

苏天赞,崔爱菊,梁瑞才,等. 海洋地球物理空间数据仓库系统的设计与实现. 地球物理学进展, 2011, 26(4):1478~1484, DOI: 10.3969/j.issn.1004-2903.2011.04.045.

Su T Y, Cui A J, Liang R C, *et al.* Design and development of marine geophysics spatial data warehouse system. *Progress in Geophys.* (in Chinese), 2011, 26(4):1478~1484, DOI:10.3969/j.issn.1004-2903.2011.04.045.

海洋地球物理空间数据仓库系统的设计与实现

苏天赞^{1,2}, 崔爱菊¹, 梁瑞才^{1,2}, 王建村^{1,2}, 王 勇^{1,2}

(1. 国家海洋局第一海洋研究所, 青岛 266061; 2. 国家海洋局海洋沉积与环境地质重点实验室, 青岛 266061)

摘 要 本文将空间数据仓库技术引入到海洋地球物理空间数据的管理和应用上来,探索一套适合于海洋地球物理调查和科研工作的空间数据仓库构建方法. 本文重点从数据提取工具、空间数据仓库和应用服务器等几个方面对海洋地球物理空间数据仓库系统进行设计和开发,实现海洋地球物理空间数据的组织和管理,向用户提供便捷、直观、高效的面向应用主题的空间数据查询、分析、统计和可视化等服务.

关键词 海洋地球物理, 空间数据仓库, 数字海底, 多维数据模型

DOI:10.3969/j.issn.1004-2903.2011.04.045 中图分类号 P631 文献标识码 A

Design and development of marine geophysics spatial data warehouse system

SU Tian-yun^{1,2}, CUI Ai-ju¹, LIANG Rui-cai^{1,2}, WANG Jian-cun^{1,2}, WANG Yong^{1,2}

(1. First Institute of Oceanography, SOA, Qingdao 266061, China; 2. Key Lab of Marine Sedimentology and Environmental Geology, SOA, Qingdao 266061, China)

Abstract By conducting the technology of Spatial Database Warehouse (SDW) to the management and analysis of marine geophysical spatial data, a spatial data warehouse meeting the requirement of marine geophysical survey and has been proposed. The marine geophysics SDW system was designed and developed mainly from such aspects as data extraction tools, spatial data warehouse and application server. The system can not only aggregate and manage the marine geophysical spatial data, but also provide convenient, efficient, visible and subject-oriented services such as spatial data query, spatial analysis, spatial statistics and so on.

Keywords marine geophysics, Spatial Data Warehouse (SDW), digital seabed, multidimensional data model

0 引 言

随着“数字地球”概念的提出以及相关理论和技术迅猛发展,先进的空间信息技术开始广泛地应用在海底科学调查和研究领域的各个方面. 经过长期的探索与实践,许多海洋地学数据库和数据中心已经建立,并且为海洋科研和工作人员提供便捷、高效的数据共享服务,如美国 NGDC 的 Marine Geology & Geophysics 数据库、英国斯科特极地研究所的极

地与冰川数据库等.

目前,已建成或正在建设的海洋科学数据库主要是基于业务信息(如项目、航次等)对海洋专业要素进行组织和管理,而专业要素的空间特征则作为属性信息分别存储在各个相关的数据表中,强调数据的最小冗余度和最大一致性约束,并没有建立空间索引机制. 这种数据库能够满足基于属性信息的海洋专业要素查询和共享的应用需求. 但是,当基于空间特征对海洋综合信息进行空间检索、分析和可

收稿日期 2010-09-13; 修回日期 2010-12-01.

基金项目 国家“863”计划课题“面向网络的海洋地球物理空间数据仓库关键技术研究”(2008AA09Z304)资助.

