

El Rol de las Bases de Datos Espaciales en una Infraestructura de Datos

Mariella Gutiérrez
Universidad Católica de la Santísima Concepción
Chile
marielag@ucsc.cl

Resumen

Una Base de Datos Espacial permite describir los objetos espaciales que la forman a través de tres características básicas: atributos, localización y topología. Los atributos representan características de los objetos que nos permiten saber qué es lo que son. La localización, representada por la geometría del objeto y su ubicación espacial de acuerdo a un sistema de referencia, permite saber dónde está el objeto y qué espacio ocupa. Por último, la topología definida por medio de las relaciones conceptuales y espaciales entre los objetos, permite mejorar la interpretación semántica del contexto y establecer ciertas jerarquías de elementos a través de sus relaciones.

Por otro lado, desde el punto de vista tecnológico, una Infraestructura de Datos Espaciales debe incluir datos y atributos geográficos, metadatos, métodos de búsqueda, de visualización y mecanismos para proporcionar acceso a los datos espaciales.

En este trabajo se analiza cómo se vinculan las Bases de Datos Espaciales con las necesidades de una Infraestructura de Datos Espaciales. Para ello se estudian aspectos relacionados con las tecnologías y estándares de datos espaciales, el modelamiento conceptual y lógico, el almacenamiento de objetos espaciales y los metadatos que los describen, así como también, las operaciones del álgebra relacional y funciones geométricas que pueden operar sobre uno o un conjunto de objetos relacionados conceptual o espacialmente. Todos estos aspectos considerados fundamentales cuando se quiere compartir información espacial.

Se estudia además, diversas operaciones sobre los datos espaciales basados en un modelo de objetos, como son las operaciones de la teoría de conjunto, topológicas, de orientación y métricas. Se establece un paralelo de cómo pueden ser resueltas estas operaciones utilizando un motor de base de datos con extensión espacial y un software de sistema de información geográfico clásico. Para ello se utilizan como herramientas el RDBMS Postgresql con su extensión espacial PostGIS y el software de Sistemas de Información Geográfica ArcView.

Por último se presentan aspectos de accesibilidad y representación de los datos a través de las propias funcionalidades de las bases de datos espaciales y también de su conexión con otras herramientas.

Introducción

La evolución de los administradores de bases de datos (DBMS) hacia la incorporación de extensiones para el manejo de datos espaciales obedece a la necesidad de manejar de forma integrada tanto los datos descriptivos como los datos espaciales o geométricos de los objetos que la forman.

La mayoría de los sistemas de información geográficos adoptan una arquitectura en la cual los datos son administrados a través de dos modelos de datos heterogéneos: un DBMS para los datos descriptivos y un módulo específico para la administración de datos espaciales. Esto puede provocar inconsistencias, dificultad para la validación, integración y recuperación de los datos (Rigaux 2002), (Zlatanova 2006).

Los administradores de bases de datos espaciales (SDBMS) permiten la adopción de una arquitectura integrada, en la cual el administrador de datos es extendido para almacenar tanto la descripción de los objetos como su geometría. El lenguaje de consulta (SQL) es extendido para manejar nuevos tipos de datos (puntos, líneas y polígonos) y son incorporadas nuevas funciones que permiten la selección y recuperación de los datos, no sólo por criterios alfanuméricos, sino que también aplicando criterios espaciales a través de relaciones topológicas, de orientación, medición, entre otras (Rigaux 2002)

Por otro lado, un aspecto clave en una Infraestructura de Datos Espaciales es proveer mecanismos que permitan buscar, recuperar, compartir e integrar datos espaciales. Para el cumplimiento de este objetivo se plantea como un elemento clave el adoptar estándares para la definición de los elementos espaciales (features), su modelamiento (feature schema), tener una completa y consistente descripción de los datos, lo que se conoce como metadatos y también sistemas de metadatos que permitan la gestión de los mismos.

