

嵌套关系 DB 实现时态 DB 系统的方法*

吴扬扬 陈锻生

(华侨大学计算机科学系, 泉州 362011)

摘要 讨论嵌套关系表达时态关系、嵌套元组演算表达时态元组演算的能力, 提出一种用嵌套关系数据库(DB)实现时态数据库系统的方法。

关键词 嵌套关系数据库, 时态数据库, 嵌套元组演算, 时态元组演算

分类号 TP 311.13

传统的数据库时时都在变化, 旧的数据被删除、修改, 新的数据被插入, 数据库中保存的只是当前的信息, 这对某些应用是不合适的。在实际应用中, 要求数据库不仅提供当前, 而且提供过去(历史)的数据, 数据库系统要具有处理时态信息的功能。时态数据库就是为适合这种需要而提出的。时态数据库的研究不能抛弃和脱离现有的数据库和数据库发展的主流, 应考虑如何在主流数据库模型上增加处理时态信息的能力。时态数据库的对象是三维的实体, 用传统的关系数据库(二维表)模拟它是勉强的, 因为一个对象的历史需要用几个 INF 元组才能表示^[1]。近年来, 非第一范式的理论、实现和应用研究均取得巨大进展, 出现一些基于这种模型的数据库管理系统^[2~4]。非第一范式方法的本质在于允许一个关系的属性是另外一个关系, 它可以描述大而复杂的结构, 用它来描述三维对象是比较方便、直观的。

1 嵌套关系数据库及其元组演算

M. A. Roth 等人提出的嵌套关系数据库模式, 是一个形如 $R_j = (R_{j1}, R_{j2}, \dots, R_{jn})$ 的规则集合, 其中 $R_{jn} (1 \leq j \leq n)$ 可以出现在其它规则的左边。也就是说, 允许一个关系的属性是另一个关系, 但不能出现递归定义。对这样一个数据模型, Roth 扩充关系元组演算定义了嵌套元组演算。嵌套元组演算表达式的一般形式为 $\{t | \psi(t)\}$, 其中 t 是元组变量, ψ 是由下列原子公式及运算符组成的公式, i 元组变量 t 可写成 $t^{(i)}$ 。

原子公式可为下列四种形式。(1) $s \in r$, 其中 s 是元组变量, r 为关系名, 表示 s 是 r 中的元组。(2) $s \in t[i]$, 其中 t, s 均为元组变量, 表示 s 是 t 的第 i 个分量所指关系的元组。这里, t 的第 i 个分量是一个元组集合。(3) $a \theta s[i], s[i] \theta a, s[i] \theta t[j]$, 其中 s, t 均为元组变量, a 是常数, θ 为比较运算符($=, >$)。常数 a 可以是单值, 也可以是值的非空集合, 非单值常数只能用“=”号。(4) $s[i] = \{u | \psi'(u)\}$, 其中 ψ' 是带有自由变量 u 的公式, 表示 s 的第 i 个分量是满足

* 本文 1993-11-03 收到

ψ' 的元组集合.公式可含运算符: $\neg, \wedge, \vee, \forall, \exists$.例如,设嵌套关系 r 的模式为 $R=(A, B, E'), B=(C, D'), D'=(D), E'=(E)$. (1)找出 r 中属性 C 的某个值等于 c_1 ,且属性 D 的某个值等于2的元组.有 $\{t \mid t \in r \wedge (\exists s)(s \in t[2] \wedge s[1]=c_1) \wedge (\exists u)(u \in s[2] \wedge u[1]=2)\}$. (2)在 r 的每一个子关系 B 中,删去属性 D 不大于6的值,得到一个新元组集合 r' . $\{t^{(3)} \mid (\exists s)(s \in r \wedge t[1]=s[1] \wedge t[3]=s[3] \wedge t[2]=\{u^{(2)} \mid (\exists v)(v \in s[2] \wedge u[1]=v[1] \wedge u[2]=\{w^{(1)} \mid w \in v[2]) \wedge w[1]>6\})\})\}$.设 r 为下述(i),则(1),(2)的结果分别为(ii),(iii).

