# 第 1 章　　绪 论

时间是自然界无所不在的客观属性，所有信息都具有相应的时态属性。随着数据库与信息技术的深入和发展，信息系统面临许多新的应用和新的需求，对时态信息处理的需求越来越迫切。时态信息处理已成为许多新一代数据库与信息系统的关键技术，特别是在电子政务、电子商务、数据仓库、数据挖掘、决策支持系统等信息系统中扮演着日益重要的角色，在某些系统中，信息的时态性还起着关键性的作用。

自 20 世纪 80 年代初时态数据库技术产生以来，经过大量计算机工作者的研究和实践，时态数据库在基础理论、数据模型、数据库语言、应用技术等方面都取得了丰盛的成果，已经成为数据库与信息系统一个重要和充满活力的学科方向。本章首先简要讨论时态应用的实际需求，然后介绍时态数据库技术的产生与发展，讨论时态数据库研究和应用现状及主要不足，最后介绍本书的目的和组织。

## 1.1　时态应用的实际需求

时态数据库应用不断渗透到数据库和管理信息系统等许多应用领域，本节以电子政务、工作流技术、数据仓库与数据挖掘领域为例，说明时态信息处理的应用需求。

### 1.1.1　电子政务中的时态应用

在电子政务中，政府提供和保存着大量的信息，这些信息的有效处理显得特别重要。例如，政府发布的政府招标公告，规定了一个进行竞标的时间区间，超过时间的标书视为无效；政府制订的政策法规在不断的完善中，更新也比较快，这也要求电子政务系统可以自动地对失效的信息进行相应的处理。

但是，目前多数电子政务的信息处理技术沿用电子商务和传统应用的信息处理方式，我们经常可以看到政府网站上的过期信息，这些过期的信息不但给公众造成了很大的困扰，而且也影响了政府职能的有效行使。出现这种情况的原因就是在电子政务系统中没有很好地处理政务信息的有效性。

需要全本请访问 www.ertongbook.com

例如，表 1.1 是一个招标公告信息的传统模式，包括两类信息，一类是信息内容，一类是控制信息（这里发布时间就是控制信息）。招标公告的竞标是有时间期限的，显然，这样的一种信息表示方式不能体现信息的有效性。但是，如果对其进行改进，添加时态区间属性和状态属性，如表 1.2 所示，那么，这既描述了招标公告的时间点时态属性（发布时间），又描述了其时间区间属性（有效时间）。这就需要在系统中支持时态信息处理。设计支持时态信息处理的时态电子政务(Temporal Electron Government ）系统已成为新一代电子政务的崭新课题和重要方向。

表 1.1 招标公告的传统信息模式

| 编　号 | 类　型 | 发布单位 | 发布时间($T_0$) |
|---|---|---|---|
| 1 | 计算机设备招标 | 东山区政府 | 2002-12-12 |

表 1.2 招标公告的时态信息模式

| 编　号 | 类　型 | 发布单位 | 发布时间 | 有效时间 | 状　态($S_i$) |
|---|---|---|---|---|---|
| 1 | 计算机设备招标 | 东山区政府 | 2002-12-12 | 2002-12-12～2002-12-30 | 0 |

本书第 9 章将给出应用实例：工资时态知识数据库系统，也就是一个典型的时态电子政务应用。在该系统中，不仅数据有时态性，而且知识也有时态性。例如，一个人员的工资需要根据他的学历、职务及其获得时间等信息（时态性数据），依据相应时期的工资政策（时态性知识)来确定。

### 1.1.2 工作流的时态应用

工作流(Workflow）是工作任务在多个人或单位之间的流转。在计算机网络环境下，这种流转实际上将表现为信息或数据在多个人之间的传送。按照工作流管理联盟（Workflow Mangerment Coalition，WfMC）的定义，工作流指的是“业务过程的部分或全部在计算机应用环境下的自动化”，在此过程当中，文档、信息或任务按照一定的过程规则流转，实现组织成员之间的协调工作，以期达到业务的整体目标。可以简要地说：工作流包括定义、监测、控制、优化，以及相关的逻辑支持的复杂过程，是一个过程、事件、资源的有机结合的三维概念。

工作流中的时态属性主要包括结构性和整体性的时态属性和局部性工作项的时态属性两种。工作流作为一个过程流，其运行过程与时间有密切的关系。目前，将工作流作为整体机构的时态关系的研究已经取得一定的成果，例如，用 TPN(Time Petri Net)对于工作流的时序关系进行研究，可以分析整个工作流的最早完成时间，每个工作项的最早、最迟开始时间，有效时间，截止时间等。

局部性工作项的时态属性是指：在工作流的实际运行中，对于每一个工作流的运行实例，其中运行的数据都是不同的，工作流模型要对这些具体的运行实例进行有效的控制，就要求包含对具体的工作流中特有的时态属性进行处理。

