

# 空间数据建库研究

胡圣武, 张光胜, 王宏涛

(河南理工大学 测绘学院, 河南 焦作 454000)

**摘要:** 采用空间数据库的基本理论, 结合北京市国土资源与房屋管理信息系统中建库的实例, 研究了空间数据库的数据性质, 空间数据库建库的原则、标准和步骤; 探讨了空间数据库的组织管理; 分析了空间数据库数据的无缝集成、元数据的组织、空间数据库增量备份和增量更新等关键技术。结果表明: 空间数据库还存在空间数据的组织模式、空间数据的安全性、空间数据的快速更新等问题, 还需要进行大力研究。

**关键词:** 空间数据; 数据库; 标准; 管理

中图分类号: P208 文献标志码: A 文章编号: 1672-6561(2007)02-0199-06

## Database Establishment of Spatial Data

HU Sheng-wu, ZHANG Guang-sheng, WANG Hong-tao

(School of Surveying and Geodesy, Henan Polytechnic University, Jiaozuo 454000, Henan, China)

**Abstract** The paper researches data property, principle, standard and step of spatial database and discusses framework management of spatial database and analyses key technology that are seamless integration of spatial database, metadata structure, spatial database copy and spatial data update adopting spatial database basic theory combination instance about establishing of Beijing land source and house management information system. The result indicates that it exists many problems in building spatial database that are spatial data structure mode, spatial data security and spatial data updating, therefore, it needs vigorous research.

**Key words:** spatial data; database; standard; management

## 0 引言

空间数据库建库是建立 GIS 的一个重要环节, 建库的好坏直接影响到 GIS 的功能, 因此, 空间数据库建库是关系到 GIS 成败的关键。笔者就空间数据库建库的基本理论和主要问题, 结合北京市国土资源与房屋管理信息系统的实例进行研究。

## 1 空间数据库管理方式<sup>[1]</sup>

### 1.1 集中式管理

所有空间数据集中存放于中心数据库, 并从全局数据关系统一考虑设计的数据库。

目前, 空间数据库大部分按集中式管理, GIS

应实现以下要求:

(1) 系统应实施数据统一管理。  
(2) 系统应能方便地实现各部门数据向本系统的转移。

(3) 采用关系数据库或对象关系数据库管理空间数据和属性数据, 确保空间和非空间数据的一体化集成; 实现真正的图文一体化集成; 实现空间数据的“无缝”组织。

(4) 需要解决历史数据管理问题。北京市国土资源与房屋管理信息系统的空间数据库就是采用集中式管理。

### 1.2 分布式管理

各空间数据有自己独立的数据库服务器, 完全

收稿日期: 2006-07-21

基金项目: 国家自然科学基金项目(40474003)

作者简介: 胡圣武(1970-), 男, 湖南津市人, 副教授, 博士, 从事 GIS 基础理论和图像处理技术研究。E-mail: hushengwuzhu@163.com

自治。

## 2 逻辑设计

### 2.1 空间数据库内数据的性质

一般 GIS 系统最终的数据应该包括空间数据、非空间型业务数据和管理维护型数据。

#### 2.1.1 空间数据

空间数据也是具有公共地理定位基础的数据<sup>[1]</sup>。如在北京市国土资源与房屋管理信息系统框架定义内的空间数据, 主要包括两大部分: 一部分是为国土房管局各部门所共享的基础空间数据, 由于本局的地籍图数据要素齐全, 基础地形图的参考功能相对弱化, 但考虑到国土房管局在北京市的测绘管理主导作用, 这部分数据必须建设完整; 另一部分是为各系统应用部门业务应用的空间数据, 比如地籍图、房产图、分户图、土地利用规划图等。基础空间数据可看作整个系统的空间定位核心。另外, 业务型空间数据的存在也是必然的, 尽管这些数据跟基础空间数据息息相关, 但在具体的国土房管业务应用中, 参与直接应用的不是基础空间数据, 而是业务型空间数据。这时候, 业务型空间数据是作为应用分析的核心而存在。由此可见, 系统内存在一个从基础空间数据到业务型空间数据再到业务应用分析的一个层层递进的支持关系, 三者缺一不可。

