

文章编号: 1007-4619 (2000) 02-0157-04

# 空间数据仓库的技术与实践

赵霏生, 杨崇俊

(中国科学院 遥感应用研究所, 北京 100101)

**摘 要:** 空间数据仓库(Spatial Data Warehouse)是 GIS 技术和数据仓库技术相结合的产物,它大大扩展了 GIS 的应用功能,为逐渐兴起的全球变化和区域持续发展研究以及复杂的商业地理分析提供了强有力的支持。该文主要从空间数据仓库的起源、特征、体系结构、关键技术与应用等方面来探讨空间数据仓库这一全新的 GIS 应用。

**关键词:** 数据仓库;GIS;空间数据仓库

**中图分类号:** TP311.132 **文献标识码:** A

## 1 空间数据仓库的出现

当代地学问题的提出和解决常常要求多学科多部门之间的协作。因此,空间信息的共享与互操作、分析与综合在对于地学综合问题的研究中便显得日趋重要。传统的 GIS 应用系统一般是面向某一个具体应用、由日常的工作流程驱动的,数据往往处于采集时的原始状态;系统应用也只是对业务数据进行增、删、改等事物处理操作<sup>[1]</sup>和简单的空间查询与分析。为了更好地适应当今全球变化和可持续发展研究的需要,需要一个统一的信息视图将来自不同学科的相关数据按照相应的主题转换成统一的格式,集成、存储在一起,然后通过各种专业模型从多个角度去了解这个世界。空间数据仓库正是为了更好地满足当今地球科学研究的需要而产生的。

空间数据仓库和一般的空间数据库在物理本质上均是对数据高效地存贮。空间数据仓库是建立在传统的数据库管理系统之上,依靠它们管理数据的存贮,而不管它们是集中式的,还是分布式的;是松耦合的,还是邦联式的。二者之间的差别在于它们面向的应用不同,因此在数据的组织、集成上存在着较大的差异。空间数据库(源数据库)负责原始数据的日常操作性应用,提供简单的空间查询和分析;空间数据仓库则根据主题通过专业模型对不同源数据

库中的原始业务数据进行抽取和聚集,形成一个多维视角,为用户提供一个综合的、面向分析的决策支持环境<sup>[2]</sup>。这样,空间数据仓库和空间数据库各自只做自己所擅长的,系统的性能达到最佳。另外,空间数据仓库较好地引入了时间维的概念,可根据不同的需要划分不同的时间粒度等级,以便进行各种复杂的趋势分析,如土地覆盖变化研究、全球气候的变化趋势等,以支持政府部门的宏观决策。

目前空间数据仓库已成为国外 GIS 界研究的热点,并已被应用到多个项目,如澳大利亚的土地管理系统<sup>[3]</sup>、苏格兰的资源环境信息系统<sup>[4]</sup>等。由于空间数据仓库管理的数据量非常大,支持决策的空间计算非常复杂,因此研究主要着重于以下几方面<sup>[5-7]</sup>:

- (1)空间数据结构,即如何高效地存贮和检索数据;
- (2)多源数据的集成,即空间、时间、属性数据的集成以及异构数据之间的集成;
- (3)空间数据的管理,即如何将空间数据放入到大型的商业数据库中;
- (4)空间数据挖掘,即决策支持模型方法的研究。

## 2 空间数据仓库的主要功能特征

数据仓库是面向主题的、集成的、具有时间序列

收稿日期: 1999-01-04; 修订日期: 1999-09-15

基金项目: 中国科学院“九五”重大及特别支持项目《中国资源环境遥感信息系统及国情速报》All rights reserved. <http://www.cnki.net>

作者简介: 赵霏生(1971-),男,北京市人,1994年在中国科学院自然资源综合考察委员会获硕士学位,现在中国科学院遥感应用研究所攻读博士学位,从事分布式网络 GIS 研究,发表论文3篇。

特征的数据集合,以支持管理中的决策制定过程。空间数据仓库则是在数据仓库的基础上,引入空间维数据,增加对空间数据的存贮、管理和分析能力,根据主题从不同的 GIS 应用系统中截取从瞬态到区段直到全体地球系统的不同规模时空尺度上的信息,从而为当今的地学研究以及有关环境资源政策的制定提供最好的信息服务。