例如，在一个工厂建设项目环保局审批的工作流模型中，可以定义每一项工作的时间特性,这里,假定环保局的审批工作必须在 60 个工作日内完成。但是，在具体的运行中，可能会有这样的情况，经过预审批出来的审批意见是有时间特性的，这就是说预审批意见是有时间限制的,假设其为 30 天。当工作流经过了预审批后到达环保局审批时，如果环保局在 30 天内不能对这个申请进行审批，那么预审批的意见就自动失效，申请被驳回；如果在 30 天内审批通过，那么审批流程可以继续进行。

在上面的工作流例子中，就既存在结构性的时态属性：环保局审批工作的完成期限（60 个工作日）；也有工作项的时态属性：预审批意见的有效期限( 30 天 )。

事实上，工作流中还存在其他多种形式时态属性，时态工作流（ Temporal Workflow ）是目前新一代工作流技术的研究热点。

### 1.1.3 数据仓库中的时态应用

随着数据库技术，特别是数据仓库的发展和应用，出现了所谓“数据丰富，知识贫乏”的问题；人们不再满足对数据库进行查询、检索等简单的操作，而是希望从中发现知识，以辅助高层人员的决策（即数据挖掘）。数据挖掘的任务是从大型数据库或数据仓库中提取人们感兴趣的知识,这些知识是非平凡的、隐含的、事先未知的、潜在有用的信息,通常是以概念、规则、规律、模式、约束、可视化等形式来表示。

在数据仓库中通常存放有大量的时态数据，例如，超市交易记录中的交易时间，病历中的检查和诊断时间等。时态数据反映了事物发生发展的过程，有助于揭示事物发展的本质规律。在数据挖掘中结合时态数据的特点可以挖掘出更有用的信息。例如，在一个电子商务系统中，由于顾客一般在 7：00AM～9：00AM 买早餐，如果我们对这个时间的数据进行挖掘,就会发现规则 牛奶→咖啡；但是，如果对一天的数据进行挖掘，就可能不能发现牛奶→咖啡这条规则。因为可能除了 7：00AM～9：00AM 这个时间段，很少有人买牛奶就买咖啡，因此不能达到支持度或／和置信度，从而不能发现该规则，也就失去了一条有用的规则。

所以，在数据挖掘中必须考虑时间的问题，选取有效时间进行挖掘至关重要。时态在数据挖掘应用中主要集中在规则的有效性问题上，另外，还有一个重要应用是时序数据挖掘，例如证券的走势等。

## 1.2 时态数据库的起源与发展

时态信息的需求与技术实际上一直伴随着数据库技术的发展而产生和发展。关于时态数据库的起源与发展，国际时态数据库权威学者 A. Tansel、J. Clifford、S. Jajodia、A. Segev 和 R. T. Snodgrass 等在 1993 年共同主编的《Temporal Databases: Theory, Design, and Implementation》中收集了此前时态数据库几乎所有的重要成果，我国学者唐常杰 1999 年在《时态数据库的沿革、特色与代表人物——时态数据库 20 年回顾之一》等文献中也对时态数据库技术前 20 年的发展做了系统回顾。但是，到 21 世纪的今天，时态信息处理的研究与应用出现了许多新的进展。我们通过多年的研究，参阅了大量文献，特别是国内外最新研究动态和应用成果，将时态信息技术的起源与发展分为 3 个时期：开创期(1982 年以前)、理论与模型探索阶段(1983～1993 年)、应用与发展阶段(1994 年以来)。

### 1.2.1 开创与形成时期

时态特性是信息的客观存在，早在 70 年代就有人关注到时态信息的应用。例如，1970 年，G. Wiederhold 和 J. F. Fries 研制的医疗系统在处理时态信息方面做了较早的尝试；1977 年，Kahn Ketal 在《Artificial Intelligence》上发表的《Mechanizing Temporal Knowledge》反映了早期关于时态信息的基础性研究工作。到 20 世纪 70 年代末、80 年代初，数据库技术的日渐成熟和大容量高速存储设备的发展为时态据库技术的产生与应用创造了条件。加州大学洛杉矶分校 J. Ben Zvi 和纽约大学的 J. Clifford 在 1982 年的博士论文是时态数据库技术开创期的两个标志性成果。

J. Ben Zvi 在 1979～1982 年间对时态数据库技术进行了系统的研究，在 1982 年完成博士论文《The Time Relational Model》。其主要贡献包括：提出了时态数据库模型，以时间区间作字段值，提出了非 1NF(N1NF)的 TDB(Temporal Data Base，时态数据库)突破了数据库字段值只能是一个数或串的局限，引入了后来被称为双时态（Bitemporal)的概念，即用有效时间(Valid Time) 表示被管理对象在库中的生命周期，用事务时间(Transaction Time)表示数据库本身的历史，引入了时态索引结构。这些概念和思想一直沿用至今，标志着时态数据库的形成。

