

Bases de Datos Temporales, Espaciales y Espacio-Temporales

Nicolás Del Piano

Índice general

| | |
|-----------------------------------|----|
| Resumen | 2 |
| Introducción | 3 |
| Bases de Datos Relacionales | 4 |
| Bases de Datos Temporales | 5 |
| Bases de Datos Espaciales | 9 |
| Bases de Datos Espacio-Temporales | 10 |

Resumen

Este trabajo presenta una introducción hacia las Bases de Datos Temporales, Espaciales y Espacio-Temporales. Las Bases de Datos Temporales (*Temporal Databases*) están diseñadas para la captura de información que varía en el tiempo. Las Bases de Datos Espaciales (*Spatial Databases*) fueron concebidas por la necesidad de registrar el cambio geográfico y físico de cierta información. Por último, las Bases de Datos Espacio-Temporales (*Spatio-Temporal Databases*) son el resultado de la unión de las capacidades y propiedades ofrecidas por ambos tipos de Bases de Datos. Primero se presentarán conceptos de Bases de Datos clásicas, para luego abordar más claramente los temas centrales de esta monografía. El segundo capítulo presenta de una manera detallada las Bases de Datos Temporales, el tercero lo hace para las Espaciales, y el cuarto para las Espacio-Temporales. Por último se abordan tópicos generales relacionados con estos tipos de Bases de Datos.

Introducción

Hoy en día, la cantidad de información que manejan las corporaciones y empresas es gigantesca. Por esta razón, es necesario el uso de una herramienta que provea una forma de gestionar adecuadamente esta información. Este es el propósito de las Bases de Datos: brindar al usuario una forma de controlar el acceso, almacenamiento y administración de los datos de la entidad en cuestión.

Con la aparición de nuevas tecnologías, el surgimiento de nuevas necesidades fue inevitable, implicando que las Bases de Datos Relacionales no sean una bala de plata (aunque sean las más usadas actualmente) para resolver todos los problemas de gestión de datos. Surgieron conceptos como Minería de Datos, Data Warehouse y Big Data: la información ya no tiene la misma dimensión que antes. Fue entonces cuando el Modelo Relacional clásico necesitaba extenderse para representar eficientemente datos que varíen en tiempo y espacio.

Las Bases de Datos Temporales se encargan del dominio del tiempo y su relación con los datos, permite analizar la historia y controlar la validez temporal de los mismos. Una gran variedad de aplicaciones del mundo real manejan datos variables en el tiempo: control de inventario, registros médicos, operaciones bancarias, sistemas de información geográfica, gestión de reservas, aplicaciones científicas, etcétera. Esta necesidad de referencias temporales justifica la creación de un modelo de datos temporal.

Las Bases de Datos Espaciales extienden el modelo para representar el dominio espacial, con estructuras que puedan identificar un objeto en el espacio. Deben permitir la descripción de objetos espaciales mediante tres características: atributos, localización y topología. Además deben proveer tipos de datos espaciales para estructurar entidades geométricas en el espacio. Existen diversas áreas donde la gestión de información geométrica, geográfica o espacial es crucial: Sistemas de Información Geográfica, Bases de Datos multimedia, imágenes satelitales, ciencias ambientales, astronomía.

El objetivo de las Bases de Datos Espacio-Temporales es extender los modelos de información espacial para incluir el tiempo y describir en forma más dinámica la realidad que se quiere representar. El modelo espacio-temporal abarca aplicaciones demográficas, ecológicas, relacionadas con marketing, militares, urbanísticas y de fenómenos naturales, entre otras.

Bases de Datos Relacionales

Bases de Datos Temporales

El tiempo es un aspecto importante para los fenómenos del mundo real: los eventos ocurren en momentos de tiempo específicos.

A veces nos interesa saber con cierta certeza cuándo ocurrió tal evento, y poder compararlo con otros para obtener información de interés.

Muchas de las áreas donde se aplican las Bases de Datos tienen naturaleza temporal:

- Control de inventario.
- Registros médicos.
- Sistemas de información geográfica.
- Operaciones bancarias.
- Data Warehousing.
- Sistemas de control de reservas (aerolíneas, hoteles, etc).
- Aplicaciones científicas.