Si bien existen estándares para la información geográfica (ISO/TC211) y en particular para los metadatos (ISO/TC211 19115), se plantea como un problema el tener sistemas eficientes para administrar los metadatos y también mecanismos para que los datos puedan ser compartidos en línea (Cao 2002), (Phillips 1998).

Con respecto a los sistemas de gestión de metadatos, éstos han sido definidos de manera que permiten el almacenamiento, la actualización y la búsqueda de datos espaciales a través de interfaces interactivas, generalmente a través de su implementación en lenguaje XML.

Su finalidad es permitir que los usuarios que requieran datos espaciales puedan conocer dónde pueden encontrarlos, qué características tienen, cuándo fueron actualizados, en qué formatos están disponibles y cómo pueden acceder a ellos (ISO/TC 211 19115 2000) (GSDI 2004). Sin embargo, no es un objetivo fundamental el proveer los datos en línea y cuando lo hacen, proveen el acceso a datos espaciales siguiendo los protocolos y formatos definidos por los estándares para la transferencia de datos espaciales vectoriales y raster (DIGEST, SDTS, DXF, entre otros). Deberían ser definidos también metadatos para la búsqueda y transferencia de datos espaciales almacenados en repositorios de bases de datos espaciales a través del lenguaje SQL y una conexión ODBC (Open Database Connectivity) o JDBC (Java Database Connectivity), por ejemplo. Para esto, sería necesario incorporar a los metadatos la dirección IP del servidor de bases de datos, el nombre de la base, un nombre de usuario y password. Esta iniciativa ha sido parcialmente implementada en la infraestructura de datos espaciales canadiense (CGDI) en la cual se plantea que cada proveedor de datos espaciales que desee compartir sus datos, instale un servidor de búsquedas que permita recuperar datos geográficos a través de consultas a los metadatos almacenados en un Database Server (Geoconnections 2004).

A continuación se presentan algunos avances en el campo de las bases de datos espaciales que favorecen su incorporación a una infraestructura de datos espaciales, analizados desde la perspectiva de la adopción de los estándares para la descripción, representación e implementación de información geográfica.

Descripción de la Información Geográfica (Metadatos)

Un tema crucial para incorporar las capacidades de una base de datos espacial a una infraestructura de datos espaciales (IDE), es que ésta esté desarrollada siguiendo los estándares para la descripción, definición espacial, operación y acceso a información geográfica.

En particular, las iniciativas existentes para la creación de estándares de datos espaciales, llevadas a cabo por la International Standard Organization (ISO) y el OpenGIS Consortium (OGC) proveen estándares para la información geográfica.

Para la descripción de la información geográfica, existe el estándar de los metadatos, Geographic Information – Metadata (ISO/ TC 211 19115 2000), el cual se aplica al catálogo datos, clearinghouse y a la descripción de datos espaciales. Los metadatos están estructurados en una serie de paquetes de metadatos que agrupan las descripciones en diferentes categorías. La IDE de Chile se basa en estas categorías y reagrupa su contenido en categorías según se presenta en el Sistema Nacional de Información Territorial (SNIT 2006)(Tabla 1).

ISO/TC211 19115	SNIT
1. Descripción	1. Información de Contacto para los Metadatos
2. Restricciones	2. Identificación del Dato
3. Calidad de los datos	3. Coordenadas Geográficas
4. Mantenimiento	4. Sistema de Referencia
5. Representación Espacial	5. Restricciones
6. Sistema de Referencia	6. Distribución del Dato
7. Contenido	7. Formato de Distribución del Dato
8. Representación gráfica	8. Opciones de Transferencia Digital
9. Distribución	9. Información de Imagen
10. Extensiones a los metadatos	
11. Esquema	
12. Extensiones de datos espaciales y temporales	
13. Citas y partes responsables	