作者简介 苏天赞,男,1977 年生,博士,副研究员,主要从事海洋地球物理资料处理、海洋地理信息与虚拟现实方面的研究.
(E-mail:sutianyan@fio.org.cn)

视化表达等功能时,系统效率会明显降低,甚至无法满足应用需求。例如,当需要根据空间约束条件(如经纬度范围)查询专业要素时,往往需要首先遍历项目数据表,然后对每一个项目中所有专业要素的空间属性信息与空间约束条件进行比较,找出符合空间约束条件的空间对象,再将空间对象的属性信息返回到客户端,这就相当于检索一条记录需要进行若干步的操作。而且,在没有建立空间索引机制的情况下,空间属性信息与空间约束条件进行比较的执行效率会随着数据量的增加明显降低。

20 世纪 90 年代初,在传统的数据库(Database, DB)的基础上,数据仓库(Data Warehouse, DW)的概念被提出,并且得到了迅速的发展和应用。许多专家和学者对其进行了定义^[1,2],简单概括起来,数据仓库是一个来自于异构数据源、基于不同的细节层次(或粒度)进行存储、面向主题的、集成的、不可更新的、随时间不断变化的数据集合,用以更好地支持企业或组织在经营管理中的决策分析。空间数据仓库是具有空间维和空间度量的数据仓库,能够实现分散的、各自独立的包含空间信息的数据集进行统一集成和管理,形成统一空间标准(如坐标系统、投影方式、比例尺等)和空间索引机制,支持面向主题的大数据量的空间分析和空间数据挖掘等操作^[3]。经过十多年的发展,空间数据仓库的相关理论和技术得到了蓬勃的发展^[4~8],并且已经出现了一些成熟的空间数据仓库产品,如美国 ESRI 公司的 ArcSDE、美国 MapINFO 公司的 SpatialWare、美国 Microsoft 公司的 TerraServer 等,从而促使空间数据仓库在众多领域展开研究、示范和实质性应用^[9~17]。

本文将空间数据仓库技术引入到海洋地球物理空间数据的管理中,针对其中的关键技术问题进行探索和应用,研究适合于海底专业要素的空间管理和空间决策支持的空间数据仓库技术,为海洋资源规划、管理、海域地质勘探提供高效的空间数据分析服务,满足“数字海洋”、“数字海底”的发展需求。

1 系统设计

1.1 总体设计

海洋地球物理空间数据仓库系统的体系结构如图 1 所示。海洋地球物理空间数据仓库系统主要包括数据源、数据提取工具、空间数据仓库、应用服务器以及客户端等五部分,其中主要的设计工作在数据提取工具、空间数据仓库、应用服务器三部分。

1.2 数据提取工具

由于空间数据仓库的数据来自于多个不同的数据源,这些数据源采用的格式和标准不一致,尤其是空间参考标准,如空间坐标系、投影方式、比例尺、空间网格大小等。需要数据提取工具从不同的数据源中提取数据,并且对这些数据进行必要的转换、标准化、集成,使其符合海洋地球物理空间数据仓库的模型结构、空间标准和语义规范。数据提取的方法包括:数据清洗、数据转换和数据聚集。数据清洗是指对数据源中的数据内容和结构进行提炼,删掉空间数据仓库主题应用不需要的内容,适当增加派生性冗余字段(由现有的字段派生,便于空间分析和空间数据挖掘等操作的运行),进行一致性和完整性检查。数据转换主要包括两部分,一部分是对数据结构的转换,将数据源中的格式和编码转换成空间数据仓库中统一设计的格式;另一部分是空间转换,将空间数据的空间参考标准和空间存储格式转换成统一的标准。数据聚集主要是根据空间数据仓库设计的多维数据模型,将经过提炼和转换后的数据进行聚集操作(通常采用 group-by 的 SQL 查询操作符)^[18],将数据进行集成。

1.3 空间数据仓库

海洋地球物理空间数据仓库根据一定的应用主题,设计空间数据仓库的数据模型、粒度分割和存储结构,实现对海洋地球物理空间数据的集成和管理,并且能够快速响应来自应用服务器端的高端应用请求。下面以海洋地球物理调查工作量统计(即测线长度的计算)为应用主题,介绍海洋地球物理空间数据仓库的设计方法。