空间数据仓库为了决策支持的需要,主要具有以下几方面功能特征:

(1)空间数据仓库是面向主题的。传统的 GIS 数据库系统是面向应用的,只能回答很专门、很片面的问题,它的数据只是为处理某一具体应用而组织在一起的,数据结构只对单一的工作流程是最优的,对于高层次的决策分析未必是适合的。空间数据仓库为了给决策支持提供服务,信息的组织应以业务工作的主题内容为主线。主题是一个在较高层次将数据归类的标准,每一个主题基本对应一个宏观的分析领域<sup>[1]</sup>。例如土地管理部门的空间数据仓库所组织的主题有可能为土地覆盖的变化趋势、土地利用变化趋势等;而按照应用来组织则可能是地籍管理、土地适宜性评价等。很显然,按照应用来组织的系统不能够为土地管理部门制定决策提供直接、全面的服务,而空间数据仓库的数据因其面向主题,具有“知识性、综合性”,所以能够为决策者们提供及时、准确的信息服务。

(2)空间数据仓库是集成的。空间数据仓库的建立并不意味着要取代传统的 GIS 数据库系统。空间数据仓库是为制定决策提供支持服务的,它的数据应该是尽可能全面、及时、准确、传统的 GIS 应用系统是其重要的数据源。为此空间数据仓库以各种面向应用的 GIS 系统为基础,通过元数据刻画的抽取和聚集规则将它们集成起来,从中得到各种有用的数据。提取的数据在空间数据仓库中采用一致的命名规则,一致的编码结构,消除原始数据的矛盾之处,数据结构从面向应用转为面向主题。

(3)数据的变换与增值。空间数据仓库的数据来自于不同的面向应用的 GIS 系统的日常操作数据,由于数据冗余及其标准和格式存在着差异等一系列原因,不能把这些数据原封不动地搬入空间数据仓库,而应该对这些数据进行增值与变换,提高数据的可用性,即根据主题的分析需要,对数据进行必要地抽取、清理和变换。最常见的操作有语义映射、获取瞬像数据、实施集运算、坐标的统一、比例尺的变换、数据结构与格式的转换、提取样本值等。

(4)时间序列的历史数据。自然界是随着时间而演变的,事实上任何信息都具有相应的时间标志。为了满足趋势分析的需要,每一个数据必须具有时间的概念。

(5)空间序列的方位数据。自然界是一个立体的空间,任何事物都有自己的空间位置,彼此之间有着相互的空间关系,因此任何信息都应具有相应的空间标志。一般的数据仓库是没有空间维数据的,不能做空间分析,不能反映自然界的空间变化趋势。

### 3 空间数据仓库的体系结构

空间数据仓库是储存、管理空间数据的一种组织形式,其物理实质仍是计算机存储数据的系统,只是由于使用目的不同,其存储的数据在量和质以及前端分析工具上与传统 GIS 应用系统有所不同。空间数据仓库按照功能分为以下几部分:

(1)源数据。空间数据仓库为了支持高层次的决策分析需要大量的数据。这些数据可能分布在不同的已有的应用中,存储在不同的平台和数据库中。

(2)数据变换工具。为了优化空间数据仓库的分析性能,源数据必须经过变换以最适宜的方式进入空间数据仓库。变换主要包括提炼、转换、空间变换。数据提炼主要指数据的抽取,如数据项的重构、删去不需要的运行信息、字段值的解码和翻译、补充缺漏的信息、检查数据的完整性和相容性等<sup>[8]</sup>;数据转换主要指统一数据编码和数据结构、给数据加上时间标志、根据需要对数据集进行各种运算以及语义转换等;空间变换主要指空间坐标和比例尺的统一、赋予一般数据空间属性。数据变换工具为数据库和空间数据仓库之间架起了一座桥梁,使源数据得到了增值和统一,最大限度地满足了空间数据仓库高层次决策分析的需要。元数据是数据仓库的核心,它是关于数据的数据,定义了数据的存贮模型和数据结构以及转换规则和控制信息等。