#### 2.1.2 非空间型业务数据

非空间型业务数据从数据性质上体现为属性数据, 内容包括表格型数据、文档数据和多媒体数据。按逻辑结构可分为结构化数据和非结构化数据。结构化数据主要指有一定结构、可以划分出固定的基本组成要素、以表格形式表达的数据, 可用关系数据库的表、视图表示, 如各种申请表、审批表等; 非结构化数据是指没有明显结构、无法划分出固定的基本组成元素的数据, 主要是一些文档、多媒体数据, 如申请材料、各种文件、法规等。

#### 2.1.3 管理维护型数据

管理维护型数据的产生主要是从空间数据库的稳定性和可维护性上考虑的, 以往的系统在维护方面往往依赖于大量的文档说明和标准引述。

管理维护型数据包括元数据、分类编码数据、生僻汉字和各类标准等。元数据即数据的数据, 也就是对现有数据的解释、定义和描述, 比如对数据库中表结构的说明、字段名的说明和空间数据的分层方式等。在海量的数据存储中, 元数据对数据库

的管理和维护提供了清晰的结构。元数据不是自然存在的, 是从现有数据中提取出来的。

### 2.2 关键数据的组织模式

#### 2.2.1 GIS 数据管理

地形图、地籍图等 GIS 数据, 由大量空间对象组成, 以往 GIS 数据管理软件均以文件形式存储, 空间对象被人为分割, 而实际上空间数据在物理上是边缘无缝的。近年来, GIS 数据管理普遍开始向通用大型关系数据库靠拢, 这样可以充分利用 RDBMS 已有的数据管理功能实现海量空间数据存贮与管理、事务处理(Transaction)、记录锁定、并发控制和数据仓库等功能, 利用扩展的 SQL 语言对空间与非空间数据进行操作, 同时可以方便地实现长事务和版本管理。尤其使空间数据与非空间数据得以集成在统一的数据平台, 从而促使 GIS 应用与一般应用的无缝集成。

#### 2.2.2 CAD 图形数据管理

利用关系数据库管理空间数据的关键在于面向对象的空间数据模型的采用。面向对象的空间数据模型的采用改变了原有 GIS 中图形与属性分离的概念, 反映空间对象的几何图形数据只是作为一个属性字段与其他非空间属性存贮于关系数据库表的一行中。这种数据模型可以方便地定义空间对象之间、空间对象与非空间对象之间的关联关系和规则, 能更好地对现实世界建模。

#### 2.2.3 文档数据管理

传统的文件系统主要基于操作系统提供的功能, 不适于在网络环境尤其是 Internet 环境下的文件管理与访问, 缺乏内容管理(Content Management)功能, 文件访问安全控制比较困难, 文件管理与备份容易出错。多数主流数据库管理平台都提供了将一个文件字段的形式存贮在关系表(Table)的一个记录中, 关系表中的其他字段可以用于描述文档的属性, 利用关系数据库提供的功能解决了传统文件系统存在的问题, 并且扩展了文件系统的功能。

### 2.3 数据库规划

数据库规划的目的是将国土资源、矿产资源、房地产和测绘管理中的纷繁复杂的信息进行高度概括, 分类管理, 提高数据的清晰度。

北京市国土资源和房屋管理信息化数据库包括:

#### 2.3.1 土地基础信息数据库

(1) 土地利用(现状)数据库。

- (2)土地利用规划数据库。
- (3)地籍数据库。
- (4)地价数据库。
- (5)农用地分等定级数据库。
- (6)建设项目用地数据库。
- (7)土地开发、复垦与整理数据库。