1.3.1 数据模型

构建适合于一定应用主题的数据模型是海洋地球物理空间数据仓库设计的主要内容。数据仓库通常采用多维模型(也可称为“数据立方体”或超立方体)来构建,这种模型能够便于决策支持工具更准确、更迅速地从数据仓库中挖掘信息。多维数据模型有三个核心的概念:事实(fact),维(dimension)和度量(dimension)。“事实”代表关于所感兴趣的“维”的状态,而“度量”则是“事实”的一个属性。“度量”是一个因变量,依赖于每一个“维”(自变量)的信息。例如,对于“海洋地球物理调查工作量”这一“事实”,“专业”、“区块”和“测线”可能是它所关心的三个“维”,而“测线总长度”则是由这三个“维”确定的“度量”,这样由(“重力”,“DW01 区块”,“测线对象”)三个“维”的值确定的“度量”值“9223 公里”就是“事实”的一个属性。多维数据建模方法能更直接地反映

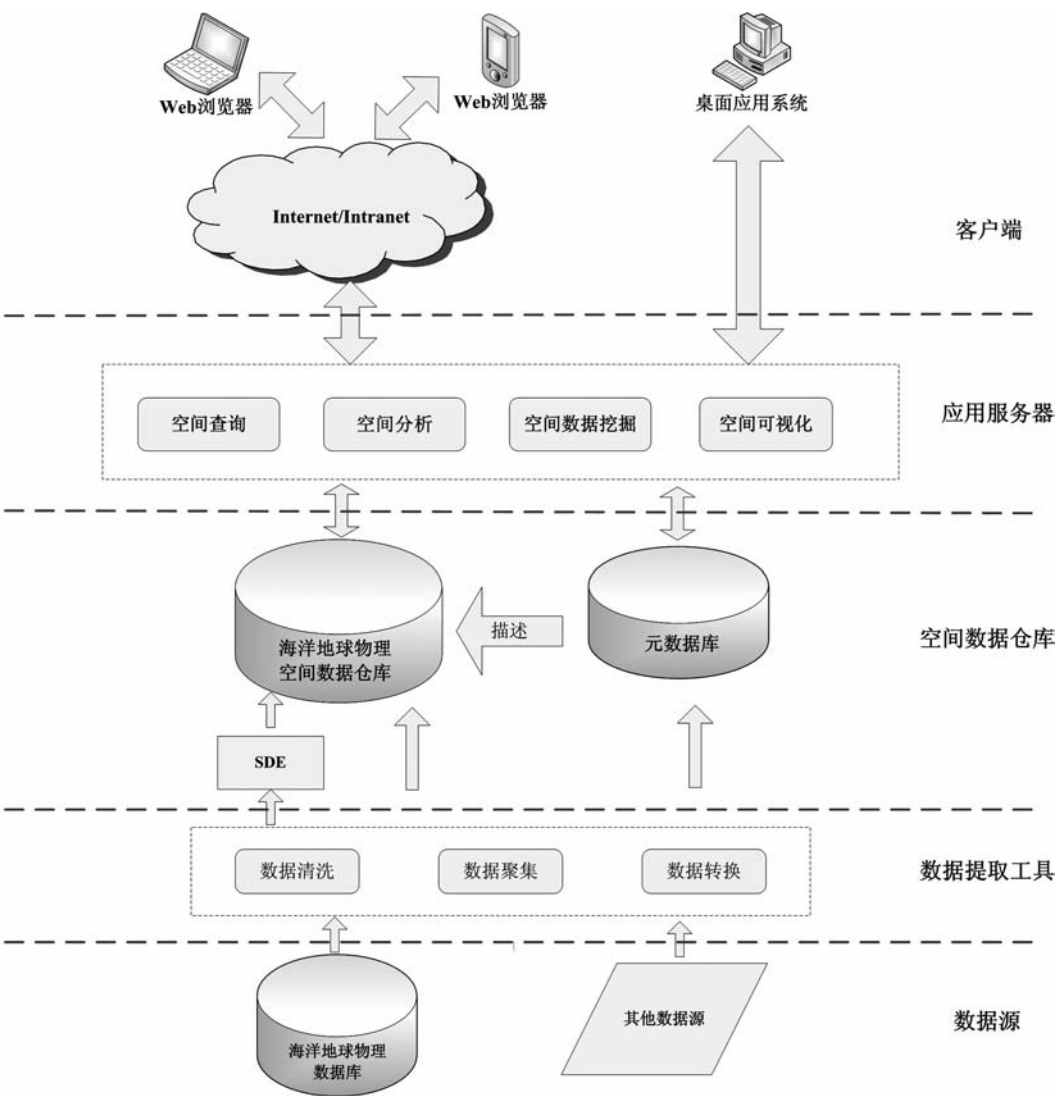


图 1 海洋地球物理空间数据仓库系统的体系结构
Fig. 1 Framework of marine geophysical spatial data warehouse

我们对信息和行为的理解方式(比如,先确定自变量,然后找出受影响的因变量),这样能够更好的对数据仓库进行挖掘.

海洋地球物理空间数据仓库的多维数据模型与普通的数据仓库不同,主要按照空间特征来组织数据,构建多维数据立方体.空间维是空间数据仓库表达空间对象的基础.在海洋地球物理调查工作量统计应用主题中,“测线”维就是空间维,它的值对应空间对象,包含该空间对象的空间信息(例如空间坐标系、投影方式、坐标值等),工作量统计的计算主要就是测线空间对象的度量长度进行统计.另外,如果需要反映空间对象的变化特征,还需要考虑增加时间维.

多维数据模型可以采取三种数据结构来建模:星型模型(star schema),雪花模型(snowflake schema)

图 2 工作量统计应用主题的星型多维数据模型
Fig. 2 Star schema multidimensional data model of statistics of geophysical survey