Relaciones no temporales

En la Figura 3.1 puede observarse una tabla relacional no temporal. Cada tupla

| Id | Nombre | Estado | Sueldo |
|----|--------|--------|--------|
| 1 | Juan | Activo | 5700 |
| 2 | Manuel | Activo | 2300 |

Figura 3.1: Tabla relacional no temporal.

representa un hecho verdadero *ahora*. Solo hay un estado representable de la Base de Datos: *el actual* (*current snapshot*). A medida que el tiempo transcurre, los datos se van actualizando y modificando. Con este modelo, perdemos información. Las Bases de Datos convencionales representan el estado de la información en un instante de tiempo dado. Aunque la Base de Datos es actualizada, estos cambios son vistos como modificaciones del estado actual y los datos obsoletos son borrados. Por lo tanto, solo podemos utilizar la información actual de la Base de Datos. Esto genera un problema cuando queremos responder preguntas involucradas a intervalos de tiempo: *¿Cuáles empleados percibieron un aumento el mes pasado?*

Bases de Datos Temporales: Definición

Un *DBMS Temporal* es un Sistema de Gestión de Bases de Datos que proporciona herramientas para el manejo y control de Bases de Datos Temporales.

Una *Base de Datos Temporales* es una Base de Datos que tiene dimensión del tiempo a través del almacenamiento de datos temporales.

Proporcionan un marco que mantiene la historia de los cambios que se produjeron en la fuente de datos. Están diseñadas para la captura de la información que varía en el transcurso del tiempo (puede apreciarse esta relación en la Figura 3.2).

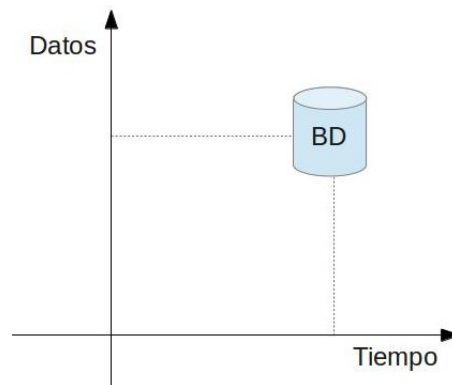


Figura 3.2: Relación tiempo y datos.

Datos Temporales

Un *Dato Temporal* es un dato convencional al que se le asocia un período de tiempo para expresar valores temporales en la Base de Datos.

Este agregado de información temporal se denomina *time-stamping*. Al asociar el tiempo con la información, es posible almacenar diferentes estados de una base de datos.

Dimensión del Tiempo

El tiempo es infinito, continuo y multidimensional [?]. Las computadoras no pueden representar información continua, de hecho, se aproximan discretamente. Así, para representar una noción del tiempo, se lo transforma en un conjunto discreto con una cierta granularidad. Por ejemplo, la sentencia *Homero Simpson nació el 12 de Mayo de 1956* tiene una granularidad de días.

Las Bases de Datos Temporales almacenan dos dimensiones de tiempo:

- Tiempo Válido
- Tiempo Transaccional

El *Tiempo Válido* representa cuándo un hecho tiene validez, es decir, es verdadero en el mundo real. El Tiempo Válido de un evento es el tiempo de un reloj en el que ese evento ocurrió [?]. Es independiente de si dicho evento fue registrado o no en la Base de Datos. Los Tiempos Válidos pueden encontrarse en el pasado, presente o futuro. Una de las características es que todos los eventos tienen asociado un Tiempo Válido, pero no necesariamente son registrados. Además brindan

la capacidad de gestionar la historia de la Base de Datos.

El *Tiempo Transaccional* registra el período de tiempo donde un hecho fue almacenado en la Base de Datos. Permiten realizar consultas que muestren el estado de la Base de Datos en un tiempo específico. Este tiempo está acotado en ambos extremos; la creación de la Base de Datos y el tiempo presente, es decir, los Datos Transaccionales viven solamente dentro de la vida de una Base de Datos. Una de las capacidades interesantes es que permiten volver hacia un estado anterior (*roll-back*), ya que almacenan datos de las operaciones que se fueron haciendo.

