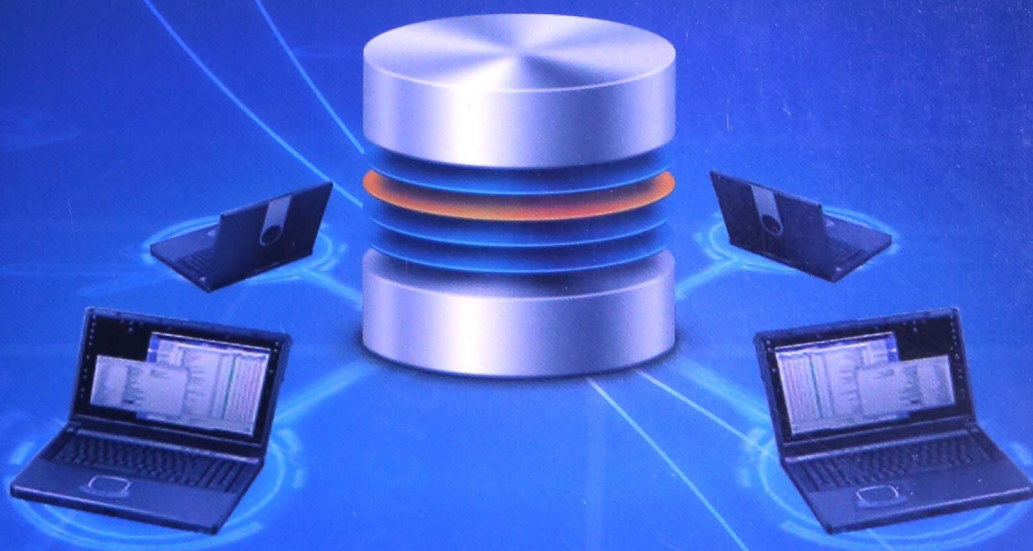


分布式空间数据 集成与查询优化技术

Distributed Spatial Data Integration and
Query Optimization Technology

朱欣焰 陈静 向隆刚 周春辉 陈迪 著



测绘出版社

分布式空间数据集成与 查询优化技术

Distributed Spatial Data Integration and
Query Optimization Technology

朱欣焰 陈静 向隆刚 周春辉 陈迪 著

测绘出版社

· 北京 ·

© 朱欣焰 2013

所有权利(含信息网络传播权)保留,未经许可,不得以任何方式使用。

内 容 简 介

分布式空间数据库系统研究内容涉及地理空间数据模型、异构空间数据集成、空间索引和高性能计算等方面。本书在阐述分布式空间数据库系统相关概念的基础上,系统地介绍了分布式空间数据库的组成和分布式空间数据的组织与管理,讨论了分布式空间数据库的集成管理方法和异构栅格数据的无缝可视化集成方法,重点探索了基于分割分片的空间片段连接优化方法及分布式空间无缝处理算法,完成了分布式异构空间数据库原型系统的设计与开发并进行相关的实验验证和分析。

本书系统地总结了作者在该领域近几年的研究成果,内容新颖,理论性强,可作为地理信息及相关专业的研究人员、高校教师、研究生及高年级本科生的参考用书。

图书在版编目(CIP)数据

分布式空间数据集成与查询优化技术/朱欣焰等著. —北京:测绘出版社, 2013.11

ISBN 978-7-5030-3222-6

I. ①分… II. ①朱… III. ①地理信息系统—分布式
数据处理 IV. ①P208.2 TP274

中国版本图书馆 CIP 数据核字(2013)第 281543 号

责任编辑	赵福生	封面设计	李 伟	责任校对	董玉珍	责任印制	喻 迅
出版发行	测绘出版社	电 话	010-83543956(发行部)				
地 址	北京市西城区三里河路 50 号		010-68531609(门市部)				
邮政编码	100045		010-68531363(编辑部)				
电子信箱	smp@sinomaps.com	网 址	www.chinasmp.com				
印 刷	三河市世纪兴源印刷有限公司	经 销	新华书店				
成品规格	169mm×239mm						
印 张	13.5	字 数	260 千字				
版 次	2013 年 11 月第 1 版	印 次	2013 年 11 月第 1 次印刷				
印 数	0001-1500	定 价	46.00 元				

书 号 ISBN 978-7-5030-3222-6/P·675

本书如有印装质量问题,请与我社门市部联系调换。

前 言

地理空间数据具有多时空、多尺度、多语义、多源性等特征,有效地表达和组织管理异构空间数据一直都是地理信息系统(GIS)研究的核心课题。作为地理信息系统的基础,可以说地理信息系统中的绝大多数技术问题都是由地理空间数据引发的。由于目前 GIS 广泛应用于各领域和部门,空间数据格式种类繁多且更新较快,数据量巨大,有基础地理数据、各种土地专题数据、矿产资源数据等,因此,采用传统的集中管理方式存在较大难度和限制。空间数据在生产、管理、维护、应用服务等方面的分布特征,决定了空间数据管理必然向分布式方向发展。不同数据生产者和使用者之间要建立合理的管理、维护与共享机制,因此需要分布式空间数据存储和管理能力。

