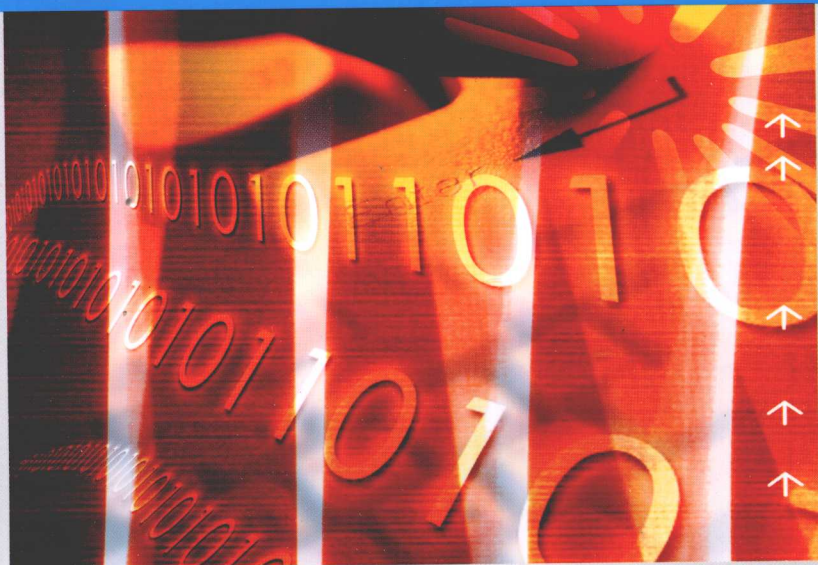


时态数据库

设计理论

郝忠孝 著



科学出版社

www.sciencep.com

TP311.13/374

2009

时态数据库设计理论

郝忠孝 著

科学出版社

北 京

内 容 简 介

本书是在作者三十余年来对数据库理论研究的基础上撰写的。书中系统论述和分析了时态数据库数据组织理论和基于时态 ER(实体-联系)模型的概念设计研究成果以及作者提出的若干新的概念、方法、算法。

本书共分七章。主要内容包括:时态数据库设计研究的概况、时态模块模式与时态模块、时态函数依赖和集细于关系、时态类型和它的封闭集处理、TFD 的推导规则、时态类型处理中的细于关系矩阵、时态类型集的封闭集及其公共最大下界、有限导出时态类型集、*glb* 操作的实现算法。通过计算一个时态类型集的封闭集以及它的细于关系矩阵可以在计算机上实现对时态类型的相关处理。讨论了基于全序、偏序 TFD 集的时态模式规范化,基于 TMVD 的时态模式的规范化,强全序时态模式中多值依赖问题,基于时态 ER 模型的时态数据库设计。

本书可作为计算机科学与技术、数据库及相关专业的高年级本科生教材或硕士生选修课教材,也可供从事上述领域研究的博士生、科研人员及工程技术人员参考。

图书在版编目(CIP)数据

时态数据库设计理论/郝忠孝著. —北京:科学出版社,2009
ISBN 978-7-03-024704-9

I. 时… II. 郝… III. 数据库系统-程序设计 IV. TP311.13

中国版本图书馆 CIP 数据核字(2009)第 090784 号

责任编辑:耿建业 汤 枫/责任校对:陈玉凤
责任印制:赵 博/封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

丽源印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009 年 6 月第 一 版 开本:B5(720×1000)

2009 年 6 月第一次印刷 印张:12

印数:1—2 500 字数:224 000

定价:40.00 元

(如有印装质量问题,我社负责调换〈路通〉)

作者简介



郝忠孝，教授，山东蓬莱人，1940年12月生，中共党员，曾任原东北重型机械学院副校长，齐齐哈尔大学副校长，哈尔滨理工大学学术委员会主席。现任哈尔滨理工大学博士生导师、哈尔滨工业大学博士生导师（兼）。原机械电子工业部有突出贡献专家、享受国务院政府特殊津贴、全国优秀教师、