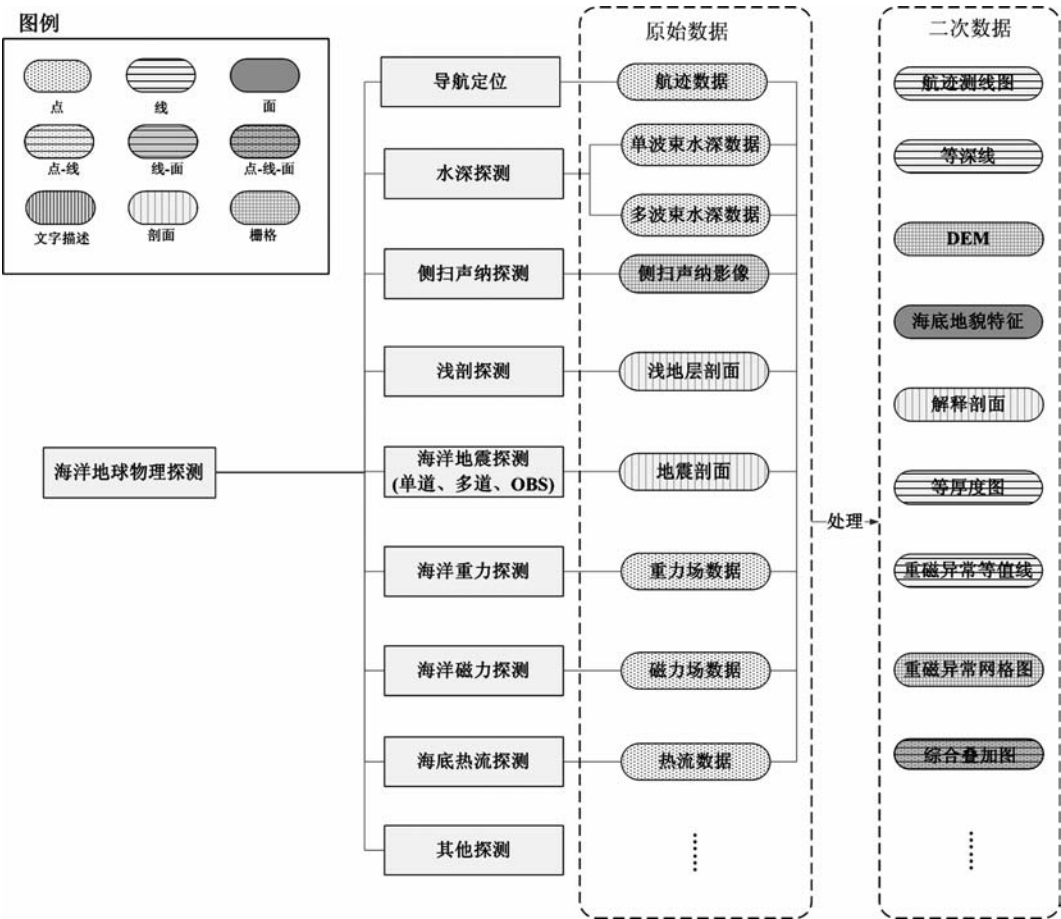


图 3 海洋地球物理数据分类和空间表达方式
Fig. 3 Classification and spatial formats of marine geophysical data

和事实星座(the Fact Constellation)^[19]. 海洋地球物理空间数据仓库主要采用星型模型来建模. 星型模型有一个中央“事实”表,该表包含多个“维”(列),每一个“维”的属性指向对应“维”表中某一记录的外键. 图 2 是海洋地球物理调查工作量统计应用主题的星型多维数据模型.

1. 3. 2 粒度分割

在多维数据模型中,每一个“维”有许多值,每一个值对应维度坐标轴上的一个位置. 如果将一个“维”上的变量根据不同的粒度进行重构,将产生不同层次级别的信息

(如年细分为季,季细分为月,月细分为周,周细分为日). 另外,不同级别粒度的组合也可以表达相同的“维”(如年-月-日,年-季-周). 粒度越小,数据越详细,综合度也越低;相反,粒度越大,级别就越高,数据综合度也越高,但数据内容也越含糊. 海洋地球物理空间数据仓库对数据粒度的分割主要分两类:属性粒度分割和空间粒度分割. 属性粒度分割主要针对属性维进行分割,例如项目可以细分为区块,项