Tabla 1: Categorías del Diccionario de Datos de los Metadatos

Cada una de las categorías presentadas en la tabla anterior se detalla a través de un conjunto de atributos, los cuales de acuerdo a su importancia pueden ser obligatorios u opcionales. Además, para optimizar las búsquedas, muchos de los valores de los atributos están definidos por dominios de datos. Un ejemplo del uso de dominios que parece particularmente importante, es el metadato que describe el contenido en la norma ISO y área temática definida dentro de la categoría identificación del dato en el diccionario del SNIT. A través de estos dos campos no sólo se clasifica la información espacial, sino que también se establece una correspondencia entre un dato definido por un estándar global y su aplicación en un contexto local. La siguiente tabla muestra las categorías temáticas de los datos espaciales definidas por ISO y el SNIT (Tabla 2). La correspondencia se establece por el usuario al completar los metadatos. Una vez determinada el área temática de acuerdo

al SNIT, se selecciona del listado de categorías ISO la que mejor represente el conjunto de datos, pudiéndose también agregar uno o más campos para asociar el dato a más de una categoría (SNIT 2006).

Esta clasificación de los datos espaciales en categorías preestablecidas, adquiere gran relevancia cuando se quiere compartir información a través de bases de datos espaciales heterogéneas. Si dos o más bases de datos espaciales adoptan el estándar para la descripción de datos y en particular su clasificación en categorías ISO, la implementación de herramientas de búsqueda que utilicen esta descripción como un criterio de selección primario lograría acotar el conjunto de datos involucrado en la consulta, lo que debería tener un efecto positivo en la calidad de la respuesta y el tiempo involucrado.

Categoría ISO	Área Temática del SNIT
Agricultura	Información Territorial Base
Biología	Infraestructura
Límites	Social
Climatología / Meteorología /Atmósfera	Recursos Naturales
Economía	Propiedad
Elevación	Patrimonio
Ambiente	Planificación Territorial
Información Geocientífica	
Salud	
Imágenes basadas en la cubierta de la tierra	
Inteligencia Militar	
Alta Mar	
Localización	
Océanos	
Catastro y planificación	
Sociedad	
Infraestructura	
Transporte	
Telecomunicaciones	

Tabla 2: Categorías de Datos Espaciales (SNIT 2006)

Definición de Objetos Geométricos (Spatial Schema and Simple Feature Specification)

Además del estándar para la descripción de la información geográfica, si se desea acceder a información geográfica desde un repositorio de datos espaciales, se requiere que este repositorio adopte también estándares de definición de objetos geométricos y de operadores y funciones para el tratamiento de los datos.

Los motores de bases de datos espaciales como el Oracle Spatial y PostGIS adoptan el estándar del OpenGIS que define datos geométricos de tipo point, linestring, polygon, multipoint, multilinestring, multipolygon y geometrycollection (ISO/TC 211 19107 2001) (ISO/TC 211 19125-2 2001) (OGC 2006) (Postgis 2006) (Oracle 2005) fundándose en un modelo espacial basado en objetos (Shekhar 2003, p.26-31). En el caso de PostGIS un atributo geométrico de un elemento espacial sólo puede almacenar uno de estos tipos, a diferencia de OracleSpatial que puede almacenar más de uno (por ejemplo, polygon y

multipolygon en una misma columna. Esto presenta un problema en la implementación de funciones geométricas que pudieran retornar elementos de más de un tipo (Martínez 2005). Además, OpenGIS define dos formas de representar los objetos espaciales, adoptadas por PostGIS: (1) (WKT) Well-know text como los ejemplos anteriores y, (2) (WKB) Well-know binary. Las dos formas guardan información del tipo de objeto y sus coordenadas. Además la especificación OpenGIS requiere que los objetos incluyan el identificador del sistema de referencia espacial (SRID), el cual es requerido cada vez que se inserta una instancia en un objeto espacial. El modelo conceptual de los objetos espaciales está formado por tres tablas (figura 1): una que define el objeto en si, y otras dos tablas de metadatos para definir el sistema de referencia espacial, llamada spatial_ref_sys y la columna o atributo geométrico, llamada geometry_column (OGC 2006)(ISO/TC 211 19125-2 2001).