省共享人才专家、省级学科带头人、省计算机学会副理事长。

主要研究领域：①空值数据库理论。在国内、外首次提出了空值数据库数据模型，完成一系列相关研究，形成了比较完整的理论体系，完成了国内、外第一部该方面的论著《空值环境下数据库导论》。②数据库 NP-完全问题的求解问题。首次基本解决了求全部候选关键字、主属性、基数为 M 的候选关键字、最小候选关键字等问题，完成了《关系数据库数据理论新进展》一书。③数据库数据组织的无环性理论研究。首次给出了归并依赖集、关联度等概念。并在此基础上对无 α 环、无 β 环、无 γ 环的分解条件与规范化理论研究方面有了突破性进展，著有《数据库数据组织无环性理论》。④时态数据库理论研究。系统提出并完成了时态数据库中基于全序、偏序、多粒度环境下的各种时态理论问题研究，完成了《时态数据库设计理论》的论著。⑤主动数据库理论研究。完成了国内、外第一部该方面的论著《主动数据库系统理论基础》。⑥空间、时空数据库理论研究。

完成了国家、省部级项目 10 项；获省部级科技进步奖一、二、三等奖 5 项。发表学术论文 200 余篇，其中，国家一级论文 130 余篇、在《计算机研究与发展》上发表个人学术论文专辑两部，被 SCI、EI 等检索 80 余篇。1991 年发表学术论文数居中国科技界第五位（并列）。著书五部。

前 言

由于现实世界是不断演变进化的,时间是反映现实世界信息的基本组成部分,因而大多数数据库应用程序都有时态的特性。例如,地震资料分析应用程序、天气监测应用程序、天气预报分析程序、资源管理应用程序、银行等财经类的应用程序、项目管理等记录性应用程序。传统的数据库管理系统对时态信息的存储、处理和操作都十分有限,正因为传统数据库缺乏对时态数据的支持,因而在很多方面产生了问题。例如,它把时间数据作为一个字段的值进行存储和管理,只反映了对象某一个时刻或当前时刻的信息和状态,不联系对象的历史、现在和将来,无法将对象的历史、现在和将来作为对象的一个发展过程来看待,而这样做无助于解释事物发展的本质规律。抓住事物的发展趋势这一点对于决策支持系统这类应用程序来说是很基本、很重要的;同时要求管理数据库系统中元事件的时态信息,例如数据库被查询修改的时刻、时间区间。多用户系统中对锁定排队以及资源竞争协调的时标等,这些时态数据也有助于提高数据库系统的可靠性和效率。随着数据库技术的不断发展,人们开始逐渐意识到必须为时态数据建立时态数据库的模型,或者在现有的数据库模型上加以改造,于是提出了时态数据库的概念。1982年美国的 Zvi 在他的论文“The Time Relational Model”中首次提出了时态数据库和几个重要的概念,这是时态数据库形成的标志。1993年美国的 Tansel, Clifford 等所著的 *Temporal Databases: Theory, Design and Implementation* 一书中首次对时态数据库给出了较为全面的论述。

所谓时态数据库,是指能够处理时间信息的数据库。传统数据库只记录了数据的当前状态,在现实情况改变时,数据库也发生变化。而时态数据库不仅存放对象的现状,而且存放对象过去的一切状态,并且可以根据对象现在和过去的状态推测其未来可能的状态。

时态数据库技术已成为一个诱人且活跃的研究领域。时态数据库在宏观上有下面两个方面的特性:①动态性。传统的数据库系统对数据进行静态或准动态的数据库管理。在数据更新时,过时的数据将从数据库中删除,这就不能反映出现实世界的动态过程。例如,李明的专业技术职务是 1991.5~1994.3 为助教,1994.4~2001.6 为讲师,2001.7 至今为副教授。在时态数

数据库中,过时的数据不再从数据库中删除,对历史数据也可以进行更新,使系统和现实世界一直保持着全方位的动态交换。②全面性。由于时态数据库是所有数据的集合体,可以提供任何时刻和时间段的数据,这是传统数据库无法比拟的。时态数据库为历史、当前和将来进行对比、分析、监测和预测预报提供了丰富的数据,从而为预测预报系统、决策支持系统和其他分析系统服务。

目前,对时态数据库已进行了大量的研究。当前对时态数据库的许多方面如在数据模型、实现技术方面取得了许多重要成果,但是对于时态数据库设计理论等方面的研究尚不够深入,这直接影响着一个具有优良性能的时态数据库的设计和应用。为此,本书重点讨论了时态数据库设计理论等方面的问题,力求对具有优良性能的时态数据库的设计和应用有所帮助。