同样是在 1982 年，纽约大学的 J.Clifford 完成了他的博士论文《A Logical Frame Work for the Temporal Semantics and Natural Language Querying of Historical Database》，并发表了一系列论文。他的主要贡献是对历史数据库做了开创性的工作，提出了 HRDM(His-

torical Relational Data Model)模型。他研究了在关系、元组、字段值上加时态信息的技术细节，引入了历史关系模型、历史关系代数，将传统关系数据库当作历史数据库的一种特例，即当把区间缩小为一点[Now, Now]时的历史数据库。此后的 20 多年来，J. Clifford 博士一直活跃在时态数据库前沿。可以说时态数据库是在 1982 年正式形成的。

### 1.2.2 理论与模型探索阶段

1982 年以后，TDB 的研究开始走向繁荣时期。在此后的 10 年中，计算机学术界对时态数据库给予了极大的关注，发表了数以千计的论文(例如，以 EI CompendexWeb 进行检索，1982～1994 年间包含“Temporal Database(时态数据库)”和“Temporal Information (时态信息)”的论文有 2000 多篇)；学术界提出了数以百计的时态数据库模型和时态信息处理方法。此阶段，国际上一些重要大学和研究机构涌现出一大批研究学者，形成了一批专门的时态数据库研究集体。最主要的研究集体集中在美国纽约大学、依阿华州立大学、加州大学伯克利分校、南加利福尼亚大学、亚利桑那大学、惠普公司和 Bell 公司的专业实验室等。

这个时期标志性成果是 A. Tansel、J. Clifford、S. Gadia、S. Jajodia、A. Segev 和 R. T. Snodgrass 在 1993 年共同编辑出版的《Temporal Databases: Theory, Design, and Implementation》，该书被称为“世界第 1 本关于时态数据库专著”。该书对此前国际时态数据库技术的研究给出了全面的总结。书中列出 20 年来 TDB文献 500 余篇，文献作者遍及美国、欧洲、中国、日本、印度、瑞士、土耳其等。收录了国际时态数据库方面重要学者的研究成果和时态数据库模型。书中列出的 13 种最有影响的时态数据库模型，除了前面提到 J. Ben Zvi 和 J. Clifford 的模型外，其他 11 种如下：

- TempSQL, Sharshi. K. Gadia&S. Nair, 1985。
- IXRM(Interval-extended Relational Model ), Nikos A. Lorentzos, 1987。
- TRM(Temporal Relational Model)& TSQL, S. B. Navathe, 1987。
- HSQL(Historical Query Language), N. L. Sarda, 1987。
- TQuel, R. T. Snodgrass, 1985。
- TRC(Temporal Relational Calculus), Abdullah Tansel, 1992。
- TEER(Temporal Query Language for Enhanced Entity Relationship Model), R. Elmasri, 1985。
- TDM (Temporal Data Model Based on Time Sequeuce), Arie Segev & Arie Shosham, 1988。
- OODAplex (Object Oriented Aplex), U. Dayal, 1989。
- Object History, S. Ginsburg, Tanaka, C. J. Tang, 1983。

需要全本请访问 www.ertongbook.com

- Temporal Deductive Databases，Marianne Bandinet 等，1989。

这段时期的重要特征是时态数据库理论的研究和时态数据模型的探讨，因而模型多，但原型系统少，实际应用少。

### 1.2.3 应用与发展阶段

经过 20 世纪 80 年代末的“百花齐放”，1993 年 A. Tansel、J. Clifford 等人的《Temporal Databases：Theory，Design，and Implementation》可以说是划上了一个“分号”。此后，学术界的观点和认识逐步趋向“统一”，从早期的基于关系代数，到后来的 Calculus-based、Datalog-based、OO 等，现在基本上都采用扩充 SQL 模型，时态模型没有新的突破。1994 年后人们开始进入“思索”阶段，如何将时态数据模型“标准化”和“产品化”是这个时期的重要目标，而广泛应用和实践则是这个时期的重要特征。

在“标准化”方面，代表性成果是 R. T. Snodgrass 等人在 1994 年提出的双时态数据模型 TSQL2，TSQL2 是对 SQL-92 语言标准的时态进行扩充，并提出将 TSQL2 的相关结构集成到 SQL3 标准。1996 年已将增加有效时间和事务时间的 SQL/Temporal 方案分别于 5 月和 10 月提交到 ISO 和 ANSI 委员会。但是，时态数据模型标准化的道路还需要探索。美国 Christian S. Jensen、J. Clifford、R. T. Snodgrass 等于 2000 年合作的《Temporal Database Management》是这阶段的重要文献。