La implementación en lenguaje SQL de las tablas de metadatos de objetos espaciales, son definidas en los estándares de especificación de objetos (OGC 2006),(ISO/TC 211 19125-2 2001). Las bases de datos espaciales adoptan estos estándares pues utilizan como lenguaje básico el SQL, sin embargo cada uno de los motores puede implementar diferentes variaciones lo que va en desmedro de la interoperatividad (Martínez 2005).

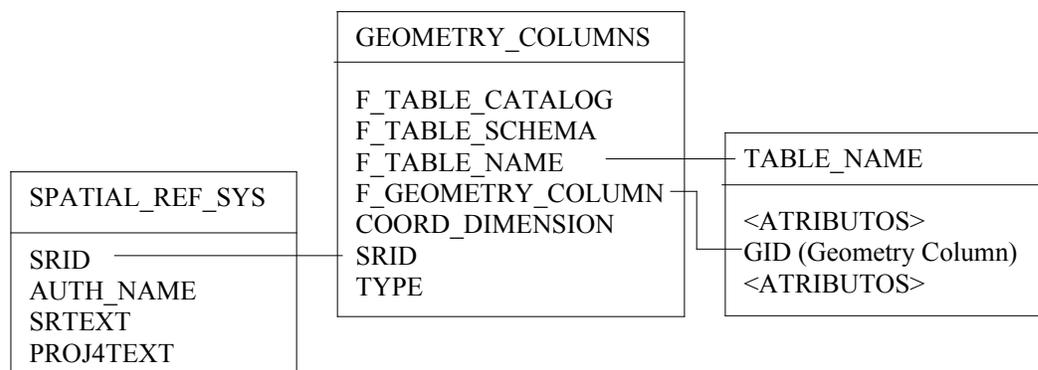


Figura 1: Definición de un objeto espacial en el estándar OpenGIS.

La tabla spatial_ref_sys contiene un identificador numérico y una descripción textual del sistema de coordenadas espacial de la base de datos. Esta definición es concordante con la que incluye el metadato para el sistema de referencia de la norma ISO/TC211 19115. En PostGIS esta tabla viene con los datos de los sistemas de referencia espacial en formato WKT según el estándar establecido por el EPSG (European Petroleum Survey Group) el cual también es utilizado por el SNIT (OpenGIS Specificacion), por lo que al ingresar un dato espacial, su columna geométrica debe hacer referencia a identificador del sistema de referencia que corresponda al dato (SRID), lo cual asegura la consistencia de los datos. En el SNIT el metadato del sistema de referencia se compone de cuatro atributos Datum, Huso, Proyección y Elipsoide los cuales no hacen referencia a valores dentro de un dominio. Esto quiere decir que la búsqueda de datos espaciales dentro de un sistema de referencia común se hace más compleja al no existir una correspondencia directa entre los valores de los datos.

Por otro lado, la tabla geometry_data_column define la tabla del elemento espacial a través de tres atributos, como son el nombre de la base de datos, del esquema y de la tabla. Además incluye el nombre de la columna geométrica, su dimensión (2D ó 3D), el

identificador del sistema de referencia espacial (SRID) y el tipo de dato espacial (POINT, LINESTRING, POLYGON, MULTYPOINT, GEOMETRYCOLLECTION).

Esta definición de datos espaciales, basada en el estándar, permite además la implementación de operaciones y funciones geométricas y/o topológicas para el tratamiento de los datos, lo cual también está definido en OpenGIS Simple Feature Specification for SQL (OGC 2006).

Definición de operaciones y funciones sobre los datos Geométricos (Simple Feature Specification for SQL)

Así como es importante la descripción y la definición de la información espacial, es de igual importancia la definición de operaciones y funciones que permitan su real implementación y uso. El estándar para la implementación de la especificación de objetos geométricos en SQL (OGC 2006) define tres categorías de funciones sobre objetos de tipo geométrico: (1) funciones básicas que permiten la descripción de un objeto geométrico, (2) funciones de consulta de relación espacial ente dos geometrías y, (3) funciones que implementan operadores espaciales.