近年来,随着计算机网络技术和信息高速公路建设的飞速发展和应用,网络技术为数据分布存储提供了基础,它可以将异地配置的若干空间数据存储站点连接起来。但不同地方、不同类型的数据采用了不同的 GIS 软件建库,导致数据之间不能互相访问,产生了许多信息孤岛。在这种情况下,就要求数据服务具备极强的分布式和异构化特性,即数据可以分散于多个服务器,而客户端的访问能够跨越异构网络、异构操作系统平台、异构数据库平台。另外,由于空间数据服务本身具有时空变化的特点,如果要提供快速、全面的空间分析和空间图形界面,就要求整个系统具有足够高的性能。

基于网络的分布式 GIS 已成为当前研究的热点问题。各 GIS 厂商也相继推出了自己的分布式 GIS 产品,为实现开放式、分布式 GIS 提供了基本的环境。分布式 GIS 是小到政府办公、企业部门生产管理,大到国家空间信息基础设施(national spatial information infrastructure, NSII)、全球空间数据基础设施(global spatial data infrastructure, GSDI)和数字地球(digital earth, DE)等一切信息系统的支撑技术。分布式地理信息系统的研究内容涉及地理空间数据模型、空间索引、空间元数据、高性能计算等方面。本书在简要介绍分布式空间数据库基本概念的基础上,重点阐述了作者近年来在分布式空间数据库无缝和查询优化等方面的研究成果,提出了相关的基础理论,希望本书的出版能进一步丰富分布式空间数据库理论,为分布式空间数据服务集成与应用的深化提供理论和方法上的支持。

全书共有 7 章,主要内容如下:第 1 章为绪论,主要介绍了地理空间数据的基本特征及在空间数据组织和管理中存在的无缝问题。第 2 章主要在介绍分布式数据库系统的基础上,讨论分布式空间数据库的分片、分布,以及分布式空间数据库

存在的跨边界有缝问题。第3章主要针对异构空间数据的冲突,讨论分布式空间数据模式集成方法。第4章主要介绍了分布式异构空间栅格数据的集成方法,包括栅格分布式不规则三角网(TIN)与规则格网(GRID)无缝可视化集成、分布式矢量和栅格数据集成可视化。第5章主要介绍了分布式空间数据库跨边界查询的实现原理,对分布式空间片段连接进行分类研究,重点讨论基于分割分片的跨边界空间片段连接的优化问题。第6章在片段连接优化转化规则的基础上,解决如何实现跨边界透明无缝高效的查询问题。第7章主要建立跨边界连接优化的异构分布式空间无缝查询处理实验系统,进行算法验证及实验结果分析。

本书主要内容在研究和出版过程中得到了国家自然科学基金项目(40971232, 41271401)和国家863计划(2007AA12Z201)的支持。本书第1、2、5章由朱欣焰撰写,第3章由向隆刚撰写,第4章由陈静撰写,第6章由朱欣焰、周春辉撰写,第7章由周春辉、陈迪撰写。在本书的撰写过程中,一批富有朝气的青年学者为此做出了积极的贡献,无论是在资料的收集、整理中,还是在具体的实验系统实施中,都有他们的身影,包括芮维博士后、夏宇博士、苏科华副教授、黄家凯硕士、付雷硕士等。在此,对他们表示衷心的感谢!

分布式GIS是地理信息系统发展的一个重要分支,也是必然趋势,希望本书的出版能起到抛砖引玉的作用。由于时间仓促,相关理论和技术还在不断地完善和更新中,书中难免存在疏漏和不足之处,恳请读者批评指正。

作 者

2013年7月

目 录

第 1 章 绪 论	1
§ 1.1 地理信息系统与空间数据库	1
§ 1.2 空间数据无缝组织与管理	3
§ 1.3 分布式空间数据库	11
第 2 章 分布式空间数据库组成	18
§ 2.1 分布式数据库系统	18
§ 2.2 分布式空间数据库系统	25
§ 2.3 空间数据分片与分布	33
§ 2.4 分布式空间数据库系统跨边界有缝问题分析	45
第 3 章 分布式空间数据集成技术	53
§ 3.1 空间数据集成研究概述	53
§ 3.2 异构空间数据冲突分类与表达	60
§ 3.3 基于扩展 GML 的异构空间数据集成	65
第 4 章 分布式空间栅格数据集成可视化技术	69
§ 4.1 分布式空间栅格数据集成概述	69
§ 4.2 异构栅格数据多尺度集成可视化规则	70
§ 4.3 分布式异构栅格空间数据集成方法	73
§ 4.4 分布式 TIN 与 GRID 无缝可视化集成	78
§ 4.5 分布式矢量和栅格数据集成可视化	87
第 5 章 跨边界空间片段连接优化方法	91
§ 5.1 分布式空间数据库连接优化概述	91
§ 5.2 基于分割分片的空间片段连接分类	92
§ 5.3 空间片段间第一类连接优化	97
§ 5.4 空间片段间第二类连接优化	107
§ 5.5 空间片段间第三类连接优化	109
§ 5.6 空间片段间第四类连接优化	119