### 2.3.2 矿产资源与地质环境基础信息数据库

- (1)矿业权管理数据库。
- (2)矿产资源储量数据库。
- (3)矿产资源规划数据库。
- (4)矿产资源开发利用数据库。
- (5)地质环境管理数据库。

### 2.3.3 基础地理信息数据库

- (1)房屋基础信息数据库。
- (2)其他基础信息数据库。

## 3 空间数据库建库

数据库是一切处理的基础,建库工作在整个系统建设中是一项技术要求高、难度大的基础环节,良好的数据库设计与建库才能够保障数据的合理有效性。

### 3.1 建库原则<sup>[2]</sup>

#### 3.1.1 空间数据库设计原则

(1)数据冗余度小,共享程度高。充分利用存储空间,减小投入。

(2)数据独立性强,使应用子系统对数据的存储结构与存取方法有较强的适应性。

(3)设计结果符合各项规范指标要求。

#### 3.1.2 空间数据库保护原则

建立数据库的保护机制,用以负责阻止一切物理破坏和读写破坏,并能以最快的速度使其恢复工作,是数据库建设与使用顺利实施的必要条件。

(1)完整性原则。通过实时监控数据库事务(主要是更新和删除操作)的执行,来保证数据项之间的结构不受破坏,使存储在数据库中的数据正确、有效。

(2)并发性原则。当多个用户程序并发存取同一个数据块时,应对并行操作就需要进行控制,从而保持数据库数据的一致性,例如不致因为多名用户同时调阅某户籍资料并进行编辑而产生该数据资料的歧义。

(3)安全性原则。通过检查上机权限对局各处室及下属单位的不同级别数据库用户进行数据访

问与存取控制来保障数据库的安全与机密。

#### 3.1.3 数据库备份原则

根据各个数据库的实际需要定期或实时对数据进行本地与异地备份,同时备份应与更新同步。

### 3.2 建库标准

#### 3.2.1 数据标准

- (1)定位标准。
- (2)数据分类标准。
- (3)编码体系和代码标准。
- (4)各数据库与文件命名标准。
- (5)元数据标准。
- (6)符号标准。
- (7)数据格式与交换标准。
- (8)数据质量标准。
- (9)数据处理标准。

#### 3.2.2 操作标准

- (1)建库作业流程与技术规定。
- (2)建设验收标准。

### 3.3 数据库建设步骤

#### 3.3.1 空间数据库建库计划

数据库建设是一项系统工程,每一步都关系到以后系统运行状况的好坏,所以在建设数据库之前,每一步的工作内容及参数都必须考虑周到并形成文档,作为数据库建设的指南。

#### 3.3.2 数据建模

数据的种类繁多且彼此之间的关系错综复杂,如果之间的关系没完全理顺,对于以后的系统都是一种灾难。数据建模是对所有数据整理熟悉的过程,也是数据库设计中表结构定义的基础,所以数据建模是关系到系统生命力的重要因素。

#### 3.3.3 数据库设计

在北京市国土资源与房屋管理信息系统中就完成基础空间信息数据库等数据库的数据模型设计、数据的逻辑结构以及数据的物理结构设计,确定基础空间信息数据的分层方案与编码方案。

#### 3.3.4 制定数据库建设的各项指标、规范与建库指导书

针对上述数据库设计的内容,在保证入库图形数据经过严格接边与平差、具备统一的拓扑关系以及实现属性数据与图形数据无缝连接的前提下,制定源数据获取方案、数字化作业流程与数字化质量要求书,包括确定从其他有关部门获得的数字化数据是否需要同构转换和数据通过怎样的输入

方式与技术指标约束入库模拟等工作方式。

### 3.3.5 空间数据入库

这是一个费时的工 作, 且是一个重复性劳动, 需要具体工作人员有耐心和兢兢业业的工作精神。

### 3.3.6 技术监督与验收

在建库过程中, 由技术人员配合完成数据监督与查错, 从而监督数据入库操作、数字化作业的规范性, 并对拟入库数据依照验收标准执行检查。

## 3.4 空间数据库迁移方案

由于技术的更新和硬件设施的更新换代非常快, 因而存在现有系统运行环境向新系统运行环境转换、低版本数据库向高版本数据库转换以及两个不同数据库之间进行转换时, 数据库中的数据(包括结构定义)需要被转移并使之正常运行, 这就是数据库中的数据迁移。对于结构定义相同的同种数据库管理系统不同版本之间的升级, 这种迁移比较简单。但对于不同的数据库平台之间的转换, 更重要的是不同结构定义之间的转换, 二范式向三范式的转换, 各种约束的重定义等, 迁移工作的复杂度就大大增加了。

