

数据库在地理信息系统中的应用研究

丁力 汪小林 许卓群

(北京大学计算机系, 北京, 100871)

摘 要: 地理信息系统 (GIS) 的重要任务就是海量空间属性管理, 数据库作为高效的数据管理工具, 是 GIS 重要研究课题。本文提出 GIS 数据管理的两个重要问题——表达问题和效率问题, 然后分析比较一些现存的 GIS 数据管理策略, 指出包装模型的实用性。

关键词: 地理信息系统 对象-关系数据库 表达 效率

1. 前言

地理信息系统 (GIS) 就是用于存储、检索、操作、分析和显示地理相关信息的计算机化系统^[1]。目前 GIS 产业的应用领域分布广泛, 得到较大规模推广应用的就有: 地籍与房产管理; 农业、土地和水资源规划; 林业; 资源调查; 环境保护; 大型功能工程项目的前期准备和实施控制; 公用事业管理; 市政建设与管理; 灾害应急模拟等^[2]。

GIS 的一项关键任务就是海量地理信息的数据管理。目前的 GIS 产品主要有两种数据管理策略: 文件系统或数据库系统。

二进制文件结构简单, 表达能力强大, 可以很方便表示复杂结构。其缺点也很明显: 数据冗余度大, 难于共享数据, 容易造成数据不一致性; 程序与数据缺乏独立性, 系统不易扩充^[3]。

数据库系统较好地改进了文件系统的不足。为了解决多用户, 多应用共享数据的需求, 数据库系统做了很多工作: 减少数据冗余, 提高扩充能力; 支持数据与程序的独立性; 提供数据控制功能, 如数据的安全性控制, 数据的完整性控制, 并发控制^[3]; 提供网络功能, 如 ODBC, WEB 服务。但它也存在数据访问速度慢的问题。

随着 GIS 应用领域的扩展, 数据管理、网络服务、信息安全机制、可互操作性和智能搜索都是设计 GIS 的重要因素。文件系统的简单数据管理功能明显不足, 而数据库系统则提供了比较完善的解决策略, 可见数据库在 GIS 中大有用武之地。

2. GIS 中的数据管理问题

2.1. 表达问题

表达问题指在数据库中如何表达 GIS 应用中复杂多样的空间对象。空间对象是直接或间接地与地球表面的一个位置 (以及时间, 方向) 相关的事物, 它包括空间属性和用户属性两部分。空间属性记录了空间对象的几何信息 (位置、方向等), 而用户属性则记录了空间对象的被用户关注的信息。空间对象具有如下特征: 个体结构复杂; 存在复杂的拓扑关系; 空间属性与用户属性关联紧密。一般来说, 表达问题包括:

- 空间对象的表示与存储的一致性: 空间对象在 GIS 应用中和数据库存储中的如何表示, 两种表示模型是否一致。如果不一致, 如何维持映射关系。
- 空间属性与用户属性的关联问题: 同一个空间对象的空间属性与用户属性是否分开存储。如果是, 则它们是如何关联起来的; 如果不是, 又该如何统一管理。

2.2. 效率问题

效率问题指如何提高空间对象在数据库中的访问效率。效率问题包括：

- 独占资源的效率问题：如何提高在独占资源的操作中提高数据访问速度。例如在一张地图指定区域内搜寻空间对象的操作中，如何缩短操作时间。
- 多用户共享资源的效率问题：如何在一个支持共享资源的 GIS 中提高多用户对数据的访问速度。例如在多用户对同一张地图的访问中，如何提高速度。

3. GIS 中常见的数据库应用策略

我们采用数据库作为 GIS 数据管理的解决方案，就是希望通过数据库的特性解决上述两个问题。由于存在多种数据库系统模型，他们的实现策略也存在差异。

3.1. 关系数据库

关系数据库是最成熟的数据库产品，它以强大的关系理论为支持。其主要优点包括：物理数据存储与逻辑数据库结构的独立性；多样、简易的数据访问能力，从而提供了高效数据访问的可能性；相当灵活的数据库设计；数据存储和冗余的减少。^[3] 利用关系数据库主要有两种策略：

策略一 不透明类型模型（Opaque Types Model）

- 设计思想
把空间对象的整个空间属性直接存储在长二进制字段（LOB）中。只能通过特定的函数来访问空间属性并解释其内容结构。
- 评价
此策略实现了把空间对象存入数据库的目标。但是，空间属性对用户不透明，需要解释使用；且不支持继承，无法表示一些复杂的拓扑关系，而且每次访问数据都是对整个 LOB 的存取。

策略二 纯关系模型（Pure Relational Model）

- 设计思想
把构成空间对象的每个基本元素（点、线段）都作为一行记录存入表中，通过关系模型来维护其空间属性的内部结构以及空间对象之间的拓扑关系。
- 评价
此策略同样可以把空间对象存入数据库。但此方法难以复杂的空间对象，一个空间对象可能要涉及多张表，成百上千条记录，查询难以设计且速度较慢。

3.2. 面向对象数据库

面向对象数据库是面向对象思想与数据库管理系统相结合的产物。它摒弃了原有的关系理论，具有完全面向对象的特征，包括抽象、继承、封装、重载、多态等特性。具有很强的理论优势。但其商业化产品侧重点是程序设计语言对复杂数据的访问，还存在查询表达能力