(i)的结果

A	B		E'
	C	D'	E
		D	
a_1	c_1	11	e_1
		2	e_2
	c_2	1	e_3
a_2	c_2	11	e_1
		14	e_2
	c_2	16	e_3

(ii)的结果

A	B		E'
	C	D'	E
		D	
a_1	c_1	11	e_1
		2	e_2
	c_2	1	e_3

(iii)的结果

A	B		E'
	C	D'	E
		D	
a_1	c_1	11	e_1
		14	e_2
	c_2	16	e_3

关于嵌套元组演算的安全约束和完备性讨论,详见文[2].

2 时态数据库及其元组演算

S. K. Gadia 提出用 $\rightarrow 1NF$ 方法模拟现实世界中事件的时间信息,给出一个模型和基于这种 $\rightarrow 1NF$ 方法的时态元组演算.规定时间论域 T 为 $[0, \text{now}]$, now 指当前时间,时态元素是 $[0, \text{now}]$ 的有限的子区间的并.定义一个数据库模式 R ,是关系模式 $\{R_1, R_2, \dots, R_n\}$ 的有限集合.关系模式 R 上的一个时态数据库,是 $R_i (1 \leq i \leq n)$ 上时态关系 $\{r_1, r_2, \dots, r_n\}$ 有限集合. $R_i (1 \leq i \leq n)$ 上的时态关系 $r_i (1 \leq i \leq n)$,是 R_i 上时态元组的有限集合.

基于上述数据模型的时态元组演算TCAL,是这样定义的.时态元组演算表达式的一般形式为 $\{x \mid f(x)\}$,其中 x 为时态元组变量, f 是带有自由变元 x 的公式.下面介绍公式,关系表达式和时态表达式. (1) $r(x)$ 是原子公式,其中 r 是关系, x 为元组变量,表示 x 属于 r . (2) $x(A)\theta y(B), x(A)\theta_{c_2, c_1} y(B)$ 也是原子公式,其中 x, y 为元组变量, A, B 为属性, $c_1 \in \text{dom}(A), c_2 \in \text{dom}(B), \theta$ 为强比较算符(strong θ comparator) θ ,它要求被比较的对象要具有相同的时态域. (3)若 f 是公式,则 $\neg f$ 也是公式. (4)若 f, g 是公式,则 $f \wedge g, f \vee g$ 也是公式. (5)若 f 是公式, x 为自由变元,则 $(\exists x)f(x), (\forall x)f(x)$ 也是公式. (6)若 v 是由 T 中的时态元素,则 v 是一个原子时态表达式. (7)若 v_1, v_2 是时态表达式,则 $(v_1), \bar{v}, v_1 \cup v_2, v_1 \cap v_2, v_1 - v_2$ 均为时态表达式. (8)若 $f(x)$ 是具有自由变元 x 的公式,则 $\perp x f(x)$ 是一个时态表达式,表示求 x 的时态区间. (9) $x:v$ 是公式,其中 v 是时态表达式, x 是元组变量,表示 x 的时态包含在 v 之中. (10)若 $f(x)$ 是带有自由变元 x 的公式,则 $\{x \mid f(x)\}$ 是关系表达式.

例如,设有关系模式EMP(姓名,工资,部门)和MANAGER(部门,主任). (1)列出张三当主任时,电子系所有职工的姓名及工资. $\{x(\text{姓名}, \text{工资}) \mid (\exists y)(\text{EMP}(y) \wedge y(\text{部门}) = \text{电子系})$

$\wedge x(\text{姓名})=y(\text{姓名}) \wedge x(\text{工资})=y(\text{工资}) \wedge y: \perp z(\text{MANAGER}(z) \wedge z(\text{主任})=\text{张三})$. (2)
给出在计算机系还没有主任时, 数学系的职工名. $\{x(\text{姓名}) | (\exists y)(\text{EMP}(y) \wedge x(\text{姓名})=y(\text{姓名}) \wedge y(\text{部门})=\text{数学系}, \wedge y: \perp z(\text{MANAGER}(z) \wedge z(\text{部门})=\text{计算机系}) \sim)\}$. 关于时态元组演算安全表达式及公式的解释, 详见文[5]. S. K. Gadiate 在文[5]中对 T 是 $[0, \text{now}]$ 上的离散量这一条件作了讨论, 指出这一假设可简化许多概念, 保证 T 的任一子集都是时态元素, T 上的任一时态元素均可表示成 T 上的闭区间的并, 并且是有限集合. 为简化讨论, 下面的论述均假设 T 是离散量.

