

# 基于 RDBMS 建立层次数据模型

刘实 (内蒙古大学计算机科学系)

**摘要:**本文介绍在 RDBMS 中建立层次数据模型的两种方法,结合实例对两种方法的优缺点进行了分析和讨论。

**关键词:**RDBMS 关系数据库 层次模型

## 1. 引言

数据库中的数据从整体上看是有结构的数据,按其数据结构的组织方式的不同,可以分为三种数据模型,层次模型、网状模型、关系模型,相应的数据库则称为层次数据库、网络数据库、关系数据库。

早期的数据库均采用层次模型或者网状模型。从 1972 年 E. F. Codd 提出关系模型的概念,首次用数学的方法来研究数据库的结构和数据操作,将数据库的设计从以经验为主提高到以理论为指导。相应的关系数据库管理系统也逐渐成熟起来。目前人们普遍使用的数据库管理系统如 XBASE 系列、ORACLE、INFORMIX 等数据库管理系统,均是关系数据库管理系统(RDBMS)。

关系数据模型以及 RDBMS 以其数据独立性高、数据冗余度小、严格的数学理论指导、操作简单灵活等特点,已经逐渐取代了层次、网状模型而成为主流数据库产品。但在实际应用中,许多数据结构在逻辑上是一种层次关系。在 RDBMS 中,如何建立层次模型才能更好地提高效率,是本文要讨论的主要内容。

## 2. RDBMS 下建立层次数据模型

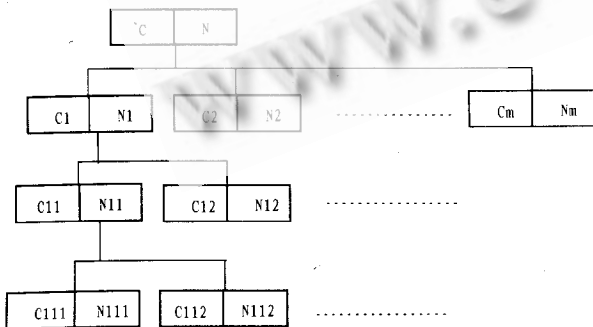


图 1 层次数据模型

(1)层次数据模型。层次数据模型实际上是一种树结构,图 1 给出了一个层次结构的示例。其中的  $C_i$  表示编码,  $N_i$  表示名字。对于实际的层次数据模型,其层数和结点个数一般是确定的,这有利于在微机上进行数据组织。

(2)RDBMS 中建立层次模型的两种方法

①集中组织法。以数据在每一层中的编码的拼接作为关键字组织数据库,其关键字的长度是各个层次上编码的长度之和。各个记录的关键字等长。上层关键字中长度不足的部分补'0'。在图 1 中,假设第一层的编码长度为 3 位,第二层和第三层的编码长度均为 1 位,第四层的编码长度为 2 位。则根结点的编码为 9990000(其中的每个 9 表示一位数字),  $N_1$  的编码为 9991000,  $N_{11}$  的编码为 9991100,  $N_{111}$  的编码为 9991101, 依次类推。相应的关系数据库模式为:模式名(编码, 名称, ...)如表 1 所示。

表 1 图 1 的集中组织数据库

编码	名称
9990000	N
9991000	N1
9991100	N11
9991101	N111
9991102	N112
.....	

这种组织方法虽然通过编码的前缀可以看出其层次关系,但由于不同层次的数据没有明显的区别,操作效率受到一定的影响。

②分层组织法。每层建立一个数据库,层次之间的关系通过数据库的关键字来建立联系。对于图 1 所示的层次模型,可以用 4 个数据库来表示。

·第一层数据库。库名(一级编码, 名称)。

- 第二层数据库。库名(一级编码,二级编码,名称)。
- 第三层数据库。库名(一级编码,二级编码,三级编码,名称)。
- 叶子数据库。库名(一级编码,二级编码,三级编码,四级编码,名称)。

按照这种方法组织数据库,各个数据库的候选关键字是唯一的,均符合 3NF 范式。数据库之间的操作可以通过关系运算实现。由于是分层组织数据库,层次关系明确,易于操作。

### 3. 实例及分析

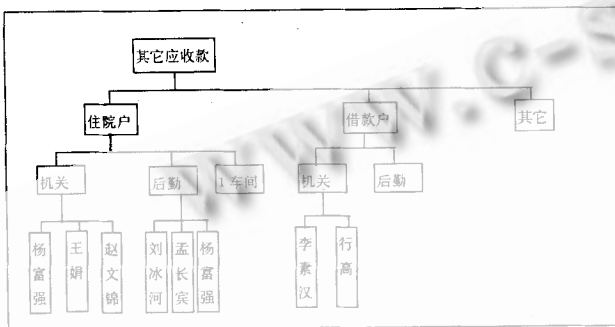


图 2 其他应收款的层次结构

图 2 给出了在财务管理系统中“其他应收款”及其下级科目的层次结构,我们给出两种组织方法下建立的数据库,并对其操作效率进行简单的分析。

(1)集中组织法建立的数据库。按照关键字拼接的方法集中组织数据库,为了讨论的方便,增加了科目余额字段。数据库如表 2 所示。

为了增加可以表示的中间结点的个数,这里的二级、三级科目均采用字母来编码。该库中的记录个数即层次结构的结点个数之和。对于一个实际的单位来讲,此库的记录往往有数百条之多。因此此库占用存储空间较多。