(3)空间数据仓库。源数据经过变换进入空间数据仓库。空间数据仓库以多维方式来组织数据和显示数据。维是人们观察现实世界的角度,但多维数据库中的维并不是随意定义的,它是一种高层次的类型划分<sup>[9]</sup>。为了获得较高的系统性能,维屏蔽掉了许多原始数据,决策分析所需的综合数据预先已被统计出来放在其中。空间维和时间维是空间数据仓库反映现实世界动态变化的基础,它们的数据组织方式是整个空间数据仓库技术的关键。在实际

分析过程中,可按需要把任一维与其它维进行组合,以多维方式显示数据,供人们从不同角度来多方位地认识世界。空间数据仓库的数据组织方式可分为基于关系表的存储方式和多维数据库存储方式。基于关系表的数据模型主要有星型和雪花模型<sup>[10]</sup>;多维数据库数据模型主要是超立方体结构模型。空间维数据的具体表现形式为空间对象的名称和指向空间对象的指针。空间数据仓库的大多数计算分析主要集中在空间多边形的融合和分离。

(4)客户端分析工具。空间数据仓库的目标是提供决策支持,它不仅需要一般的GIS查询和分析工具,更需要功能强大的分析工具,它是空间数据仓库的重要组成部分。客户端分析工具按照功能可以划分为查询型、验证型、挖掘型,主要采用旋转、嵌套、切片、钻取和高维可视化分析技术,以多维视图的形式展现给用户,使用户能直观地理解、分析数据,进行决策支持。

## 4 空间数据仓库的应用实例<sup>[11]</sup>

中国资源环境遥感信息系统及农情速报(简称CRERSIS)是中国科学院“九五”特别支持项目。该系统的数据来自于中国科学院长期积累的多学科、多尺度、多类型的科学资料和信息;系统的主要任务是进行全国主要农作物的农情速报和中国可持续发展能力及资源环境方面的国情分析。该系统是面向决策支持的,数据量大,数据类型多样,既有空间特性又有时间特性,需要规范、沉淀、提炼和集成,以便得到有用的信息和知识,在设计上需要按照不同的决策主题来组织数据,因此CRERSIS比较适合利用空间数据仓库技术来完成。

CRERSIS中的数据来自于不同的平台,如Arc/Info、Intergraph/MGE,类型多种多样,有资源环境背景数据、土地资源数据、生态环境数据、卫星影像、统计数据、气象信息、专题数据等。为了将这些中国科学院多年研究积累的丰富数据更好地服务于决策支持,CRERSIS按照农作物长势监测、农作物种植面积监测、农作物估产、水灾遥感监测、旱灾遥感监测、生物量监测、中国生态环境分析、中国土地资源时空变化分析8个主题来组织数据。

CRERSIS采用统一的数据模型将空间数据存贮到关系型数据库中,数据组织采用多维数组结构进行数据存储,并有维索引及相应的元数据管理文件与数据相对应。为了优化查询、分析过程,采用了星

形模型,使空间数据仓库的复杂查询可以直接通过各维的层次比较、上钻、下钻等操作完成。

从气象卫星接收到数据处理,数据采集与更新部分以现有系统为基础,定期收集来自不同平台的,如Arc/Info、ERDAS、PCI、SPLUS等,与8个主题相关的图形、图像和属性数据,将其转换到统一的平台之上进行集中统一管理。

数据管理和调度部分通过元数据集中统一地管理空间数据、属性数据和其它非结构化的数据。为了统一管理空间数据和属性数据,在数据管理和调度部分采用了ESRI公司的SDE(Spatial Data Engine)平台和Oracle数据库。SDE采用了C/S结构,适合多用户的并发查询;解决了在关系数据库中存贮连续、大型的空间数据的问题;同时,SDE提供了强大的在线分析功能,可以进行集成式的空间数据分析,并可利用ArcView、MapObjects等现有的工具软件进行空间和数值数据的查询和显示。

在查询和分析部分,通过各种应用模型和可视化分析工具,用户可借助网络根据系统的提示能同时对图形、图像、表数据、文本信息进行查询;可根据区域、时间序列组织数据;从不同角度浏览、查询各主题的数据、信息和分析结果。系统选用了BusinessObject等软件工具完成数据挖掘、建立应用模型、数据汇总和详查工作。