不强，查询效率比较低的问题^[4]。目前 OODB 还难以实用于 GIS 应用。

3.3. 对象-关系模型数据库

对象-关系模型把关系理论与面向对象思想有机结合。不但继承了关系系统的优良特性，留住了使用关系系统的老用户；还提供使用面向对象方法建模的能力，满足了新用户数据库表达能力的要求。其主要特征包括支持基本类型扩充；支持基本类型扩充；支持继承；支持规则。^[4]

策略三 空间扩展模型（Spatial Extension Model）

空间扩展的主要目的就是想建立一个空间对象表达模型，把空间属性和用户属性的存储管理统一在 DBMS 之下，由 DBMS 完成 GIS 的数据层功能。参考 DB2 Spatial Extension^[5]，ORACLE Spatial Extension^[6]，可以归纳模型特性。

- 设计思想

定义一组数据类型以描述空间对象的空间属性，如点，线，多边形，点簇，线簇，多边形簇。采用继承方式（DB2）或关系方法（ORACLE）描述数据类型之间的关系。提供在数据类型上的预定义的函数以及空间索引，同时支持用户扩展。

- 评价

此策略重点在数据库上，希望通过增强数据库的空间功能来从根本上解决问题，是一种长远打算。但还存在技术不成熟的问题，如不能提供对象-关系系统应具有的所有功能；其性能/价格问题可能会影响应用范围。

策略四 包装模型（Wrapped Model）

- 设计思想

在数据库上附加一层中间件。包装层实现了从空间对象模型到数据库存储的映射。对上层（GIS 应用）的接口，提供完整的空间对象管理功能，统一管理空间属性和用户属性。对下层（数据库）的接口，则将空间属性与用户属性分开处理，空间属性的管理功能由包装层和 DBMS 共同实现，用户属性的管理功能则由 DBMS 直接实现。

包装层的主要任务就是分析并执行空间对象访问命令；为 GIS 应用提供一个一致且稳定的接口。

- 评价

由于支持面向对象方法，包装层为空间对象提供了统一的访问接口；同时包装层可以实现 DBMS 不能完成或 DBMS 能完成但效率不高的部分。但每次数据访问中都加入了中间层的开销。

4. 结束语

策略一、二是数据库引入 GIS 的一种尝试，还不够完善。一个重要问题就在于对空间属性存储粒度的划分上。另外关系数据库不能直观地提供空间对象的存储管理，还需要 GIS 应用提供一部分数据管理功能，如空间属性查询与分析。

策略三、四明显由于前两个方案，且各具特点。前者立足于数据库开发者的角度，考虑用 DBMS 完成 GIS 的数据管理功能。而后者立足于 GIS 应用开发者的角度，考虑产品可用性以及效率问题。前者还处于完善阶段，代表了未来 GIS 数据管理的发展方向。而后者则

较为成熟，代表了当前 GIS 数据管理的有效方法。

经过对比可以看出，包装模型是一种以性能换功能的方法，是权宜之计，但它在当前应用中相当有效。其特点也是很鲜明的：GIS 应用接口稳定，使得基于它的 GIS 应用也具有较稳定的稳定性；易于剪裁，采用构件方式实现可以进行多层次的剪裁，以取得良好的性能/功能搭配；数据库接口通用，不依赖于特定的数据库产品；易于扩展，可以针对效率问题在包装层提供一定的优化，如智能 cache 技术、空间索引过滤和查询优化，可以针对领域应用扩展功能，还可以由此实现 Distributed GIS 和 WebGIS。当然，引入包装层会增加数据操作的开销，而且构造并维护一个好的包装层也需要大量的人力物力投入。

总之，包装模型不失为目前数据库在 GIS 中应用策略的最佳选择之一。而事实也是这样的，国内外的 GIS 厂商的产品通常都采用包装模型策略与数据库结合。

参 考 文 献

- 【1】David M. Mark, Nicholas Chrisman, Andrew U. Frank etc., The GIS History Project, http://www.geog.buffalo.edu/ncgia/gishist/bar_harbor.html
- 【2】陈俊、宫鹏，《实用地理信息系统——成功地理信息系统的建设与管理》，科学出版社，1998
- 【3】萨师煊、王珊，《数据库系统概论》，高等教育出版社，1991
- 【4】Michael Stonebraker, Dorothy Moore 著，杨冬青、唐世渭、裴芳等译，《对象关系数据库管理系统——下一个浪潮》，北京大学出版社，1997
- 【5】Judith R. Davis, 《IBM's DB2 Spatial Extender: Managing Geo-Spatial Information Within The DBMS》，IBM Corporation's white paper, 1998
- 【6】Oracle8 Spatial Cartridge User's Guide and Reference, Release 8.0.3

作者简介：

丁力，男，硕士生，1975 年 12 月生。主要研究方向为地理信息系统、分布式计算、软件 Agent 等。

汪小林，男，博士生，1972 年 11 月生。主要研究方向为地理信息系统、分布式计算、软件 Agent 等。

许卓群，男，教授，博士导师，1927 年出生。主要研究方向为地理信息系统、分布式计算、并行计算、人工智能等。

Survey on Database Usage in GIS

Ding Li, Wang Xiaolin, Xu Zhuoqun

(Department of Computer Science, Peking University, Beijing, 100871)

Abstract: Manipulating huge data is an important goal of Geographical Information System (GIS). Database is best for Data Manipulation in GIS. This paper issues two vital problems, representation of spatial data and system efficiency. Then it analysis four solutions and points out the Wrapped Model, one of effective solutions, is better by comprehensive comparation.

Keywords: GIS, ORDB, Representation, Efficiency