#### (2)分层组织的数据库

##### ①一级科目数据库(YJKMK.DBF)

科目号	科目名称
119	其他应收款

表 2 其他应收款的数据库

科目号	科目名称	科目余额
1190000	其他应收款	
119A000	住院户	
119AA00	机关	
119AA01	杨富强	2000.00
119AA02	王娟	2500.00
119AA03	赵文锦	200.00
119AA04	郭绣花	1500.00
.....		
119AB00	后勤	
119AB01	刘冰河	2010.00
119AB02	孟长宾	300.00
119AB03	杨富强	400.00
.....		
119B000	借款户	
119BA00	机关	
119BA01	李素汉	3000.00
119BA02	行高	440.00
.....		
119BB00	后勤	
119BB01	游宝三	4000.00
119BB02	梁关朗	234.00
.....		

##### ②二级科目数据库(EJKMK.DBF)

科目号	二级科目号	科目名称
119	A	住院户
119	B	借款户
119	C	其他

##### ③三级科目数据库(SJKMK.DBF)

科目号	二级科目号	三级科目号	科目名称
119	A	A	机关
119	A	B	后勤
119	A	C	一车间
.....			
119	B	A	机关
119	B	B	后勤
119	B	C	一车间
.....			

(接下页)

④明细科目库(MXKMK.DBF)

科目号	二级科目号	三级科目号	明细科目号	科目名称	科目余额
119	A	A	01	杨富强	2000.00
119	A	A	02	王娟	2500.00
119	A	A	03	赵文锦	200.00
119	A	A	04	郭绣花	1500.00
.....					
119	A	B	01	刘冰河	2010.00
119	A	B	02	孟长宾	300.00
119	A	B	03	杨富强	400.00
.....					
119	B	A	01	李素汉	3000.00
119	B	A	02	行高	440.00
.....					
119	B	B	01	游宝三	4000.00
119	B	B	02	梁关朗	234.00
.....					

(3)两种方法的比较。下面以两个操作为例,比较一下两种结构的优缺点。

①访问记录。在集中组织法中,由于各层记录不区分,要访问某个记录时,一般有这样两种方法。

·建索引之后顺序访问。由于记录个数一般较多,效率很低。

·按关键字设置过滤器。例如要访问住院户某人的信息,可设过滤器如下:

```
SET FILTER TO SUBS(科目号,1,5) = = 119AA
```

访问所有二级科目,过滤器如下:

```
SET FILTER TO SUBS(科目号,5,3) = = 000
```

设置过滤器之后,查询时面对的是满足条件的记录的子集,可以适当加快查询速度。

在分层组织法中,访问中间结点与下级结点无关,可以直接在上层数据库中进行操作。访问明细科目时,与中间结点也可以不发生关系,直接对子集操作,速度要快。

②查询结点名称。例如,“显示科目余额在 2000 元以上的住院户人员名单”,现在来看一下两种方法的区别。

·集中式。由于存储的科目名称仅仅是当前结点的名称,所以显示出来的名称无法说明其所在的单位。如住院户中有两个杨富强,其中的一个科目余额为 2000.00 元,但在科目名称中说明不了他是哪个单位的。如果在科目名称的前面加上其上级结点的名字,如科目名称改为“机关杨富强”“后勤杨富强”,虽然科目名称一目了然,将造成大量的数据冗余。

·分层式组织方法。在这种方法中,可以通过数据库的关联操作在各个数据库之间建立关联,需要上级结点的科目名称时,可以很方便地显示出来。方法如下:

```
SELE 0
USE YJKMK ORDER TAG KMH && kmh 为以科目号
为索引建立的索引标志(key 科目号)
SELE 0
USE EJKMK ORDER TAG EJKMH && key 科目号 + 二
级科目号
SELE 0
USE SJKMK ORDER TAG SJKMH && key 科目号 + 二
级科目号 + 三级科目号
SELE 0
USE MXKMK ORDER TAG MXKMH && key 科目号 +
二级科目号 + 三级科目号 + 明细科目号
SET RELA TO 科目号 INTO YJKMK
SET RELA TO 科目号 + 二级科目号 INTO EJKMK AD-
DI
SET RELA TO 科目号 + 二级科目号 + 三级科目号 INTO
SJKMK ADDI
DISP ALL FOR 科目余额 > = 2000.00 YJKMK. 科目名
称,EJKMK. 科目名称,SJKMK. 科目名称,科目名称,科目余额
这种方法没有数据冗余,节省存储空间,加快访问速度。
```

### 4. 结语

从以上的讨论可以看出,采用分层组织法,虽然比集中组织法用到的数据库要多,但在访问效率、数据冗余、数据存储等几个方面,均比集中组织法的效果好。因此,建议在 RDBMS 中建立层次数据模型时,尽量采用分层组织法以提高程序的效率和软件质量。