPHA is described. Now the system has been applied to the management of students' archives and statuses in our university.