在“产品化”方面的代表性成果是瑞士 TimeConsult 公司于 1998 年推出的 TimeDB 2.0，目前已发布 Beta 4。TimeDB 是一个双时态关系数据库系统，支持平台 Java、JDBC。它基于 SQL 查询语言，支持时态查询语言 ATSQL2(Applied TSQL2)。ATSQL2 基本模型源于 R. T. Snodgrass 的 TSQL2 语言，并结合丹麦 Aalborg 大学的 Michael Boehlen 提出的 ChronoLog 模型和瑞士 Andreas Steiner 的 Bitemporal ChronoSQL 模型。TimeDB 作为商业数据库 Oracle 系统的前端运行。ATSQL2 的语句(包括查询、更新和断言等)被编译成标准的 SQL-92 语句在后台执行。这种方法确保了在不同的平台和不同的数据库管理系统中的兼容性。但是目前 TimeDB 还在完善之中，仍然没有能够真正“产品化”。

这个时期的另一个重要特征是“时态信息的应用”。 20 世纪 80 年代中后期，由于计算机相关技术的迅速发展，特别是网络与多媒体等技术的发展，使得应用领域的许多新需求能够实现。时态信息的应用就是一个重要方面，其主要应用领域包括：地理信息系统、农业信息系统、电信信息系统、电子政务、电子商务、智能决策支持系统，以及数据仓库与数据挖掘等，特别是时空(Spatio-temporal) 信息技术和多媒体信息技术的时态应用。但是，由于时态数据库理论与模型还不够成熟，时态信息产品还没有形成，所以这时期的应用大多都是只借助时态数据库的一些概念，时态信息的管理与操作的实现还是采用传统的数据库技术与相

关应用领域的技术（GIS、Multimedia 等）相结合来完成的。

## 1.3 时态数据库的研究现状

### 1.3.1 研究文献分析

国内时态数据库的研究和应用也越来越受到关注，关于时态信息处理技术的文献越来越多，相比 20 世纪 80 年代而言，20 世纪 90 年代以 10 多倍的速度增长。我们主要以《中国学术期刊网》、EI CompendexWeb 以及相关网站为例对相关论文做了分析。

论文以《中国学术期刊网》为例，查询得到国内 1994 年以来，以‘时态数据库’为关键词的论文 21 篇，全文中有‘时态数据库’的论文有 30 篇；关键词包含“时态信息”的论文有 7 篇，全文中包含‘时态信息’的论文有 65 篇。从论文内容分析，关于时态数据库技术的一般介绍性论文约占 30%，多媒体中时态数据技术应用约占 15%，在地理信息系统、农业信息系统、电信信息系统等专业信息系统中的涉及时态信息处理的论文约 15%，其他主要是时态数据理论与模型的阶段性研究成果，也有 4 篇数据挖掘中的时态信息应用，关于时态知识的论文主要是对 13 种时态逻辑关系的扩充研究，还没有包含“时态知识数据模型”的内容。由此可见，国内专门研究时态信息模型与理论的作者和单位不多。

国际论文以 EI CompendexWeb 进行检索，在 20 世纪 80 年代，Abstract 中包含‘Temporal Database’，的论文有 64 篇，包含‘Temporal Information’的论文有 432 篇；20 世纪 90 年代至今，Abstract 中包含“Temporal Database”的论文有 848 篇，包含‘Temporal Database’的论文 3671 篇。从 2000 年以来的论文分析，一半左右论文的主要内容是关于时态数据模型和理论方面的，而时态数据的应用方面则主要是关于时空数据库和多媒体时态信息，此外，关于数据仓库中的时态信息和时态数据挖掘（Temporal Data Mining）论文明显呈增长趋势。

而且，许多网站介绍时态数据库研究文献和研究学者信息，例如，美国 George Mason 大学的 S. Jajodia 等建立的‘Temporal Database Bibliography Update”网站收集了 337 篇时态数据库论文，包括模型（Model）、数据设计（Database Design）、查询语言（Query Language）、约束（Constraint）、时间粒度（Time Granularitiy）、实现、存取方法、实时数据库、序列数据库、数据挖掘（Data Mining）、并发控制及其他共 12 类。

这些资料反映了近期时态数据库的一些主要研究方向和成果。一些重要的时态数据库论文主要集中在一些重要的研究集体和研究学者，例如时态数据库研究机构 TimeCenter

云集了美国、欧洲、澳洲、新加坡等国家的大学和研究机构的众多研究者。

### 1.3.2 主要时态数据模型

时态数据库理论研究已经走过了近 30个年头，取得了丰硕的成果。时态数据库模型很多,各有千秋。下面是几个“代表性”的时态数据模型(Temporal Data Model)的基本情况，从中可以看出这些模型在概念和机制上有许多共同的特点。

1. TempSQL

基本结构:SQL 的扩充。
支持时间类型：有效时间。
提供操作:关系表达式,时态表达式,布尔表达式。
典型查询(其中加粗字体表示关键字,下同)：

```
Select selectList
While whileExpression
From fromList
Where whereCondition
Group By groupByList
Having havingCondition
During duringExpress
```