为了达到对时态数据库设计的目的,本书以时态数据库应用设计为主线,对时态模块模式与时态模块、时态函数依赖和集细于关系、时态类型和它的封闭集处理、时态类型处理中的细于关系矩阵、时态类型集的封闭集及其公共最大下界、有限导出时态类型集等进行了讨论,给出了 glb 操作的实现算法。通过计算一个时态类型集的封闭集以及它的细于关系矩阵可以在计算机上实现对时态类型的相关处理。提出了全序时态类型集、关联集等概念和相关操作。给出了求全序时态类型 TFD 集成员籍的算法,一个具有全序时态类型 TFD 集的满足时态模块模式的保持依赖、无损连接的 T3NF 分解的多项式时间的分解算法,研究的算法对时态数据库设计来说具有普遍的意义。给出了规范程度高于时态三范式、低于时态 BC 范式的时态初等关键字范式和时态简单范式。定义了部分、传递时态、包含时态和隐含时态函数依赖概念,提出了弱时态第三范式 (WT3NF) 和一个得到保持依赖的无损连接的 WT3NF 分解的分解算法,消除了时态非主属性对时态码的部分、传递、包含和严格隐含时态依赖,该算法更加有效和易于实现,可以真正通过计算机实现多时间粒度下的时态数据库设计。讨论了非严格偏序时态类型集,偏序时态模块模式,模式、模块投影和偏序时态 BC 范式等概念,给出了避免时态类型间复杂操作的偏序时态 BC 范式的分解算法。给出了有限导出时态类型集的概念和它的求解算法,提出了一种新的关系——强细于关系及时态类型与时态类型集之间的集强细于关系、支持集等概念。讨论了时态类型集的强封闭集、属性集的有限闭包、有限依赖基和特殊有限依赖基等问题,给出了相应的基本定理、求解算法。并对 TFD 和 TMVD 逻辑蕴涵的判定、时态模式的时态 T4NF 范式分解算法等进行了讨论。对强全序时态

模式中的多值依赖问题进行了深入研究,给出了规则的时态多值依赖(RTMVD)、属性的时间粒度等概念,提出了强全序时态模块模式无损连接分解间关系的相关定理。并且对强全序时态模式中混合依赖集的规范化问题进行了讨论,给出了强全序模式下无冗余时态混合依赖覆盖算法和全序时态混合规范覆盖算法。给出了 TEERM 模型,在模型中引入了变化粒度和时间基数等新的结构,通过这些结构可以描述时态属性和联系,并且能够规范 TFD 约束。模型中没有显式地引入任何时间戳属性,并且属性的变化粒度和实体类型的相对变化粒度可以规范任意的时态类型,这隐含着 TEERM 模型可以支持有效、事务和用户定义的时间。通过提出的 TFD 生成规则以及向时态模块模式投影过程,可以容易地实现由 TEERM 模型转换成具有 TFD 约束的时态模块模式。由于传统 ER 模型中的每一个结构在 TEERM 模型中都有能够表达相同语义的结构与之对应,因此,TEERM 模型既可以用于时态数据库设计也可以用于非时态数据库设计。无论对于传统应用还是时间相关的应用,基于 TEERM 模型的数据库设计方法都能够有效地进行数据库设计。

本书的另一个目的是作者从事数据库理论研究工作三十余年,作为一种责任想把研究的一些结果留给年轻同志,使有兴趣的同行和读者深入到这一领域。同时,本书也是作者独著书的收官之作。

由于本书所讨论的内容绝大部分是新的,部分理论和方法尚未来得及实践和验证,加之作者水平有限,书中难免存在不妥之处,敬请同行和读者批评指正,谢谢!

在本书出版之际,特别感谢哈尔滨理工大学对我的鼓励和支持。同时,本书的出版得到了科学出版社的大力支持和帮助;学生姚春龙、李艳娟、万静等同学做了大量有成效的工作,特别是万静博士审校了全书,另外,本书的所有图表的绘制部分由李博涵同学完成,在此表示诚挚的感谢!