El primer grupo de funciones permite conocer la descripción de un objeto geométrico a través del tipo de geometría (POINT, LINESTRING, POLYGON, MULTYPOINT, GEOMETRYCOLLECTION), dimensión (2D ó 3D), sistema de referencia utilizado (SRID), geometría en los formatos WKT o WKB, entre otras características.

El segundo grupo de funciones está orientado a conocer el tipo de relación espacial que existe entre dos objetos geométricos. Estas funciones se implementan a través de un nombre de función que recibe como parámetros los campos de geometría de ambos objetos y retorna verdadero o falso dependiendo si se da o no el tipo de relación consultada entre éstos. Dentro de estas funciones es posible conocer por ejemplo si dos objetos son iguales, están disjuntos, se intersectan, si uno contiene a otro, si son adyacentes o se cruzan.

Por último, el tercer grupo de funciones implementa operaciones sobre una o dos geometrías las cuales se definen por un nombre de función que recibe como parámetros de entrada, el nombre de uno o dos campos geométricos y otros parámetros numéricos si se requieren, y entrega como salida una nueva geometría. Dentro de estas funciones es posible conocer la geometría resultante de la intersección, unión y diferencia entre dos geometrías, además de un buffer de una cierta dimensión alrededor de una geometría, entre otras operaciones.

A continuación se presenta la implementación de algunas de estas funciones en el SDBMS PostGIS. También se muestra qué significa realizar estas mismas operaciones en una herramienta de sistemas de información geográfica como ArcView (PostGIS 2006) (Manual ArcView 3.2a) (Tabla 3).

Función	PostGIS	ArcView
Geometría	geometrytype(geometry): string	Menu theme – Table – ver el contenido del campo shape
Dimensión	Dimension(geometry): integer	No es visible directamente desde los menús.
Sistema de Referencia	srid(geometry): integer	No es visible directamente desde los menús.

Geometría en formato WKT	astext(geometry): text	No es visible directamente desde los menús.
Disjuntos	disjoint(geometry,geometry): Bool	Not (aShape.Intersects (anotherShape)): Bool
Intersección	Intersects(geometry,geometry): Bool	aShape.Intersects (anotherShape): Bool
Adyacencia	touches(geometry,geometry): Bool	No es visible directamente desde los menús. Sólo es posible ver polígonos adyacentes cuando se hace una selección de tema sobre tema, con un único tema.
Intersección	intersection(geometry,geometry): geometry	Menu Edit – Intersect Features Menu View Geoprocessing Wizard – Intersect two themes
Unión	geomunion(geometry,geometry): geometry	Menu Edit – Union Features Menu View Geoprocessing Wizard – Union two themes
Buffer	buffer(geometry, float8[, int4]): geometry	Menu theme – Create Buffer – (theme name, distance, distance unit, output form)

Tabla 3: Funciones sobre datos geométricos

En el caso de PostGIS, si se requiere realizar una función sólo con algunos registros de una tabla, se suma a las funciones criterios de selección adicionales que permitan acotar el universo de instancia sobre el cual se van a ejecutar. Estos criterios pueden estar basados en atributos alfanuméricos o espaciales. En el caso de ArcView las consultas se ejecutan sobre todo el universo de los datos o si existen registros u objetos previamente seleccionados, entonces, la función se ejecutará sobre éstos. Esto significa que una consulta aplicando criterios de selección alfanuméricos y/o espaciales en PosGIS se ejecutará en una instrucción y en ArcView en al menos dos (una para la selección y otra para aplicar la función).

Implementación gráfica

Quizá el mayor desaliento a trabajar puramente con bases de datos espaciales es que los administradores no incluyen interfaces para la visualización y tratamiento de los datos de forma gráfica, al igual que los sistemas de información geográficos. Sin embargo, existen numerosas alternativas de software libre (OpenSource) que incorporan capacidades de conexión con bases de datos espaciales y en especial con PostGIS. Dentro de estas opciones se encuentran sistemas de información geográfica de escritorio, así como también software que proporcionan servicios de OpenGIS (Manso 2006) y paquetes de librerías (PostGIS 2006).