3 时态关系的嵌套关系表示

一个实体往往含有两类属性, 一类是随时间变化的, 如工资、所在部门, 称为时态属性. 另一类是固定不变的, 如姓名、性别、出生年月等, 称为非时态属性. S. K. Gadia 在文[5]中对 $\rightarrow 1NF$ 时态数据库要求时态元组是均匀的, 即同一元组中各属性的时态域是相同的. 这个时态域就是该元组的时态域, 指出了该元组的有效时间. 因此, 一个 S. K. Gadia 的时态数据库的关系模式可用一个带有嵌套的关系来表示, 即有效时间 (VT), 属性 1, 属性 2……属性 n . 其中, 有效时间 (VT) 又是一个包含起始时间 (ST) 和结束时间 (ET) 两个属性的关系. 时间信息的精度可通过关系模式 VT 的定义来确定, 关系中的时态属性, 也用一个嵌套关系表示, 这个关系由有效时间 VT 和属性组成. 例如, 职工关系 EMP 和经理关系 MANAGER 为表 1, 2. 其中, EMP 中的职工号、姓名、性别, MANAGER 中的部门, 都是非时态属性其余均为时态属性.

表1 关系 EMP

职工号	姓 名	性 别	所在部门
[1980, Now]0001	[1980, Now]王刚	[1980, Now]男	[1980, 1984]数学系 [1985, Now]计算机系
[1984, 1990]0002 U[1993, Now]	[1984, 1990]陈红 U[1993, Now]	[1984, 1990]女 U[1993, Now]	[1984, 1990]U[1993, Now] 计算机系
⋮	⋮	⋮	⋮

表2 关系 MANAGER

部 门	主 任	人 数
[1980, Now] 计算机系	[1980, 1982]张山 [1988, Now]王刚	[1980, 1981]40 [1982, 1985]45 [1986, Now]56
⋮	⋮	⋮

用 M. A. Roth 的嵌套关系数据库模式定义规则, 可定义一个相应的嵌套数据库模式. $\text{EMP} = (\text{VT}, \text{职工号}, \text{姓名}, \text{性别}, \text{所在部门})$, $\text{VT} = (\text{ST}, \text{ET})$ (时间假设只考虑到年), $\text{所在部门} = (\text{VT}, \text{部门})$; $\text{MANAGER} = (\text{VT}, \text{部门}, \text{主任}', \text{人数}')$, $\text{主任}' = (\text{VT}, \text{主任})$, $\text{人数}' = (\text{VT}, \text{人数})$.

上述两个时态关系可用嵌套关系表示为表 3, 4.

表3 关系 EMP

VT		职工号	姓名	性别	所在部门		
ST	ET				VT		部门
					ST	ET	
1980	now	0001	王刚	男	1980	1984	数学系
					1985	now	计算机系
1984	1990	0002	陈红	女	1984	1990	计算机系
1993	now				1993	now	
:	:	:	:	:	:	:	:

表4 关系 MANAGER

VT		部门	主任'		主任	人数'		人数
ST	ET		VT			VT		
			ST	ET		ST	ET	
1980	now	计算机系	1980	1982	张山	1980	1981	40
			1983	now	王刚	1982	1985	45
						1986	now	56
⋮	⋮		⋮			⋮		

4 嵌套元组演算的时态表达能力

从1,2节可知,嵌套元组演算和时态元组演算,是在传统的关系元组演算的基础上扩充而成的. 嵌套元组演算增加嵌套关系数据的描述,而时态元组演算则增加时态元素和时态运算的表示. 由此,我们只需讨论是否能用嵌套元组演算表示时态元素以及与时态有关的运算.