§ 5.7	空间片段间第五类连接优化	119
§ 5.8	空间片段间第六类连接优化	121
§ 5.9	基于分割分片的片段连接特点分析	126
第 6 章 分布式空间查询处理技术		129
§ 6.1	分布式空间数据库查询处理概述	129
§ 6.2	基于分割分片的分布式空间查询分解与数据本地化方法	131
§ 6.3	基于分割分片的分布式空间查询全局优化策略	142
§ 6.4	基于跨边界连接优化的分布式空间查询处理算法	150
§ 6.5	分布式异构空间数据库系统框架与无缝查询处理	165
§ 6.6	基于跨边界连接优化的分布式并行调度	174
第 7 章 实验系统构建与实验分析		181
§ 7.1	实验系统构建	181
§ 7.2	分布式空间查询处理实验与分析	192
参考文献		201

Contents

Chapter 1	Introduction	1
§ 1.1	Geographic information system and spatial database	1
§ 1.2	Seamless organization and management of spatial dataset	3
§ 1.3	Distributed spatial database	11
Chapter 2	The components of distributed spatial database	18
§ 2.1	Distributed database system	18
§ 2.2	Distributed spatial database system	25
§ 2.3	Spatial database fragmentation	33
§ 2.4	The cross-border problem in distributed spatial database	45
Chapter 3	Distributed spatial data integration technologies	53
§ 3.1	Overview of spatial data integration	53
§ 3.2	Classification and expression of the conflict between heterogenous spatial data	60
§ 3.3	Spatial data integration based on extended GML	65
Chapter 4	Visualization integration technologies of distributed spatial raster data	69
§ 4.1	Overview of distributed spatial raster data integration	69
§ 4.2	Visualization integration rules of heterogenous raster data	70
§ 4.3	Distributed heterogenous raster data integration method	73
§ 4.4	Seamless visualization integration for distributed TIN and GRID	78
§ 4.5	Seamless visualization intergration for distributed vector and raster data	87
Chapter 5	The join optimazation principle for cross-border spatial fragment	91
§ 5.1	Overview of distributed spatial database join optimization	91
§ 5.2	The classification of spatial join based on partitioning	

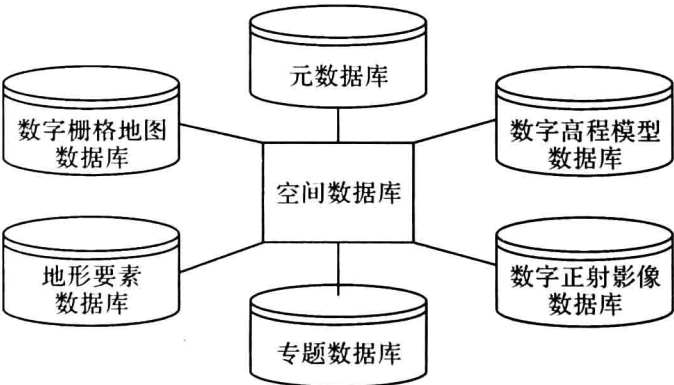
fragmentation	92
§ 5.3 The 1 st class spatial fragment joins	97
§ 5.4 The 2 nd class spatial fragment joins	107
§ 5.5 The 3 rd class spatial fragment joins	109
§ 5.6 The 4 th class spatial fragment joins	119
§ 5.7 The 5 th class spatial fragment joins	119
§ 5.8 The 6 th class spatial fragment joins	121
§ 5.9 Characteristic analysis of spatial join based on partitioning fragmentation	126
Chapter 6 Distributed spatial query processing technology	129
§ 6.1 Overview of distributed spatial database query processing	129
§ 6.2 Query decomposition and localization of distributed spatial queries based on partitioning fragmentation	131
§ 6.3 Global optimization of distributed spatial queries based on partitioning fragmentation	142
§ 6.4 Optimization algorithms of distributed spatial queries	150
§ 6.5 Architecture of distributed heterogenous spatial database	165
§ 6.6 Parallel scheduling based on cross-border fragment joins	174
Chapter 7 Experiment system and analysis	181
§ 7.1 Experiment system	181
§ 7.2 Result and analysis	192
References	201