郝忠孝

2009年3月于哈尔滨

目 录

前言

第 1 章 时态数据库设计研究的概况	1
1.1 时态数据库设计研究概况	1
1.2 时态 ER 模型的研究概况	5
1.3 时态数据库的类型和查询语言概述	6
1.3.1 时态数据库元素的类型	6
1.3.2 时态查询语言	8
1.4 本章小结	9
第 2 章 时态类型和它的封闭集处理	10
2.1 时态类型和时态函数依赖	10
2.1.1 时态类型	10
2.1.2 时态模块	13
2.1.3 时态函数依赖和集细于关系	14
2.1.4 TFD 的推导规则	16
2.1.5 时态类型处理	17
2.2 封闭时态类型集	19
2.2.1 封闭集与公共最大下界	19
2.2.2 封闭集算法	20
2.3 本章小结	29
第 3 章 基于全序 TFD 集的时态模式规范化	30
3.1 全序 TFD 集特性和成员籍	30
3.1.1 全序时态类型集的 TFD 集的逻辑蕴涵	30
3.1.2 求全序时态类型 TFD 集成员籍的算法	32
3.1.3 求时态候选关键字算法	34
3.1.4 属性集的有限闭包	34
3.2 TFD 集的化简	38
3.3 时态 TFD 集 F 规范化的基本概念	41
3.4 T3NF 分解	43
3.4.1 T3NF 分解算法	43

3.4.2 应用实例	47
3.5 时态初等关键字范式 and 简单范式	50
3.5.1 时态初等关键字范式分解	50
3.5.2 时态简单范式分解	56
3.6 本章小结	59
第4章 基于偏序 TFD 集的时态模式的规范化	60
4.1 TFD 集成员籍算法	60
4.1.1 基于有限决定集的成员籍算法	60
4.1.2 基于属性集的有限闭包的成员籍算法	67
4.1.3 基于有限导出时态类型集的成员籍算法	72
4.1.4 三种算法的比较	76
4.2 基于偏序 TFD 集的时态模式的弱时态第三范式	77
4.2.1 弱时态第三范式的定义	77
4.2.2 WT3NF 分解算法	80
4.3 WT3NF 分解算法的有效性与实用性	86
4.3.1 WT3NF 分解算法的有效性	86
4.3.2 WT3NF 分解算法的实用性	86
4.4 应用实例	88
4.5 基于偏序 TFD 集的时态模式的时态 BC 范式	90
4.5.1 偏序时态 BC 范式	91
4.5.2 偏序时态 BC 范式分解算法	94
4.6 本章小结	99
第5章 基于 TMVD 的时态模式的规范化	100
5.1 时态多值依赖的推导规则	102
5.1.1 多值依赖的概念	102
5.1.2 MVD 集 M 的依赖基	102
5.1.3 时态多值依赖的概念	103
5.1.4 时态多值依赖的推导规则	104
5.2 TFD 和 TMVD 的相互作用	109
5.3 成员籍问题	110
5.3.1 时态类型集的强封闭集	110
5.3.2 属性集的有限闭包、有限依赖基和特殊有限依赖基	111
5.4 有限依赖基和特殊有限依赖基的基本定理、求解算法	112
5.5 TFD 和 TMVD 逻辑蕴涵的判定	119

5.6 时态模式的时态 T4NF 范式分解算法	123
5.7 本章小结	126
第 6 章 强全序时态模式中多值依赖问题	127
6.1 规则时态多值依赖	127
6.1.1 规则时态多值依赖	127
6.1.2 强全序 RTMVD 推导规则	129
6.2 RTMVD 与强全序时态模块模式无损分解	132
6.3 强全序时态模式中混合依赖集成员籍	136
6.3.1 强全序模式混合依赖基算法	137
6.3.2 强全序模式混合闭包算法	142
6.3.3 强全序时态模式中成员籍问题	144
6.4 强全序时态混合依赖覆盖	147
6.5 本章小结	152
第 7 章 基于时态 ER 模型的时态数据库设计	153
7.1 TEERM 模型的结构	154
7.1.1 实体类型	154
7.1.2 属性	154
7.1.3 联系类型	156
7.1.4 指派实体类型的 TIK	163
7.2 规范 TFD 约束	164
7.2.1 规范 TFD 约束的结构	165
7.2.2 规范 TFD 的规则	165
7.3 向时态模块模式投影	166
7.3.1 几个操作	166
7.3.2 投影算法	167
7.3.3 简单分析	171
7.4 基于 TEERM 模型的数据库设计	172
7.4.1 支持传统应用	172
7.4.2 基于 TEERM 的数据库设计方法学	173
7.4.3 应用实例	173
7.5 本章小结	175
参考文献	177

第 1 章 时态数据库设计研究的概况

1.1 时态数据库设计研究概况

时态数据库的研究已有二十多年的时间,在数据模型、实现技术以及数据库设计理论等方面取得了大量的成果。时态数据库设计理论主要涉及两方面的研究:基于时态数据依赖的时态数据库规范化理论研究和基于时态 ER(实体-联系)模型的概念设计研究。下面首先介绍一下有关时态数据模型及实现技术方面的研究概况。