Dentro de los sistemas de información geográficos de escritorio, con conexión a PostGIS, se encuentra GvSIG, Udig y Quantum-GIS, entre otros. Dentro de los software servidores se encuentra MapServer, GeoServer y Deegree entre otros. Así como también existen paquetes de librerías como CEOS (de Java Topology Suite) utilizada por PostGIS para la

implementación de funciones topológicas y GDAL PostGIS que permite el manejo de datos raster.

Conclusiones

En este trabajo se presenta cómo se conectan las bases de datos espaciales con una infraestructura de datos espaciales (IDE), basándose en el análisis del nivel de adopción de los estándares para la información geográfica por parte de los sistemas gestores de bases de datos espaciales y de las IDE's. Se considera el análisis aquí presentado como preliminar y no ambicioso en el sentido de que se han tomado en cuenta sólo algunos aspectos de los estándares básicos, como son el estándar de metadatos, la especificación de objetos espaciales y su implementación en lenguaje SQL.

El paso siguiente a este análisis es profundizar más en el tema de la implementación de IDE's que permitan la búsqueda y recuperación de información geográfica almacenada en bases de datos espaciales heterogéneas. Para lo cual, es necesario abarcar temas de estándares para la transferencia de datos, la seguridad, aseguramiento de la calidad, protocolos de comunicación, entre otros. Así como también aspectos sociales y culturales que fomenten el que los propietarios de datos geográficos los compartan en un mundo globalizado.

Bibliografía

Cao Y., H. Wu and J. Bi, The Research of Metadata Management of National Resource and Environment Spatial Database. Geo-Information Science 2002, Vol 4, Nro 2, 6-10.

Geoconnections, 2004, A Developers' Guide to the CGDI: Developing and publishing geographic information, data and associated services. Geoconnections, Canada.

Global Spatial Data Infraestructure, GSDI, 2004, Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, Technical Working Group Chair, GSDI, Version 2.0, January 2004.

ISO/TC 211 19107, 2001, Geographic Information – Metadata, Internacional Organization for Standarization.

ISO/TC 211 19115, 2000, Geographic Information – Metadata. Internacional Organization for Standarization.

ISO/TC 211 19125-2, 2001, Geographic information — Simple feature access, Part 2, SQL Option. Internacional Organization for Standarization.

Martínez, J. and E. Coll, Análisis vectorial en Postgis y Oracle Spatial: Estado actual y evolución de la especificación Simple Features for SQL. JIDEE 2005, Jornadas Técnicas de la IDE en España, Noviembre de 2005.

OGC, 2006, OpenGIS Consortium Standards, <http://www.opengis.org>.

Oracle, 2005, Oracle® Spatial, User's Guide and Reference, 10g Release 2 (10.2).

Phillips, A., I. Williamson and Ch. Ezigbalike, 1998, The Importance of Metadata Engines in Spatial Data Infrastructures, AURISa 98, The 26th Annual Conference of AURISA Perth, Wester Australia, Noviembre de 1998.

PostGIS, 2006, Manual de PostGIS. <http://postgis.refractor.net/documentation/>

Rigaux, Ph., M. Scholl and A. Voisard, 2002, Spatial Databases: With Application to GIS. Elsevier Science , 24-25.

Shekhar, Sh. And S. Charla, 2003, Spatial Databases: A Tour. Pearson Education Inc. Estados Unidos.

SNIT, 2006, Sistema Nacional de Información Territorial de Chile, <http://www.snit.gob.cl>.

Zlatanova, S. and J. Stoter, 2006, The role of DBMS in the new generation GIS architecture. Rana, Sanjay; Sharma, Jayant (Eds.), In: Frontiers of geographic information technology 2006, 8,155-180.