第 1 章 绪 论

§ 1.1 地理信息系统与空间数据库

地理信息系统(geographic information system,GIS)是用于采集、存储、管理、分析和描述整个或部分地球表面(包括大气在内)与空间和地理分布有关数据的空间信息系统,是从地理学、测量学、制图学、遥感学、图形图像学引用电子计算机技术后自然形成的一门边缘科学(李德仁,1997),它的发展与其他相关学科的发展密切相关。自 20 世纪 60 年代开始发展,经历了 20 世纪 90 年代以来的快速发展时期,GIS 已经得到了长足的进步,特别是计算机和网络技术的飞速发展,GIS 已从原来的特定的应用领域,发展为与老百姓日常生活密切相关的技术,在国民经济各领域、政府和军事部门得到了广泛应用,在全球范围内形成了一个很大的产业。

地理信息系统的核心是空间数据库(龚健雅 等,2004)。空间数据库是 GIS 中空间数据的存储场所,数据库通常是数据库管理系统的简称,一个完整的空间数据库系统包括空间数据库、空间数据库管理系统和数据库应用系统三个组成部分(黄杏元 等,2001)。其中空间数据库指的是 GIS 在计算机物理存储介质上存储的与应用相关的地理空间数据的总和,可以以一系列特定结构的文件形式和数据库形式组织在存储介质上。空间数据库管理系统(SDBMS)则是指能够对物理介质上存储的地理空间数据进行逻辑和语义上定义、提供必要的空间数据库查询检索和存取功能,以及能够对空间数据库进行有效的维护和更新的一套软件系统。空间数据库库体总体构成如图 1.1 所示。



地理空间数据(空间数据)是指表征地理圈或地理环境固有要素或物质的数量、质量、分布特征、联系和规律的数字、文字、图像和图形等的总称。空间数据具有空间、专题(属性)、时间等基本特征(龚健雅,2001;陈俊等,1999)和非结构化、空间关系、分类编码、多尺度、海量性等特征。空间关系主要包括拓扑关系、方向关系和度量关系。空间数据类型主要包括:地图数据、影像数据、地形数据、属性数据和元数据。空间数据的表达主要有矢量表达和栅格表达两种,矢量表达又分拓扑数据模型和无拓扑关系的数据模型(面条数据模型)。

空间数据库与传统数据库的主要区别之一是:传统数据库系统管理的是不连续的、相关性较小的数字和字符数据,而空间数据库管理是连续的地理空间数据,具有很强的空间相关性(汤国安等,2000)。由于地球表面是连续的,表达地理空间信息的空间数据具有连续性,因此,空间数据库需要研究连续表达的问题,也就是无缝与跨边界问题。

到目前为止,对空间数据库还没有一个统一的普遍接受的定义,对空间数据库管理系统应该具备什么样的能力说法不一。Güting(1994)认为空间数据库系统首先是一个数据库系统;要提供对支持的数据模型进行空间数据类型定义的能力,提供支持空间数据类型及空间关系的操作查询语言;支持空间数据类型的实现,提供至少包括空间索引和有效的空间连接(spatial join)算法的能力。有效的空间索引是空间数据库系统必不可少的强制性要求,以便能够快速从大量数据中进行检索;还应该支持通过一定的空间关系对两个不同的类进行对象连接,并以最有效的方式对笛卡儿积进行过滤,以便提高效率;SDBMS 提供对 GIS 和其他应用的底层技术,提供基本的应用能力,不强调 SDBMS 直接可用于面向 GIS 的应用中。

空间数据库管理系统的实现可以建立在传统的数据库管理系统基础上,也可以从头开始开发。目前通常有两种实现方法:一种是直接对常规数据库管理系统进行功能扩展,加入空间数据存储与管理能力,如 Oracle Spatial;另一方法是在常规数据库管理系统上添加一层空间数据库引擎,以获得常规数据库管理系统之外的空间数据存储和管理的能力,如 ESRI 的 SDE(spatial database engine)、GeoStar SDE、SuperMap SDE 等。GeoStar SDE 在 Oracle、SQL Server 等关系数据库的基础上扩充了空间数据库管理的能力。Shekhar 等(2004)给出了一种在对象关系数据库管理系统(OR-DBMS)上搭建的 SDBMS 的体系结构,如图 1.2 所示。图中给出了扩充功能的主体的核心技术、DBMS 的接口和空间应用接口。常见的空间索引一般是自顶向下、逐级划分空间的空间索引,比较有代表性的包括格网空间索引、四叉树索引、BSP 树、KDB 树、R 树,而格网索引、四叉树索引、R 树索引最为常用。在空间应用接口中,Shekhar 等把可视化接口作为空间数据库的一部分。