任何数据库设计都离不开具体的数据模型。经过多年的研究,目前已经提出了几十种时态数据模型,其中绝大部分是基于关系模型的,只有少数模型是基于面向对象模型的。时态数据库一般要支持两种无关的时间类型:有效时间和事务时间。有效时间是指现实世界中信息有效的时间;事务时间是指数据库中更新所发生的时间。在这些模型中,一些模型两种时间类型都支持,一些模型只支持有效时间(通常称历史数据库),而有一些模型只支持事务时间。

在研究人员提出大量的时态数据模型的同时,时态数据查询语言的研究也取得大量的成果,出现了近几十种查询语言。这些查询语言大多是时态关系查询语言,也出现了一些时态面向对象查询语言。不同的查询语言是基于不同的数据模型的,在这些数据模型中,多数是基于 1NF(第一范式)关系模型的。实际上,由于关系数据模型具有完整的数学形式化的描述和较为完备的理论基础,因此,大量的时态数据模型是基于关系模型的。目前市场上占主导地位的数据库管理系统软件都是基于关系的,采用时态关系模型有利于时态数据库的实现和产品化。

在时态数据库实现方面,有两种实现方法:①在 DBMS 中内部创建时态支持,这种方法需要对 DBMS 进行时态扩充,使其能满足相应的时态数据模型,并要支持新的数据类型;②在用户和现有的 DBMS 之间建立一个中间件,用它来接受用户应用的时态请求,并将时态数据查询语句转换成 DBMS 所支持的查询语句(例如 SQL 语句)。对于第①种方法,应用代码可以简化且更容易维护,在很大程度上可提高编程的效率,并且绝大多数的数据处理从应用转

移到 DBMS 来处理,这样可以得到良好的性能。但当前的数据库产品提供的内建的时态支持只限于预定义的与时间相关的数据类型。这些扩展可以使用户定义新的时间数据类型。然而时态的支持远远超出只是数据类型的要求,时态支持应当在扩展查询语言的同时,在优化和处理中封装时态操作。但是,开发一个内部创建时态支持的 DBMS 是一个令人生畏的任务,这种设计只有具有基础的代码可以修改,并可获得大量资源的 DBMS 供应商才可能实现。这样就导致人们采用一种分层的方式,即第②种方法。这种方法实现较为简单,不需要对 DBMS 进行更改,对于关系的 DBMS 只需把用户的时态应用所要求的时态查询语句转化为 SQL 语句,需要的话可以利用时态查询优化技术在中间件对查询进行优化。

从上面的讨论可以看到,关系模型在时态数据模型、查询语言和实现等各方面的研究中都处于核心地位。而且现在的较有影响的数据库产品大多是基于关系模型的,因此,近年来基于关系数据模型的时态数据库设计理论研究成为重要的研究领域。时态数据库设计理论的研究主要包括两个方面:基于时态数据依赖的时态数据组织理论研究和时态 ER 模型的研究。

数据依赖是数据库设计理论中的一个核心概念,通过它可以规范属性之间满足的固有的语义约束。在关系数据库设计中,最常见的两种依赖约束是函数依赖 FD 和多值依赖 MVD。研究这两种依赖的关系数据规范化理论的目的是为了消除数据冗余和各种异常现象。对于时态数据库,传统的数据依赖不能很好地适应时态数据的动态性。为了能够刻画属性之间满足的动态的语义约束,通过扩展传统的依赖理论,已经提出了多种时态数据依赖。下面分别介绍一些主要的时态数据依赖以及基于它们的范式理论。

(1) Vinau 于 1987 年提出了动态函数依赖 DFD 的概念,他把时态关系看作在时间上的快照关系的序列,每一个新的快照关系是从前一个快照关系通过同时影响一些(或全部)属性的全局更新得到的。例如,对于一个 DFD: $A \rightarrow \hat{C}$,表明对于任意两个元组在更新前具有相同的 A 值,更新后必须有相同的 C 值,DFD 反映了时态数据的一种动态的约束。

(2) Navathe 和 Ahmed 在 1989 年基于 TRM 模型通过时态独立属性的概念提出了另一种时态数据依赖。任意两个时间变化属性是时态独立的,如果它们的值不总在相同的时间改变。Navathe 和 Ahmed 提出的时态范式认为一个时态关系中不应该包含时态独立属性,这样可以减小数据冗余。