Estas dos dimensiones son ortogonales. Un Modelo de Datos que no soporte ninguno de estas dimensiones se denomina *snapshot*, ya que captura solamente una imagen de la Base de Datos. Si se brinda soporte para Tiempo Válido, entonces se cuenta con una Modelo de Datos *histórico*, mientras que uno que soporte Tiempo Transaccional solamente se denomina *rollback*. En caso de que ambas dimensiones estén soportadas, se denomina *bitemporal*.

Ejemplo

Base de Datos Bitemporal

Incluyen ambos tiempos (Válido y Transaccional) lo que les permite proveer información histórica, a la vez que brindan la capacidad de hacer roll-back de los datos. En la siguiente figura se puede apreciar un ejemplo:

| Nombre | ViveEn | Valid-From | Valid-To | Transaction-From | Transaction-To |
|---------|------------------|-------------|-------------|------------------|----------------|
| Señor X | Springfield | 12-May-1956 | ∞ | 13-May-1956 | 16-Aug-1980 |
| Señor X | Springfield | 12-May-1956 | 2-Aug-1980 | 16-Aug-1980 | ∞ |
| Señor X | Arroyos Cipreses | 3-Aug-1980 | ∞ | 16-Aug-1980 | 20-Apr-2004 |
| Señor X | Arroyos Cipreses | 3-Aug-1980 | 20-Apr-2004 | 20-Apr-2004 | ∞ |

Figura 3.3: Ejemplo de una relación bitemporal.

Extensiones Temporales

Hay dos formas de extender el modelo relacional para especificar requisitos temporales.

La forma mostrada en los ejemplos antes mencionados se denomina **marcaje de tuplas**. Este método es muy común en el modelo relacional. Se utiliza un atributo especial para indicar la validez de una tupla: se indica *desde* y *hasta* para representar intervalos de tiempo.

$$(attr_1, \dots, attr_n) \rightarrow (attr_1, \dots, attr_n, temp_attr_1, \dots, temp_attr_m)$$

Una de las desventajas que tiene esta forma es que una entidad puede estar representada por varias tuplas, por lo que no puede lograrse una representación 1:1 de la realidad lo que podría generar información redundante.

La segunda forma se denomina **marcaje de atributos** y usa atributos multivaluados. La idea es que la marca de tiempo y la entrada referenciada se almacenen

en el mismo atributo de forma anidada. Al mismo tiempo que permite la correspondencia 1:1 entre entidades y hechos reales, dificulta las actualizaciones y no cumple la 1NF.

Operadores de Allen

Necesitamos una forma de comparar datos temporales. Allen (1983) propone un conjunto de operadores temporales lógicos para comparar intervalos de tiempo. El operador es una función de tipo: $I_t \times I_t \rightarrow \{True, False\}$

Modelo Relacional Temporal

El modelo relacional temporal incorpora la semántica temporal en el modelo relacional utilizando marcaje de tuplas como extensión temporal.

Alguna de las características que debe ofrecer son:

- Un tipo de datos para períodos de tiempo, incluyendo la habilidad de representar períodos infinitos.
- Soporte para tiempo válido, transaccional y tablas bitemporales.
- Tiempo transaccional controlado por el sistema.
- Consultas temporales.
- Predicados y operadores que actúen sobre intervalos de tiempo.

Lenguaje de consulta temporales

Una Base de Datos Temporal es un repositorio de información temporal. Un lenguaje de consulta temporal es cualquier lenguaje de consulta para bases de datos temporales.

Las propiedades de un lenguaje de consulta temporal son:

- Semántica declarativa
- Implementación eficiente.
- Independencia en la representación.
- Expresividad en la consulta.

Algunos ejemplos importantes son: TSQL, TQuel, HRDM, Backlog. Muchos otros son derivados de éstos.

DBMS Temporal

Los DBMS Temporales deben respaldar un lenguaje de definición de datos temporales, un sistema de restricciones temporales, un lenguaje de manipulación de datos y un lenguaje de consulta temporal.

Algunos ejemplos de DBMS Temporales son: TimeDB, Oracle Workspace Manager, PostgreSQL, IBM DB2, entre otros.

Bases de Datos Espaciales

Bases de Datos Espacio-Temporales