对数据库管理人员来说, 数据库数据迁移极具挑战性, 一旦措施不当, 珍贵的数据资源将面临丢失的危险。要成功地实现数据库数据平滑迁移, 需要周密计划和充分准备, 并按照一定的步骤来完成。空间数据库迁移方案为:

### 3.4.1 设计数据迁移方案

设计数据迁移方案主要包括研究原有数据的结构、来源、数据项定义、取值等现状, 研究新旧数据库结构的差异, 评估和选择数据迁移的软硬件平台、选择数据迁移方法、选择数据备份和恢复策略、设计数据迁移和测试方案等。

### 3.4.2 编制数据迁移程序

数据迁移是把数据从原有系统迁移到新系统中, 这中间存在着数据格式和存储结构的不同, 所以, 需要一系列程序辅助迁移, 尽量减少人工参与, 避免人为的主观性错误。

### 3.4.3 进行数据模拟迁移

根据设计的数据迁移方案, 建立一个模拟的数据迁移环境, 它既能仿真实际环境又不影响实际数据, 然后在数据模拟迁移环境中测试数据迁移的效果。数据模拟迁移前也应按备份策略备份模拟数据, 以便数据迁移后能按恢复策略进行恢复测试。

### 3.4.4 数据模拟迁移

根据设计的数据迁移测试方案测试数据模拟

迁移, 也就是检查数据模拟迁移后数据和应用软件是否正常, 主要包括数据一致性测试、应用软件执行功能测试、性能测试、数据备份和恢复测试等。

### 3.4.5 准备实施数据迁移

数据模拟迁移测试成功后, 在正式实施数据迁移前还需要做好完全数据备份、确定数据迁移方案以及安装和配置软硬件等工作。

### 3.4.6 正式实施数据迁移

按照确定的数据迁移方案正式实施数据迁移。

### 3.4.7 测试数据迁移效果

按照数据迁移测试方案测试数据迁移效果, 并对数据迁移后的数据库参数和性能进行调整, 使之满足数据迁移后实际应用系统的需要。

### 3.4.8 移植系统应用软件

将实际应用系统的应用软件移植到数据迁移后的数据库系统上并使之正常运行。

### 3.4.9 运行应用系统

在正式实施数据迁移成功并且数据库参数和性能达到要求后, 就可以正式运行应用系统, 并投入实际使用。

## 4 空间数据库的组织管理

一般而言, 对于空间数据库采用 SDE 来管理空间数据, SDE 具有数据管理上的突出优点<sup>[3-5]</sup>:

(1)数据录入和编辑更加精确。数据的录入和编辑可借助数据库系统功能, 所以可以减少差错。

(2)理想的空间对象模型。地理特征如饭店位置、旅游路线、度假区等, 被作为空间对象, SDE 在描述这些对象时采用了明晰的特征(属性)和行为(方法), 使表达执行具备灵活性。地理特征通过图层这种空间连续策略进行索引, 促进了快速恢复操作, 提高数据管理效率。

(3)支持大型数据库。SDE 利用统一的数据模型, 维护关系数据库中的空间和属性数据, 管理近乎无限的空间特征, 如全国范围的土地利用现状和历史数据。

(4)快速实现过程。对复杂的空间查询, 使用互操作处理的客户机/服务器模式在网络上得以实现, 客户机与服务器共同完成这一工作。客户机主要是响应空间分析操作, 服务器则进行数据搜索和检索。这种互操作处理方法使得动态空间叠加成为可能, 当大量增加客户机的时候, 利用对称多处理结构或调整计算机缓冲区大小, 可以把客户机带