(3) Jensen 等在 1994 年提出了一种依赖理论,他们认为数据冗余存在于时态关系的快照中,他们把时态关系看作快照关系的集合。由于每个快照是

一个传统的关系,因此,可以把传统的函数依赖 FD、多值依赖 MVD 和范式的概念应用于每个快照,这样通过传统的关系模式的规范化方法可以减少每个快照的数据冗余和各种异常。实际上 Jensen 等提出的时态函数依赖就是传统 FD 的直接扩展。

(4)结合 Vinau 和 Jensen 的依赖理论,Wijzen 于 1995 年定义了四种类型的依赖:快照函数依赖 SFD、动态函数依赖 DFD、时态函数依赖 TFD 和间隔依赖 ID。与 Jensen 的时态函数依赖相同,SFD 是时态关系中的一个快照满足的依赖,而其余三种依赖是快照序列间的依赖约束;与 Vinau 的动态函数依赖相同,DFD 约束相邻的两个有效时间状态;TFD 和 ID 约束一个多有效时间状态的序列。另外,Wijzen 在这四种依赖中都集成了对象标识。基于所提出的依赖理论,Wijzen 给出了相应的时态范式的定义,并提出了分解策略。

(5)Wang 等于 1997 年基于多时间粒度提出了一个新的时态函数依赖 TFD 的概念。Wang 的 TFD 能够灵活地捕捉时间变化属性信息,并通过时态类型的概念,在一定程度上能约束属性值的变化率和支持用户定义的时间类型。例如,对于某个时态模式的两个属性 $E\#$ (员工号)和 $Salary$ (薪水)及时态类型 $Month$ (月),若该时态模式满足 TFD: $E\# \rightarrow_{Month} Salary$,则表明每个员工在一个月内有唯一的薪水。基于 TFD,Wang 对多时间粒度下的时态数据库的逻辑设计问题进行了系统研究,系统地提出了时态模式的规范化理论,并定义了时态第三范式 T3NF 和时态 BC 范式 TBCNF,并给出了相应的时态模式的分解算法。

(6)Liu 于 2000 年提出了一种称为稳定性约束的动态依赖的概念,它反映了属性间的变化快慢关系。对于稳定约束 $A \nabla B$,表明不存在这种情况,属性 B 的值变化而属性 A 的值不变,即属性 A 的值要比属性 B 的值更稳定。

从前面的叙述可以看到,Wijzen 的依赖理论已经涵盖了 Vinau 和 Jensen 的依赖概念,而 Liu 的稳定性约束除不支持多时间粒度外,对如何利用其进行时态数据库的逻辑设计还缺乏系统的理论。

下面着重分析一下 Wang 的 TFD 和范式理论在时态数据库的逻辑设计方面与 Wijzen,Navathe 和 Ahmed 所提出的依赖和范式理论相比有哪些优点:① Wijzen 及 Navathe 和 Ahmed 所提出的依赖和范式理论不支持多时间粒度;② Wijzen 及 Navathe 和 Ahmed 的逻辑设计策略中没有考虑依赖保持的问题;③ Wijzen 及 Navathe 和 Ahmed 所提出的各种依赖都可以用 Wang 的 TFD 来表示。由于 Wang 的 TFD 能够很好地捕捉时间变化信息并支持多

时间粒度,其已成为在多时间粒度下时态数据库逻辑设计的有效工具。与其他时态数据依赖相比,Wang 的 TFD 具有更丰富的语义,可以在更大程度上消除数据冗余并具有最完备的规范化理论。鉴于此,Wijssen 于 1999 年对 TFD 在支持对象标识等方面进行了进一步的扩充,使其具有更丰富的约束表达能力,但在数据库的设计方面没有进一步的研究,仍采用 Wang 提出的逻辑设计方法。另外,Wang 的 TFD 是基于时态模块提出来的。时态模块模式可以按照其他的时态数据模型直接地进行转换。因此,只要得到规范化的时态模块模式,经过转换就可以直接得到满足实际数据模型所需要的数据库模式。时态模块为访问不同的时间信息系统提供了一个统一的接口。