作者：美国依阿华州立大学计算机科学系 Shashi K. Gadia 和 S. Nair。
最初发表时间:1985 年。

2. TRM & TSQL

基本结构:TRM 的一种实现。
支持时间类型：有效时间。
操作:传统 SQL 表达式,时态表达式,特别表达式(TJOIN、TNJOIN、COMPRESS)。
典型查询：

```
Select selectList(TIME-START、TIME-END、integer)
When whenExpression(Before、After、During、Equivalent、Adjacent、Overlap、
Follows、Precedes)
From fromList
```

```
Where whereCondıtıon(Fırst、Second、Third、Last、Break)
Group By groupByLıst
Time-Slice timeSliceExpression
Moving Windows movıngWindowExpression
```

**作者：美国佐治亚理工学院计算机学院** Shamkant B. Navathe **和美国加州惠普实验室** Rafı Ahmed。

**最初发表时间：**1987 **年。**

3. HSQL

**基本结构：历史数据库管理系统** HDBMS **的一种实现。**

**支持时间类型：有效时间。**

**操作：标准操作为** PJ(Project)、SL(Select)、PR(Cartesian Project) **新增原语操作为** EX(Expand)、CL(Coalesce) .

**典型查询：**

```
From Time...To Time TimeLimite
Select selectList(COALESCED)
From fromLıst(CONCURRENT)
Where whereCondition
Coalesced On vısıble-attrıbute-list
Expane By expandByList
Group By groupByLıst
Having havingCondition
```

**作者：印度理工学院计算机科学与工程系** N. L. Sarda。

**最初发表时间：**1987 **年。**

4. TQuel

**基本结构：**Quel **的最小扩展。**

**支持时间类型：有效时间和事务时间。**

**操作：普通** Quel **操作(静态数据操作),有效时间操作,事务时间操作。**

**典型查询：**

```
range of tl is Rl
```

```
.....
range of tk is Rk
retrieve (til.Dj1,…,tir.Djr)
valid during v
where
when
as of
```

作者：亚利桑那大学计算机科学系 R. T. Snodgrass。
最初发表时间：1985 年。

5. TSQL2

基本结构：SQL-92 语言标准的时态扩展。
支持时间类型：事务时间和有效时间。
语言规范：预备语言规范见 March 1994 ACM SIGMOD Record，最后规范见 March 1994 ACM SIGMOD Record，语言指南发表在 September 1994 ACM SIGMOD Record。
作者：亚利桑那大学计算机科学系 R. T. Snodgrass。
最初发表时间：1994 年。

6. OODAPLEX

基本结构 基于 DAPLEX 函数数据模型的面向对象数据模型。
支持时间类型 有效时间(通过对象的属性表达)。
操作：所有的操作都是函数，操作可以嵌套。
典型查询：

```
for (data) in (table) where (limit)
    (collected data limit)
end
```

作者 美国 Bell 通信研究所 Gene T. J. Wuu。
最初发表时间：1989 年。

### 1.3.3 研究与应用现状

目前，时态信息技术仍处于研究和发展阶段，人们从不同的观点，提出了各种时态数据

库模型。另外，由于实际应用的需求，时态信息处理的应用领域越来越宽，在应用中也提出了许多方法和技术。这些研究与应用都大大促进了 TDB 的发展。现有的时态数据库的共性概念、研究方法及不足可以归纳如下。

(1)在现有的时态数据模型中，对时间数据的描述主要是依照 J. Ben Zvi 于 1979～1982 年提出的时间点／时间区间模型、有效时间和事务时间，以及双时态(Bitemporal)等概念，对态数据演算主要是基于 J. F. Allen 于 1984 年提出的 13 种时间区间演算（或其扩展）。主要不足是：时态数据运算体系不完备，时态关系演算还没有系统和有力的数学理论支持。

(2)时态数据模型多，一些标准在逐步提出申请，例如 ATSQL2 提交的 SQL/Temporal 标准。由于时态数据模型还不够成熟，所以还没有形成较完整的国际标准。各种模型都还存在一些不足。时态数据模型一般都是传统关系数据库的扩展，并将传统关系数据库作为特例。其主要方法是在 TDB 中增加一些运算,例如 After、Before 和 Overlap 等扩展一些操作，例如时态选择、时态投影、时态连接等。目前，大部分时态数据查询语言只是扩展当前的查询语言,如 SQL 或者 Quel，时态数据查询功能有限，效率比较低。由于种种原因，数据库厂商难以下决心选用 TDB 技术用于产品。

(3)目前，时态数据库研究取得了相当的进展，但是大多数研究局限在数据库的时态属性，而忽视了其他信息的时态属性，例如知识库的时态特征和应用。时态数据技术目前还停留在“数据”处理上，关于时态逻辑和推理方面的研究，主要优点是符号演算和推理能力强，但是信息处理能力弱，与时态数据库和时态信息处理研究相脱离。关于时态知识与逻辑方面主要包括时间区间逻辑运算的扩充等，没有涉及时态知识数据库模型。