来的性能下降到最小。

(5)与已有的应用集成。用SDE所提供的API接口,不借助传统的GIS技术即可将空间数据及其分析嵌入到自己的应用中去。

(6)特征组是连续的。在关系数据库上存储空间数据,可以实现存储非常巨大的特征组而不需要其他的空间分区或块(tiles)。

(7)对地理数据的开放式系统访问。许多用户能同时编辑地理数据,地理数据库数据模型支持许多人能在本地区域编辑特征。

(8)数据访问可以通过属性取舍或空间范围切割。网络传输的仅是需要的数据,而不是全体数据,所以系统数据访问效率可以提高。

## 5 关键技术

### 5.1 数据的无缝集成

在GIS应用数据库设计中,只有实现了数据的无缝集成,才能为功能集成打下基础。GIS数据集成类型分空间数据的无缝集成和空间数据与属性数据的无缝集成。

#### 5.1.1 空间数据的无缝集成

传统的空间数据都是基于图幅的,每一图幅以文件的方式存放起来。由于数据生产的系统误差或偶然误差,在图幅与图幅之间往往存在数据不一致的情况,如河流在图幅之间发生了错位、房屋在图幅之间不能闭合等现象,这种情况称为“图幅缝隙”。由于分幅生产的原理,这种“图幅缝隙”无法避免,以往的处理方法是增加一道接边的工序,这就造成了数据被人为的修改,增大了误差。可以设想,如果在整个数字化区域内进行数据生产,就可避免这一问题。也就是说,一个跨图幅的特征如河流,如果一次性录入,就可以保证其一致性。解决了“图幅缝隙”问题,事实上也使得本来属于同一地物的分段地物还原了其本来面目。具体实施步骤为:

(1)详细拟定空间数据库结构,提出可操作的数据库建库方案。

(2)按照数据库建库方案对原始数据进行规范化整理。不管是何种格式的数据,都要进行统一的接边、查错、分层。

(3)对于有属性表的数据如MapInfo、Arc/Info数据,进行属性整理;没有属性的数据如AutoCAD数据,单独在数据库里建立属性表。

(4)对已整理完成的空间数据进行批量入库。

(5)对属性数据进行批量入库。

(6)进行图形归一化处理,保证跨图幅的特征地物保持连续。

(7)进行属性连接。

对于空间数据的连续无缝,其关键步骤是(2)、(6)。通过步骤(2),可从视图角度保证了空间数据的连续;通过步骤(6),保证了真正的连续无缝。

#### 5.1.2 空间数据和属性数据的无缝集成

空间数据和属性数据的无缝集成。对北京市国土资源与房屋管理信息系统而言,在国土和房地产业务上是如何将宗地或房产的图形表示与其相关的产权信息高度关联起来,实现图文一体化查询的问题。对于国土管理业务,如红线划拨,比较常见的处理方法都能够将宗地图形和业务数据连接在一起。对于这种情况,在数据入库时只需对图形和属性进行相应的入库对照即可。在房产管理中,比较常见的处理是将房产平面图存成AutoCAD格式文件,相应的产权信息存放在关系数据库中,通过AutoCAD格式文件文件名进行关联。这种处理方式带来以下问题:

(1)共享问题。多用户对同一数据文件进行编辑时不受保护,造成文件损坏,数据丢失。

(2)操作问题。用户在需要了解某房产周围的情况时,必须了解其他相邻房产的编号才可以调图,使用起来极不方便。

(3)维护问题。文件方式给数据管理人员带来了极大的数据维护工作量。

鉴于以上问题,在北京市国土资源与房屋管理信息系统的数据库建设中,将CAD数据以二进制大对象的形式存贮于对象关系数据库中的行(ROW)中一个列(Column)中,行中的其他列存放房屋分层平面图和分户平面图的属性,构成一个完整房屋楼层实体(或对象)和房屋单元实体(或对象)。