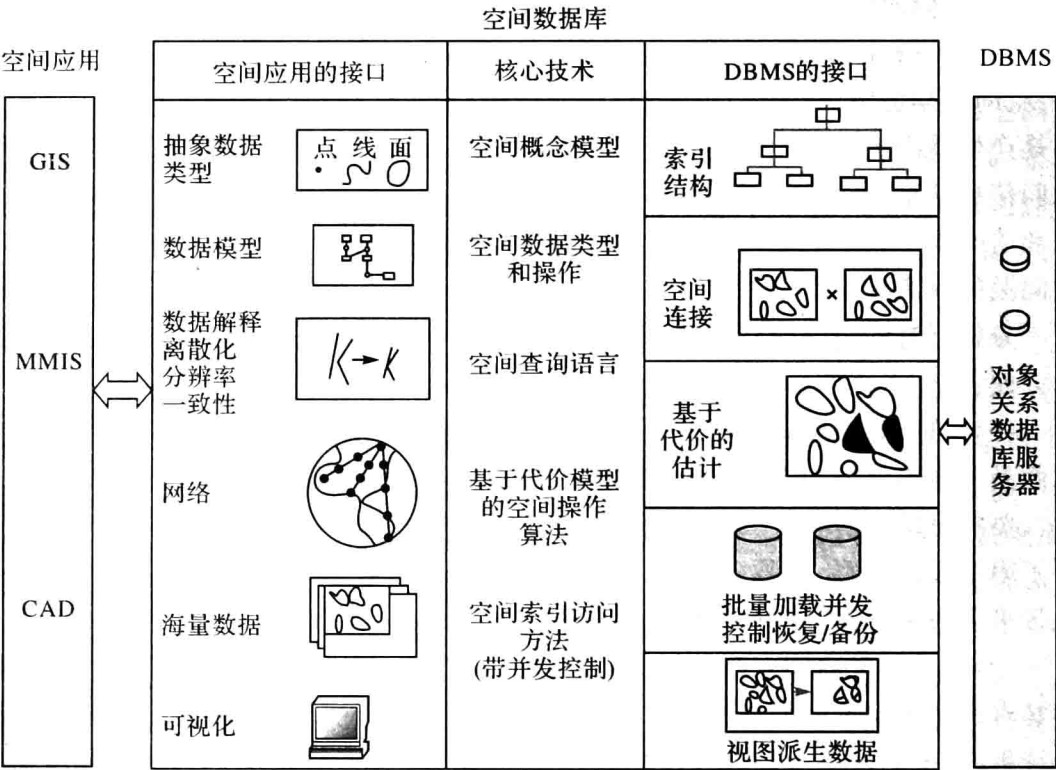


图 1.2 SDBMS 三层体系结构

实际上,空间数据管理系统的功能随着技术的发展也在发生变化,原来在应用层上实现的部分功能,可以下移到空间数据库管理层。图 1.2 的 SDBMS 三层体系结构中把可视化接口归入到空间数据库部分,就要求空间数据库系统向上一级应用透明地提供空间数据可视化表现能力,这为空间数据的一致性表示提供了很好的途径。

自从地理信息系统出现以来,围绕空间数据管理方法,空间数据库管理系统先后出现了几种不同的发展模式:文件型管理、文件-关系型管理、全关系型管理、对象-关系型管理、面向对象型管理(朱欣焰 等,2002)。目前空间数据管理主流方式是全关系型和对象-关系型,文件型、文件-关系型管理是不可缺少的方式。随着网络技术的发展,空间数据管理正在向分布式方向发展。

§ 1.2 空间数据无缝组织与管理

1.2.1 空间数据库有缝问题

地理空间的连续性及其相关性,决定了空间数据的管理必须解决无缝问题。

虽然对无缝概念目前还没有一个统一的定义,但无缝空间数据库和无缝 GIS 一直吸引大量的研究者进行研究。空间数据的无缝与空间数据的一致性往往密切相关:空间数据缝隙的产生常常来自空间数据的不一致,即空间数据库中数据的不一致是产生缝隙的主要原因之一(不一致是原因,缝隙是结果)。李德仁院士指出了空间信息共享面临的挑战:时间基准不一致、空间信息标准不一致、数据格式不一致和语义不一致的问题。空间数据无缝的最终解决还在很大程度上依赖这些不一致问题的解决。

分析已有的研究,空间数据库存在有缝问题主要原因大致可以归纳为下面几个方面:

(1)地球的椭球体与地图的平面表示的矛盾。地图是以平面为基础的,地球椭球面是一个不可展的曲面,地球在地图上的表示常常要把球面压成平面(投影变换),要想将球面与一个二维平面相对应必须要经过地图投影,必然会产生变形。为了限制投影变形,必须采用分带投影。分带投影的方式固然可以限制投影变形,但带来了投影不连续的问题,形成有缝(坐标系变换、投影变换有缝)。

(2)多比例尺之间空间数据的不连续。空间数据的多尺度特征要求空间数据库具有表达多尺度的能力。多比例尺空间数据库中,地物的表示是经过“综合”后的结果,而不是简单地缩放,所以在跨比例尺操作时(如多比例尺表现)空间数据之间会存在不连续,地物间产生有缝。

(3)空间数据采集与加工误差。空间数据的采集、加工,有不同的数据源、不同的仪器、不同的作业处理流程和不同的作业人员,几何上有误差,产生有缝,如地图的分幅采集造成图幅之间的缝隙。

(4)不同时间的空间数据之间不一致。不同时间采集加工的数据,由于时间差异,产生变化,造成新旧数据接合处的有缝。

(5)空间数据的语义差别。不同的系统、不同的空间数据模型,空间数据的语义存在差异,异构空间数据源集成时,存在语义差异,造成语义“有缝”。

(6)影像数据的色彩不连续。由于不同影像来自不同传感器、不同的成像条件,使得影像数据库中像素的色彩不连续、不均匀,形成 DOM 有缝。

(7)空间数据库中不同层间空间数据不一致。在分层表达的空间数据库中,不同层之间的空间,由于数据更新、不同时间、不同数据源等原因造成各层数据叠合时同一地物或不同地物间公共边界处存在有缝。

(8)空间数据的分布式存储与应用。存储在不同场地的分布式空间数据存在边界之间、层与层之间的几何有缝。由于空间数据几何及空间拓扑关系的相关性,要求对跨场地边界的空间查询与处理实现透明性访问,需要解决分布式空间数据库的无缝查询与处理问题。

对空间数据管理中如何实现这些无缝管理问题,已有不少可用的研究成果。

1.2.2 空间数据无缝组织与管理研究现状

空间数据库是 GIS 的核心,空间数据必须正确反映空间、时间、比例尺上的连续特性,而空间特征又包括几何、拓扑、方向和度量等关系。在讨论空间数据和 GIS 无缝时,现有研究中经常使用的术语有逻辑无缝、物理无缝、无缝组织、无缝影像库、无缝管理、无缝集成、无缝表达、色彩无缝、语义无缝、无缝浏览、无缝显示、几何无缝、尺度无缝、无缝拼接、无缝立体、无缝镶嵌等。

1. 矢量数据无缝组织与管理

在文件型、文件-关系型管理方式下,空间矢量数据一般按分层分幅进行组织和管理。在全关系型和对象-关系型方式下,仍然是分层管理,但不强调图幅概念。李爱勤等(1998)探讨了大型 GIS 地理数据库的无缝组织问题,采用纵向与横向地理数据库设计策略:纵向设计按不同比例尺的数据源建立不同层次的地理数据库;横向设计为同一精度数据源的大型地理数据库,按网络分布式无缝设计,无缝地理数据库按“工程+工作区”方式组织,提出了一种的改良 Morton 码空间索引——“分级扩散二叉树编码”,建立无缝地理数据库的空间索引机制。文中虽提到了分布式空间数据的无缝设计,但具体如何设计、实现及需要解决的问题还有待深入研究。分布式多数据库下的无缝 GIS 问题由 Laurini(1998)进行了比较系统的讨论,指出了要解决的一些冲突问题。在上述大型 GIS 地理数据库的无缝组织基础上,李爱勤(2001)进一步研究了纵向组织方法,即将自动地图综合功能或多比例尺表达与处理功能集成到 GIS 中,试图通过地图综合来实现多比例尺空间数据库,而不必建立同多个比例尺版本对应的空间数据库。当然,这种方法目前仍处在研究阶段,还没有达到实用程度。

刘纪平(1998)利用建立空间目标全库索引方法实现对被图幅线分割的对象的逻辑连接,实现对象内部(目标单元)的无缝性。这种组织方法可以正确地实现对被分割对象的完整检索,但在属性查询时仍需要进行处理,建库时需要建立好空间目标全库索引,同时在后续的操纵中要进行维护。常燕卿(2000)讨论了不同投影带下解决空间数据无缝的两种方法:①几何变换法,当可视窗口横跨两个投影带时,将不同的投影带数据统一到一个投影带;②实时投影变换法,以地理坐标存储数据,显示时始终以可视范围中心为投影的中央经线,在数据漫游中实时将可视范围内的地形数据动态投影到同一坐标体系内,数据进行实时拼接,从而避免投影的跨带问题。