(4)在应用方面,由于 20 世纪 80 年代数据库技术的迅速发展，特别是 20 世纪 90 年代多媒体技术、网络技术等的发展，所以时态信息的应用呈现出勃勃生机。在地理信息系统、农业信息系统、电信信息系统、电子政务、电子商务、智能决策支持系统、数据仓库与数据挖掘,特别是时空(Spatio-temporal)信息技术和多媒体信息系统方面，时态信息处理技术得到空间的重视和应用。但是，由于还没有成熟的时态模型和软件产品，大部分与时态相关的应用，只是借鉴当前一些时态数据模型，在实际实现中仍然只能采用传统的技术，其时态部分的解释是由应用程序而非数据库本身来进行的。

### 1.3.4 展望

新一代信息系统对时态信息处理应用技术迫切、广泛的需求，具有广阔的市场发展前景。时态信息处理理论、技术和应用的研究和应用方兴未艾。

时态数据模型朝着统一化、标准化方向发展，时态数据库查询语言朝着“产品化”方向发展。由于 SQL 是当前数据库最权威的查询语言标准，所以时态数据模型从早期的基于关系

需要全本请访问 www.ertongbook.com

代数,到后来的 Calculus-based、Datalog-based、OO 等，现在基本上都采用扩充 SQL 模型。例如 TSQL2 模型、ATSQL2 查询语言,以及提交 ISO、ANSI 的标准提议 SQL/Temporal 等都是 SQL 的扩充。Tqule 的作者 R. T. Snodgrass 也将主要研究转向基于 SQL 的时态数据模型,并提出 TSQL2 数据模型。但是，时态数据的“统一模型”工作仍然任重道远，时态信息产品化工作还大有作为。

另外，时态信息应用领域越来越广阔，时态信息需求多元化，时态信息的应用也多元化。从传统的时态数据处理到时态知识的应用，从传统的文本数据库到时空数据库、多媒体数据库系统，应用领域也渗入到科学实验信息系统、多媒体信息系统、地理信息系统、市政信息系统、电信信息系统、电子政务、电子商务、数据仓库等各个方面。时态信息需求与处理越来越复杂,完全‘统一’的时态数据模型难以实现,仍然会有多种领域范围内‘统一’的时态数据模型。此外，由于时态只是信息的重要属性，时态信息产品将朝着“嵌入式”、基于主流数据库技术与平台、面向领域的、面向应用的中间件和软件构件等方向发展。

# 第 2 章　时态基本元素模型

时间是什么？它有起点和终点吗？理论物理学家在宇宙的大爆炸的假说中指出时间开始于大爆炸的那一瞬间；至于是否有终点，他们并没有给出答案。时间的概念就像一个被压缩的橡皮筋，随着宇宙的爆炸在不断地扩展。本书是从数据库角度研究时间信息的表达、存储、管理等处理技术和应用。传统数据库在添加了时间属性后，相应的概念、性质、基本原理、基本方法都发生了变化。时间元素在时态数据库中是记录和标识元组时间属性的基本原子，因为，没有时间元素的有效准确标识，时态信息的记录是不清晰的，所以，时态元素模型是时态数据库基础之基础。

本章研究数据库中有关时间信息的基本表示方法和模型，介绍时间点、时间区间、时间跨度和时间域等时间模型；介绍时间量子、时间粒度等基本概念；讨论 Allen 的时态区间关系等；最后介绍两个特殊的时态元素 Now 和 Until Changed。

## 2.1　时态信息模型

### 2.1.1　时态信息

在现实世界中，时间无时不有，无处不在，客观世界中的事物都带有时间的属性。描述现实世界的带有时间属性的信息系统，称为时态信息系统。例如金融方面的会计、银行系统；档案管理方面的人事档案、医疗记录系统；行程安排中的飞机、火车、旅店等项目的管理系统；科学应用方面的气候、地理系统等。这些系统中记录的信息是随时间变化的，这种随着时间变化的信息称为时态信息（Temporal Information）。

时态信息需要用数据进行记录，这些用于记录时态信息的数据就是时态数据（Temporal Data）。时态数据在计算机系统中一般保存在数据库系统中。记录和管理时态信息的数据库，有别于传统的关系数据库，记录中的元组部分属性带有时变的特性，这就导致这种数据库的数据模式与传统的关系数据库数据模式差别较大，由此而产生了数据库一些基本概念、理论、性质方面的改变；这种记录时态数据反映时态信息的数据库就是时态数据库。

表 2.1 中，表示了一个单位员工的职称和工资随时间变化的情况。因为员工的工资和职称的相对稳定性，时变属性职称和工资的记录是分段的。在这里，用表格的方式记录该单位的员工工资和职称的数据就是时态数据，这些时态数据所反映出来的该单位员工的职称和工资情况就是时态信息。表中的闭区间，如[1989，1998]代表的是时间区间段，表示1989～1998年的时间区段。