### 5.2 元数据的组织

元数据是“关于数据的数据”,元数据为各种形态的数字化信息单元和资源集合提供规范、普遍的描述方法和检索工具,为分布的、由多种数字化资源有机构成的信息体系提供整合的工具与纽带。离开元数据的信息系统将是一盘散沙,无法提供有效地检索和处理,如北京市国土资源和房屋管理系统涉及的空间信息及其他房地管理信息源较多,必须建立统一的元数据库,提高系统检索和处理的速

度。元数据主要作用为:

(1)确认和检索。建立国土资源和房屋管理系统信息元数据库,可以帮助人们检索和确认所需要的信息资源,数据元素往往限于作者、标题、主题、位置、生产者等简单信息。

(2)著录描述。用于对数据单元进行详细、全面地著录描述,数据元素囊括内容、载体、位置与获取方式、制作与利用方法、甚至相关数据单元方面等。

元数据管理应结合元数据的特点,在参考一定标准的基础上,开发方便实用的元数据操作工具。元数据操作工具包括以下功能:

(1)元数据输入功能。针对各元数据子集的内容特点,为数据生产者提供一系列方便的元数据输入界面,便于数据生产者建立元数据体系。

(2)元数据编辑功能。元数据编辑包括元数据修改、记录的增加、记录的删除等操作。该功能可以反映空间数据库的更新,实现元数据的扩展。

(3)查询、检索功能。系统提供方便的菜单和按钮,使用户能快捷地浏览自己所关心的元数据信息。同时系统提供关键字查询、条件查询等方式,帮助用户定位元数据信息。

(4)元数据合并与导入。该功能实际获得其他元数据信息数据库,并从其元数据信息表中导入元数据信息,便于元数据的扩展。

(5)元数据报表输出功能。系统为用户提供报表打印输出功能,用户可以选择自己感兴趣的元数据信息,按自己要求的格式输出。

### 5.3 空间数据库增量备份和增量更新

#### 5.3.1 空间数据库的备份机制

数据库中的数据是独立于程序而存在的,无论是自然错误还是人为错误,都可能造成数据的错误,为了能够恢复改前的状态和值,数据库的操作要具有以下功能:

(1)恢复。在出错时可回到修改前状态。

(2)备份。数据库修改后,原数据应该定期备

份,这种备份又包括完全备份和增量式备份。

(3)历史数据。当数据库中的数据修改后,要存储数据的历史信息,供以后用。

基础空间数据库的备份机制完全采用大型关系数据库管理软件提供的备份机制。以空间数据库为备份的对象,短期对数据库做联机增量备份;在长周期做脱机全库备份。

#### 5.3.2 空间数据库的更新机制

空间数据库的更新就是依据规定区域内的现状,修正信息载体上相应要素的内容,以提高其精度和保持现势性的一项重要工作<sup>[6-7]</sup>。数据库的更新有方式为:

(1)同步更新。保持非 GIS 数据和 GIS 数据的同步更新。

(2)批量更新。通过基本管理系统自身的导入导出功能完成系统的更新。

## 6 结语

就空间数据库建库的基本问题进行了研究,不同的空间数据建库会有不同要求和不同问题,笔者只对空间数据建库的一般问题进行了探讨。由于空间数据建库是一个随着时代发展而发展的技术,需要不断地探索和研究。

参考文献:

- [1] 萨师煊,王 珊.数据库系统概论[M].北京:高等教育出版社,2002.
- [2] 陈述彭,鲁学军,周成虎.地理信息系统概论[M].北京:科学出版社,2000.
- [3] 刘云生.现代数据库技术[M].北京:国防工业出版社,2001.
- [4] 岳小平,鞠时光,李 芷.空间数据索引技术[J].计算机应用研究,2002,24(5):89-92.
- [5] 刘 晨,李海英.空间、时序、时空数据库索引技术分析[J].研究与开发,2002,7(8):56-60.
- [6] 吴金华,祝国瑞.空间数据仓库的认知过程[J].地球科学与环境学报,2004,26(4):67-70.
- [7] 邹逸江,吴金华.空间数据仓库的结构设计[J].地球科学与环境学报,2003,25(1):66-69.