胡鹏等(2001,2002,2005)研究了地理信息系统的空间数学基础,讨论了“无缝空间数据库”技术的若干基本问题,认为“无缝”不仅是形式上的“无缝可视”,更重要在于“无缝量度”、“无缝空间分析”、“规范的无缝集成”的实质能力上,进而将无缝分为物理无缝和视无缝;并认为真正的无缝不仅需要视无缝,

还需要真正拓扑意义上的任意方向均连续、多分辨率上均精密匹配,以达到统一度量和分析。

赵学胜(2004)针对传统平面数据模型在表示大范围数据时存在的许多问题,提出数字地球涉及的是全球数据的表示组织,这就要求地球数据在计算机内部的数字表达必须是全球的、连续的和变化的,即建立球面动态数据模型。全球网格数据模型的研究也吸引众多的研究者。由于 QTM 格网结构在全球范围是无缝的(seamless)、稳定的(stable),且具有层次结构,能有效地管理多分辨率的海量数据并进行不同层次的操作,因此 QTM 结构最受关注。不过全球网格数据模型目前还难以直接广泛应用。

针对空间数据缝隙产生的来源不同,朱欣焰等(2002)探讨了无缝空间数据库的概念,提出了空间数据库的逻辑无缝、物理无缝、逻辑查询无缝、逻辑接边、物理接边等概念;研究了基于关系数据库或对象关系数据库实现物理无缝的优势及其在存储、存取方面存在的问题。王卉等(2004)讨论了无缝 GIS 的概念框架,区分了几何无缝和物理无缝。几何无缝是指在统一坐标系下,图形显示时相邻图幅或区域的内容在视觉上不存在缝隙,是连续一致的图形。物理无缝是指一种无缝存储,存储地理数据时能实现数据非分割、非分幅存储,即地理数据能跨图幅,整体存储在关系数据库中,并已进行了几何拼接和属性拼接。指出了物理无缝存在的一些不足,认为物理无缝不是一种理想的地理信息无缝,只适用于确定信息范围的地理空间的无缝表示,并建议采用将空间数据分块存储于分布式数据库中,数据库提供相应的图块拼接信息及相应的拼接访问手段,保障空间数据在使用上的空间连贯性,而数据在逻辑使用上是无缝的。王卉等(2004)还讨论了无缝 GIS 数据生产的流程和关键技术,认为:无缝 GIS 就是在固定的比例尺下,使用户能在更广阔的空间数据基础上进行无缝显示和分析应用的 GIS。李爱光等(2005)研究无缝 GIS 空间数据组织时,指出基于拓扑关系数据模型空间数据组织方法不太适合大型无缝空间数据组织,因而使用非拓扑数据模型进行研究。

2. 影像数据的无缝组织与管理

影像数据不仅对地理信息系统中矢量数据的更新起到非常重要的作用,而且成为一种重要的数据,直接或结合矢量数据为用户提供空间信息服务。许多 GIS 软件或影像数据管理软件都可以将数字正射影像数据、遥感数据作为背景影像与矢量数据、DEM 数据进行套合显示输出,具有提供大范围的影像数据的无缝组织与管理能力,并在 Web 环境下提供使用,如 Microsoft TerraServer(<http://www.terraserver.com>)、美国 Erdas 公司的 Image Catalog(<http://www.erdas.com>)、ESRI 公司的栅格影像管理模块、Google 公司的 Google Maps 和 Google Earth 等。在国内,方涛等(1999)、王密(2001)分析了我国数字正射影像产品的空间参考特性和建立大型无缝影像数据库所带来的问题,提出了采用金字塔结构建立大型无缝

影像库的技术思路。王密提出了海量影像数据管理中分带存储模式和跨带漫游算法,解决了不同的尺度之间切换时产生的裂缝问题,同时还首次提出了基于大型无缝正射影像数据库的可量测虚拟现实(measurable virtual reality,MVR)的概念和实现方法。

海量遥感影像数据库系统的首要问题是解决数据库查询及获取的效率问题,Buzi 等(2001)提出了数据查询的零延迟目标。杨忠德等(2004)也阐述了海量影像无缝数据库管理系统实现的基本原理及关键技术,并利用面向对象数据库系统 ODBMS 实现一个海量影像数据库系统。当前,对遥感影像数据库的研究需要更加重视多数据源、多分辨率和多种空间参考的无缝影像数据管理,以及分布式的异构海量影像数据库管理模式等方面。

3. DEM 数据无缝组织与管理

数字高程模型(digital elevation model,DEM)通常的数据组织方式有不规则三角网(triangular irregular network,TIN)、正方形格网(GRID)或者称为规则正方形格网(regular square grid,RSG)和等高线(contour-based DEM)等。目前,我国省级和国家级的真实地形数据大多数都是采用规则格网结构的数字高程模型进行存储。