空间数据仓库的解决方案使CRERSIS走出此前类似系统的简单性和非实用性,解决了系统内在结构的复杂特性,以及现行系统的多异质性问题,真正达到了为决策支持服务的目的。

## 5 结 论

数据仓库技术是90年代信息技术的热点,并已逐步展开应用。空间数据仓库是在数据仓库技术基础上发展起来的,刚刚起步。它绝不仅仅是在数据仓库上加一层空间的外衣,以地图的形式来反映某些结果,而是真正地以空间数据库为基础,能够进行复杂的空间分析,反映自然界不同时空尺度下的动态变化趋势,为决策者提供及时、准确的信息。空间数据仓库将以其面向主题的数据集成、功能强大的多维分析越来越多地应用在空间决策支持系统中。

## 参考文献 (References)

- [1] Wang Shan, Luo Li. From Database to Data Warehouse [N]. Computerworld, 1996, 28. [王 珊, 罗 立. 从数据库到数据仓库

- [N]. 计算机世界, 1996, 28:(专题版)]
- [2] Bob Lambert. Data Warehousing Fundamentals: What You Need to Know to Succeed. <http://www.techweb.cmp.com/iwk.htm>
- [3] Nick Schultz, Application Integration with CORBA/JAVA/WWW/SDE. <http://www.esri.com/library/userconf/proc98/proceed/abstract/a204.htm>
- [4] Neil Brooker. Developing a Web-Based Interface to a Corporate Spatial Data Warehouse for Scottish Environmental Data. <http://www.esri.com/library/userconf/proc98/proceed/abstract/a106.htm>
- [5] Jiawei Han. Spatial Data Warehouse and Spatial Data Mining. <http://db.cs.sfu.ca/sections/publication/smmdb/smmdb.html>
- [6] Cubewerx Corporation. System Architecture of SDW. <http://www.cubewerx.com/systemarchitecture.com>
- [7] ActiveG Corporation. Spatial Data Warehouse. <http://www.activeg.com/sdw.htm>
- [8] Guo Yibin. The Basic Conception and status quo of Data Warehouse [J]. *PCWorld*, 1996, (4): 26-31. [郭宜斌. 数据仓库技术的基本概念和发展现状[J]. 微电脑世界. 1996, (4):26-31.]
- [9] Chen Yuan, Chen Wenwei. The Multidimensional Data Analysis of OLAP [N]. *Computerworld*, 1998, 22. [陈元, 陈文伟. OLAP 的多维数据分析[N]. 计算机世界, 1998, 22(专题版)]
- [10] Gao Renbo, Chen Wenwei. Data Warehouse and Data Organizing of OLAP [N]. *Computerworld*, 1998, 22. [高人伯, 陈文伟. 数据仓库和 OLAP 的数据组织[N]. 计算机世界, 1998, 22(专题版)]
- [11] The key project of CAS Ninth-Five-Year Plan. The key Technology Report of Operational Remote Sensing Information System of Resource and Environment in China [R]. 1998. [中国科学院“九·五”重大支持项目. 《中国资源环境遥感信息系统及农情速报》关键技术报告[R]. 1998.]

## The Technology and Application of Spatial Data Warehouse

ZHAO Pei-sheng, YANG Chong-jun

(Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** As the combination of Data Warehouse and GIS, Spatial Data Warehouse is a large database designed to support the needs of decision making. For this purpose, it has some new features that are different from those of the traditional database, such as integrated, subject-oriented, time and spatial variant. The emergence of Spatial Data Warehouse extends the scope of the applications of GIS and supports the research on global change, development of region and complex analysis of commerce. The system of Spatial Data Warehouse mainly includes four components. These are data source, tools of data transformation, warehouse, and analytical tools. All of these components involve some new and key technologies. We must find answers about how to define an efficient spatial data structure, integrate various data source, and find information from raw data. They are critical to the success of building spatial data warehouse.

From emergence system architecture to building an application of Spatial Data Warehouse, this article discusses this very new GIS application.

**Key words:** data warehouse; GIS; spatial data warehouse