表2.1 员工工资职称表

| 姓　名 | 职　称 | 工　资(元) |
|---|---|---|
| 李小山 | 讲师,[1989,1998]<br>副教授,[1999,2002] | 1000,[1989,1998]<br>1500,[1999,2002] |
| 张又宝 | 副教授,[1990,2002] | 1500,[1990,2002] |
| 王大江 | 教授,[1991,2002] | 2000,[1991,2002] |

### 2.1.2 时间模型

在自然界中，时间是每时每刻都存在、连续发生且一去不复的，它在时间轴上是连续存在的。从计算机科学的角度来看，如何量化时间和确定某个时间点或者时间段，这关系到如何确定时间模型。如果按照现实世界的本来面目来记录时间，则数据量过于庞大，而且人们往往并不需要获取现实世界所有时间的状态。解决这一问题的方法是建立合适的时间模型，使得它不仅能应用到数据模型中，而且能满足不同应用场合的要求。

基于对时间轴结构的选择，时间模型可以划分为如下几种模型。

1. 连续模型(Continuous Model)

连续模型把时间看作同构于实数，每一个实数对应于一个时间点。因此，在时间轴的两个时间点之间，可以存在其他的时间点。这种模型能够最精确地为时间建模，但是由于现代计算机基于数字逻辑的工作方式，所以不可能无失真地记录时间。在许多实时控制场合中，例如工业控制领域，需要记录大量随时间不断变化的数据，在这种情况下，往往采取采样的方式记录数据变化，对相邻时间点之间的数据采取插值的方法得到。

2. 步进模型(Stepwise Model)

步进模型把数据的状态看成是时间的函数。当时间点上的数据状态发生变化时才记录状态变化，否则保持不变。在这种模型下，时间序列上任一点上数据的值对应于上一次数据改变时保持的状态，如果要查询当前数据的取值，则需要回溯。例如，图 2.1 是采用步进模

型登记的某个教师身份变动信息。从图 2.1 中可以看到，李明的身份自 1993 年至今一共发生了 3 次变动，采取步进模型只要求记录 3 个点的信息。如果查询目前该人员的身份，则虽然没有登记当前的时间点，但并不是返回一个无效的空值（Null），而是沿着时间轴回溯，找到最近发生的状态变化，所以得到当前的身份为“副教授”。该模型与连续模型的不同之处在于，任意两个时间点之间不能通过插值的方法取值。




图 2.1　某教师的身份变动信息

3. 离散模型（Discrete Model）

离散模型把时间和整数映射起来，且在相邻的两个时间点之间不存在另一个时间点。任一时间点有前驱和后继时间点。在实际应用中，该模型适用于记录那些在关键时间点上才有意义的数据。例如，考虑某单位的人员工资发放总额，1～3 月的工资发放总额分别为 19 万元、18 万元和 20 万元。可以看到这组数据有如下特点。

(1)在相邻的两个月份之间不存在另一个月份，如 1 月和 2 月。

(2)相隔的两个时间点中间月份的数据，用插值的方法会得到无效数据，通过 1 月和 3 月的数据无法确定 2 月数据。

(3)采取步进模型描述可能得出错误信息。假如现在为 4 月 10 日，当月的工资尚未发放，如果回溯，就会得出 4 月工资发放总额为 20 万元的错误数据。

4. 恒定模型（Non Temporal Model）

有些数据是不随时间变化的，例如，籍贯、出生地等。这些数据只有其本身固有的属性。但是大部分数据在一种情况下没有时态属性，但在另一种情况下往往会有时态属性。例如，住址、身份、工作单位等。在一般情况下，在建模时通常没有充分考虑值随时间变化的情况。如果发生变化时，就采用最新值进行替换。

## 2.2　时间基本元素单位

时间元素（Time Element）就是指表示时间属性值的元素，时间元素在时态信息系统中

有着基础的地位，它对于正确有效的表达记录的时间属性有着重要的意义。对于时间元素来说，出于表达的需要，它有许多具体的形式，有的是离散的时间点，有的是连续的时间区间，当然还有较为复杂的情形——集合形式。

对于离散的时间点（Time Instant），属性的时变特性是离散的，每个时间点都对应记录中的属性值；对于时间段(Time Interval)，属性的变化具有某些连续的特性，用连续时间区间记录元组的属性节省了大量的空间；对于复杂的集合形式，时间元素的描述更为复杂。

### 2.2.1 时间点

基于点的时间是将时间离散化，事物或事件的时间属性用时间点表示。用时间点的形式来表示时间元素，这和系统的时间量子及时间粒度的关系较大。

基于点(Point-based)的时间元素，又称为时间点，或称时刻(Time Point)。这种描述方法是把时间看成一个个离散的时间点，这些离散化的时间点的间隔大小适度时，就可以准确地描述现实世界事件发生及变化的状况。时间点是时间轴上的一点，它是和时间粒度相关的。例如 1995 年 3 月 1 日,时间粒度精确到“天(day)”。如果系统使用的最小时间粒度是“秒(second)”，则该时间点在系统内的表示必须换算成 1995 年 3 月 1 日 0 时 0 分 0 秒。