GRID 结构的数字高程模型数据,一般都是作为传统 GIS 中的基础地理信息数据进行存储的,例如 ArcInfo 中将 GRID 作为一个图层,分成多个 Tiles 进行存储。GeoStar 3.0 的 GeoGrid 软件在采用文件系统作为数据库管理平台的同时,以图幅为单位对数字高程模型数据进行管理。一些专用的数字高程模型数据库采用了关系数据库管理系统进行数据库的管理,如全国七大江河 1:1 万数字高程模型数据库使用了 ArcView 和 Oracle 进行数字高程模型数据库的管理。为了解决不同高斯坐标带数据之间的连续显示问题,同时建立了地理坐标表示的数据库和以高斯坐标表示的数据库。GeoStar 4.0 以后的版本,数字高程模型数据可以用文件存储,也可以用关系数据库存储。

由于生产时大多数数字高程模型数据都是使用高斯分带直角坐标进行表示的,为了保持数据的原始性,比较理想的是直接使用原始数据建库。以高斯分带直角坐标系表示的数字高程模型数据,实际上表现为物理有缝的状态。为了在数据浏览和大范围的三维场景建立中给用户带来真实世界的感受,需要以一种逻辑无缝的形式对数据进行表现,即最终用户浏览和观察到的数据库应该是没有缝隙的。因此,数据库浏览和查询机制要保证呈现给用户逻辑上自然无缝的三维地形景观,需要组织好数据库数据组织上的有缝和逻辑上无缝之间的关系。

王永军(2003)对大范围海量地形数据管理涉及的相关问题进行了讨论。研究了基于关系、对象关系数据库管理系统的金字塔 DEM 数据库管理模型。在原有

的 GeoGrid 基于文件系统结构的金字塔数据库模型基础上,对金字塔的层次结构进行了改造和细化,在空间参考、数据分层、索引机制、快速浏览、大范围跨带飞行漫游支持和数据压缩等方面进行了改进。探讨了在关系数据库管理系统基础上建立具有多分辨率、多比例尺金字塔 DEM 数据库的方法,提出了相应的数据分块和索引机制,实现了基于 RDBMS 对不同坐标参考系跨越高斯分带的海量数字高程模型数据的管理。

在采用细节层次表示技术时,由于可视化时需要调用不同分辨率层次的数据,通常相邻子块的分辨率不一致会引起裂缝。在 DEM 可视化方面的缝隙及解决方法已有大量的研究,目前已得到较好的解决。

4. 空间数据无缝集成研究

空间数据集成不仅包括数据格式,还包括不同类型、不同表达、不同比例尺(分辨率)、不同时间之间的数据集成。多源空间数据无缝集成是 GIS 研究的热点之一。

宋关福等(2000)研究了多源空间数据的无缝集成问题,总结了多格式集成的几种模式。李军等(2000)将多尺度数据的集成分解为空间和时间多尺度数据集成,对两种多尺度数据传统和数据意义上的集成方法进行了详细的探讨。杨永崇等(2005)研究了土地信息系统中多分辨率数据的集成问题。王卉(2003)介绍了数据标准化模式和操作标准化模式两种多源空间数据集成途径,讨论了中间件模式实现多源空间数据无缝集成方法。此外,学者研究了基于多代理无缝上下文感知信息服务实现集成。Gupta 等(2002)通过中介器(mediation)实现多空间数据源集成。国内一些学者对基于 GML 实现异源数据集成做了许多研究。Jaeger 等(2005)将不同来源、不同地点的空间信息组合处理问题看作科学工作流,基于 Web Services 来实现分布式地理空间信息处理和集成。

语义集成是目前多源空间数据集成研究的又一个热点。语义集成是通向遥感图像和 GIS 地理空间建模的一个重要步骤。语义集成要解决数据源、系统之间语义无缝。语义差异的来源复杂,主要分为两种语义差异:同名异义(homonyms)和同义异名(synonyms)。Wiegand 等(2005)认为空间数据高度异构,不仅使数据表达和存储方法不同,而且数据查询方法也不同,分布式查询的最大问题是语义差异;虽然空间信息系统表现为一个集成工具,但 GIS 互操作性还远不能达到完全实用。OGC 定义了目录服务来支持发现、组织和存取地理信息,但未定义克服语义异构的任何方法,而语义异构问题是地理空间数据领域开放和分布环境下空间数据发现和检索方面的一个主要挑战。Sheth(1998)将互操作的发展过程分为三个阶段:第一个阶段是处理有关结构异质问题,包括查询语言差异和 DBMS 差异;第二个阶段是使用元数据向用户提供统一表达与存取异构数据源的方法;第三个阶段是关键阶段,解决语义异构问题。Sheth 认为本体、上下文、语义相互关系是