多粒度时态数据库中由于多粒度时间维的引入,使得其规范化设计极其复杂。如何为多粒度时态数据库设计行之有效的规范化方法,是近年来时态数据库理论研究人员致力研究的方向。在时态数据库的规范化研究过程及时态数据库设计的研究领域中,已经提出的一些时态数据库的范式有 Segev 的 1TNF、Navathe 的 TNF、Lorentzos 的 P 和 Q 范式、Jensen 的 T3NF 和 TBCNF、Wang 的 T3NF 和 TBCNF。其中,Segev 的 1TNF、Navathe 的 TNF、Lorentzos 的 P 和 Q 范式都是针对特定的时态数据模式,而且它们都偏离了传统的关系数据库范式,不是传统范式真正意义上的扩展;Jensen 的 T3NF 和 TBCNF 是对传统范式的扩展,但只考虑了时态关系快照中的数据冗余,没有考虑快照间的数据冗余,而且没有考虑多时间粒度的问题;Wang 的 T3NF 和 TBCNF 是对 Jensen 工作的扩展,考虑了快照间的数据冗余,而且考虑了多时间粒度的问题。但是,由于多时间粒度的引入以及时态类型间的复杂操作,使得讨论更加增加了难度。虽然 Wang 基于所提出的 TFD,分别提出了具有多时间粒度的时态数据库逻辑设计问题以及满足 T3NF 和时态 TBCNF 的时态模式的分解算法。其中,对于任何时态模式,一定可以得到一个满足无损连接、保持依赖的 T3NF 的时态数据库模式。但 Wang 的算法是指数级的,在实际的数据库设计中很难使用,这主要是因为两方面的原因:① 由于时态类型集所具有的偏序特性使得算法中的一些时态类型间的操作(例如 *Raise* 操作)在计算机中无法实现,并且一些时态类型间的操作所产生的新的时态类型很难用概念进行描述;② Wang 的算法是以时态范式判定为基础的,而对于时态范式判定至今还没有在任何文献中发现有效的算法。为克服上述问题,本书通过找到一个分解算法,使该算法避开无法实现的时态类型间的操作并且能够绕开范式判定。这将会涉及 TFD 集化简、寻找时态码和成员籍等算法。至今,还没有在任何文献中发现这些算法被系统研究。本书

建立了一个对于 *glb* 操作封闭的时态类型集;讨论了时态函数依赖集的成员籍算法;对具有全序时态类型集的时态函数依赖集进行了深入研究。在本书讨论的基础上可以对时态数据库进行有效的逻辑设计。

另外,类似于传统的多值依赖 MVD,从知识结构和依赖约束的角度应当可以定义多时间粒度下的时态多值依赖的概念,并对其语义、推导公理及在时态数据库逻辑设计中的应用等进行探索。

一个好的数据库逻辑设计目标是消除数据冗余以及插入、删除和更新异常,而要设计一个好的数据库必须讨论数据组织的理论问题,其中重要的部分之一就是数据依赖、规范化理论。

1.2 时态 ER 模型的研究概况

ER(实体-联系)模型可以为不同但相关的目的所使用,例如,作为分析工具去模拟现实世界;作为设计工具,可以描述计算机系统的数据库模式。目前 ER 模型已经开发出多个不同的版本。基于 ER 模型的关系数据库逻辑设计方法学已经为广大用户所熟知。

近年来,随着与时间相关的数据库应用系统需求的增多,对 ER 模型进行时态扩展,使其能够准确地捕捉时间变化信息,已经成为研究热点之一。经过二十多年的研究,目前已出现了多个 ER 模型的时态扩展版本和十余种时态 ER 模型。在这些时态 ER 模型中,描述最好的是 TERM 和 TERC+模型。作为时态数据库系统的分析和设计工具,现有的时态 ER 模型还存在着一些不足之处。下面以 TERM、MOTAR 和 TERC+模型为代表加以说明。在支持多时间粒度方面,尽管上述三种模型在概念级有支持多粒度的可能,但 TERM 只在原子历史级有不同的时间域,无法表达用户定义的时间粒度;MOTAR 只规范了周期属性记录的频率,对其他时间变化属性没有准确的约束;而 TERC+模型根本没有说明如何支持多时间粒度;在时态约束方面,尽管 TERM 和 TERC+模型支持时态约束,但无法灵活表达用户定义的时间变化属性之间固有的依赖约束。

由于缺乏对时态数据依赖的刻画能力,目前存在的时态 ER 模型无法和时态规范化理论相结合,因此没有形成像基于 ER 模型的关系数据库设计方法学那样完整的时态数据库设计方法。通过前面的讨论,利用时态函数依赖 TFD 约束可以对时态数据库进行有效的规范化。因此,希望一种时态 ER 模型能够规范 TFD 约束,并且能够转换成时态模块模式。目前还没有一个时态