1. 时间点系统的形式描述

令整个时间系统：$I=\langle P,\leqslant_t\rangle$，其中 $P=\{p_1,p_2,\cdots,p_n\}$ 为时间点的有限集合。

“$\leqslant_t$”表示 $P$ 上的时序，$p_i\leqslant_t p_{i+1}$表示 $p_i$ 不会出现在 $p_{i+1}$ 之后,即 $p_i$ 发生在 $p_{i+1}$ 之前或跟 $p_{i+1}$ 同时发生。由此可以得到：

$\forall p_i\in P, p_i\leqslant_t p_i$ ;关系 $\leqslant_t$ 系在 $P$ 上是自反的；

$\forall p_i,p_j\in P$，若 $p_i\leqslant_t p_j$ 且 $p_j\leqslant_t p_i$ 则 $p_i=p_j$ 关系 $\leqslant_t$ 系在 $P$ 上是反对称的；

$\forall p_i,p_j,p_k\in P$，若 $p_i\leqslant_t p_j$ 且 $p_j\leqslant p_k$ 则 $p_i\leqslant_t p_k$ 关系 $\leqslant_t$ 系在 $P$ 上是传递的；

因此，$I$ 是一个偏序。同时，对 $\forall p_i,p_j\in P$ ,或者 $p_i\leqslant_t p_j$ 或者 $p_j\leqslant_t p_i\in I$,因此 $I$ 又是一个全序。因为 $I$ 是一个全序，具备了良好的代数结构，所以这对后面 $I$ 的应用提供了良好基础。

2. 传统数据库中基于点的时间元素

在传统关系数据库中也有一些带有时间特性的属性，如“生日”、“出生日期”等，这些属性的值一般为一个时间点。在 RDBMS 中，对于这些时间值元素的管理和普通的“整型”、“实型”、“字符串型”类似。这种出现在传统关系数据库中的时间称为用户自定义时间。

传统数据库中对于用户自定义时间属性的操作也是和普通的属性操作是一样的，可以

将时间属性值看成是普通的字符串进行代数操作，比如进行两个时间属性值的先后比较（也就是对应字符串的排列先后比较），或者属性值与某类时间值进行加减等。如表 2.2 所示的数据库中含有时间的扩展项（出生日期），这是一个用户自定义时间。

表 2.2　员工工资职称表

| 工　号 | 姓　名 | 出生日期 | 职　务 | 工　资(元) |
|---|---|---|---|---|
| 019504478 | 李小山 | 19720218 | 讲师 | 1000 |
| 019504479 | 张又宝 | 19680118 | 副教授 | 1800 |
| 019504480 | 毛大江 | 19640920 | 教授 | 2100 |

对时间属性项的操作，比如查找出生日期早于 1965 年的人员，那么就可以对数据库中的时间属性值和 19650000 字符串进行比较，按字典排序进行比较，显然只有 19640920 字符串的先于 19650000，由此可见，此次查寻要找的符合条件的元组是表上的第 3 行元组 019504480 的王大江元组。当然有时为了应用的需要，要对数据库中的元组时间属性值和时间元素进行加减，例如，要查找生日在国庆节前 10 天的元组即可以采取如下做法：对表中的每一个元组的出生日期进行加上 10 天的操作，即 19720218＋10，19680118＋10，19640920＋10，结果可以看出没有一个元组的时间属性值最后几位是 1001(表示国庆节)的，可见没有一位是在国庆节前 10 天出生的。

由此看来，在带有时间属性的传统关系数据库中，对时间属性项的操作和普通的字符串操作相差无几，可以对它进行各种常规的关系数据库中的操作，而元组的时间属性的取值域与时间粒度是和应用要求相关联的。

### 2.2.2　时间区间

在基于区间(Interval-based)的时间元素中，时间的基本单位为时间段或者时间区间，即通过描述时间段的起始点和终止点来描述时间区间。时间区间是指一段时间，有固定的起止时间点，例如 1995～1996 年。时间区间的表示方法根据两端时间点是否封闭分为 4 种，如图 2.2 所示。

利用时间区间也可以方便地描述时间点。例如，令 $p_i = p_j$，这时的时间区间可以理解为延续时间为 0 的一段时间，即时间轴上的某个时间点。

在这 4 种区间的表达方式中，一般采用前端封闭，尾端开放的形式，主要是考虑时间区间兼容时间点的表示方式和时间区间的比较谓词的缘故。

例如表 2.3，在每一个元组后面加上工作的起止时间。这个起止时间值所表示的就是

需要全本请访问 www.ertongbook.com