4.1 时态元素和时态表达式对应的嵌套元组演算公式

按第3节所给的方法,一个时态元素对应 VT 上的一个关系. 例如,EMP 关系中的王刚,在数学系的时间[1980,1984]对应关系为(1);陈红对应元组的时态域[1984,1990]∪[1993,now],也对应 VT 上的一个关系为(2).

(1)VT_i

ST	ET
1980	1984

(2)VT_j

ST	ET
1984	1990
1993	NOW

T 的的任一闭区间[*t*₁,*t*₂],均可用嵌套元组演公式 VT(1)≥*t*₁ ∧ VT(2)≤*t*₂表示. 因此,任何含有~,∪,∩,—运算符的时态表达式,均可与一个嵌套元组演算公式对应.

4.2 含有⊥*x* 的时态表达式的嵌套元组演算表示

⊥*x* 表示求出元组 *x* 的时态元素,用嵌套关系表示后,相当于列出元组 *x* 的第一个分量的

所有元组. 比如, 求陈红在职的时间, 为 $\{t^{(2)} \mid (\exists x)(x \in \text{EMP} \wedge t \in x[1] \wedge x[3] = \text{陈红})\}$.

4.3 $x:v$ 的嵌套元组演算公式表示

$x:v$ 表示要求 x 的时态包含在 v 之中. 用嵌套关系表示, 相当于对 x 的查询增加一个条件, 即要求 x 的第一个分量的元组所表示的有效时间, 包含在 v 所表示的时间区间内, 例如, 要求找出1985年以前, 离退休的职工为 $\{t \mid t \in (\text{EMP} \wedge \neg(\exists s)(s \in t[1] \wedge s[2] \geq 1985))\}$.

4.4 关于强比较算符(Strong θ -comparator)

在时态元组演算中, 强比较算符要求被比较的对象, 必须具有相同的时态域. 用嵌套元组演算公式表示, 就是在比较条件上, 增加一个要求对应属性的有效时间要相同的条件, 比如, 找出人数比计算机系多部门, 为 $\{t \mid t \in (\text{MANAGER} \wedge (\exists s)(s \in \text{MANAGER} \wedge s[2] = \text{计算机} \wedge (\forall u)(u \in s[4] \wedge (\exists v)(v \in t[4] \wedge u[1] = v[1] \wedge v[2] > s[2]))))\}$.

由以上讨论可见, 能用时态元组演算表达的查询, 均可用嵌套元组演算来表达.

5 结束语

嵌套关系数据库能表示和处理时态信息, 是本文所解决的问题. 但对于如何更准确地体现时间所具有的特性, 如何改进时态关系模型以解决时态操作中的一些问题, 我们将有进一步的研究课题. 在主流数据库上增加时态信息处理, 是数据库研究的一个重要内容. 随着计算机存储技术的不断发展, 时态数据库必然具有广阔而现实的应用前景.

参 考 文 献

- 1 格兰特著. 数据库逻辑导论. 古新生等译. 西安: 西安交通大学出版社, 1988. 319~326
- 2 Roth M A, Korth H F. Extended algebra and calculus for nested relational database. ACM trans on Database Systems, 1988, 13(4): 389~417
- 3 Levene M, Lcizou G. The nested relational type model an application of domain theory to database. The Computer Journal, 1990, 33(1): 19~30
- 4 Harris E P, Kent A J, Davis R S. An environment for building graphical user interfaces for nested relational database. information systems, 1992, 17(1): 107~115
- 5 Gadia S K. A homogeneous relational model and query languages for temporal database. ACM TODS, 1988, 13(4): 418~448

A Method for Implementing Temporal DB System in Nested Relational DB

Wu Yangyang Chen Duansheng

(Dept. of Computer Science, Huaqiao Univ., 362011, Quanzhou)

Abstract In this paper, the author discusses the feasibility of expressing temporal relations in nested relations and temporal tuple calculus in nested tuple calculus; and proposes a method for implementing temporal DB system in nested relational DB.

Keywords nested relational DB, temporal DB, nested tuple calculus, temporal tuple calculus