ER 模型能够做到这一点。

针对上述的问题,本书对现有的时态 ER 模型进行扩充,其应当遵循以下原则:

(1)模型应该在更大程度上支持多时间粒度,不仅能够规范时间变化属性的时间粒度,而且还要能够规范联系中实体的相对变化的时间粒度。

(2)为了与基于 TFD 的时态规范化理论相结合,模型应当能够规范 TFD 约束。

(3)模型应当能够被合理、有效地转换成时态模块模式,以便于向其他数据模型转换。

(4)扩展后的时态 ER 模型还应该能够支持非时态信息系统的分析和设计。

这些需要对现有的 ER 和时态 ER 模型进行深入的分析,从而开发出能够同时有效地用于时态数据库和传统数据库设计的新的时态 ER 模型。

1.3 时态数据库的类型和查询语言概述

1.3.1 时态数据库元素的类型

本节将介绍时态数据库所涉及的时间、时态数据、时间粒度及和时间相关的四大类数据库类型。

(1)时态数据库涉及的三种时间类型。

① 有效时间(valid time):是指一个对象在现实的世界中发生并保持的时间,或者该对象在现实世界中为真的时间段。

② 事务时间(transaction time):是指一个对象录入数据库系统的时间,有时候也称它为系统时间(system time)。

③ 用户定义的时间(user-defined time):是指用户根据需要而输入的时间,数据库管理系统将它与数据库中其他一般数据等同对待。

例如一个员工关系(名字,级别,提升时间),元组(张辉,副教授,1997年1月1日)于1997年1月27日录入数据库,则这里提升时间“1997年1月1日”是用户定义的时间类型,它表示的是提升管理委员会同意提升的时间,有效时间就是提升生效的时间,在这个例子中(1997年1月1日,下一次提升的时间)就是有效时间,而事务时间就是将提升的信息存储至数据库的时间,即是“1997年1月27日”。

(2)时态数据库涉及的三种时态数据类型。

① 瞬时(instant):是时间轴上的固定点。例如,王三转正定级的时间是1999年7月1日。这种类型是时态领域中最基本的类型,其他类型可以通过瞬时来实现或者在某种程度上通过瞬时来模拟。除了瞬时类型之外,大多数的数据库管理系统并没有提供其他类型的时态数据类型。

② 时区(interval):是时间轴上一段时间的长度,是时间轴上不固定的但连续的部分,即两个瞬时之间的距离。时区的概念是相对的,且具有方向性,正区间表示未来,负区间表示过去。

③ 期间(period):是时间轴上一段固定的时间区间,例如,1999年7月至2000年7月就是一个期间的例子。在商业上广泛应用的DBMS和SQL92标准中都不支持这种数据类型。但是我们可以通过它们支持的瞬时数据类型来模拟期间类型。方法是用一对瞬时数据类型(t_1 , t_2)来表示,前者 t_1 表示期间的开始,后者 t_2 表示期间的结束。

(3)时间粒度:计算机对时间的存储不可能是连续的,必须以离散的形式来存储。而时间粒度是对计算机以离散的形式存储的离散化的度量,当时间粒度确定之后,粒度越小表明精确度越高,但是越小的粒度就会越增加内存存储容量。这就使用户在实际设计系统时必须根据不同应用的需求而定,当选择不同的粒度时,由于不同的粒度具有可转化的功能,所以按照需要可以进行转化,以减轻内存的负担。

时态数据库涉及如下两种时间粒度:

① 单粒度。时态数据库系统只支持单一的时间粒度。对于不同的数据库系统可能存在不同的时间粒度,但是对于一个数据库系统只认可一种时间粒度。

② 多粒度。时态数据库系统中支持两种或两种以上的时间粒度,对于不同的属性可以存在不同的时间粒度。将在后面的讨论中用到。

(4)概括地说和时间相关的数据库有下面四种类型。

① 快照数据库(snapshot database):在某一特定时间的瞬间快照来建立二维数据模型,只反映某一瞬时情况,它只支持第三种时间,即用户定义的时间。它不区分第一种时间和第二种时间,即它不区分有效时间和事务时间。由于快照数据库主要使用数据更新操作,如删除、替换、插入等对一个数据库的状态进行一次系统活动,它导致数据库的过去状态丢失并忘却,它仅仅反映现实中的一个片断。这种数据库不能称为时态数据库。

② 回滚数据库(rollback database):它支持第二种时间即事务时间来建立三维数据模型。按照事务时间进行处理,根据事务时间在系统中保持数据