目维是对区块维的综合. 空间粒度分割主要针对空间特征进行划分,比如空间比例尺,空间网格大小等. 粒度的大小直接关系到空间数据仓库的数据量和所适合的查询类型,应该将“维”表中的内容根据最小的粒度进行分割,而“事实”表则根据具体的应用主题选择相应的粒度来提取和组织数据,在满足应用需求的同时,节省存储空间.

1. 3. 3 数据存储

空间数据仓库在满足数据的完整性和一致性要求之外,还必须能够对空间数据进行组织和管理,支持空间查询和空间运算分析等操作. 海洋地球物理空间数据仓库首先选择一组基本的空间几何要素(点、线、面、栅格等)对海洋地球物理空间数据对象进行映射和建模(如图 3 所示). 然后,将根据具体的应用主题,选择统一的空间标准(如坐标系统、投影方式、比例尺)、分割粒度和数据格式(如 ESRI Shape 格式)进行提取,并且通过空间数据引擎(SDE)存储到空间数据仓库中作为空间对象维进行管理,以 Shape_ID 作为主键. 多维数据模型中的属

性维表和事实表都是以关系表的形式存储在空间数据仓库中,事实表中的空间维信息通过外键与空间对象进行关联.如图 3 所示的面向海洋地球物理调查工作量统计应用主题的星型模型,测线维表中的空间对象实际上是通过 SDE 在空间数据仓库中进行存储和管理,事实表中的测线对象实际上是指向测线维表中的空间对象的外键. SDE 能够对空间对象的空间信息和拓扑结构进行管理,并且创建空间索引,便于进行空间查询、空间几何代数运算、空间统计分析等操作,可以快速的对事实表中的度量(“测线总长度”)值进行运算.

1.4 应用服务器

海洋地球物理空间数据仓库应用服务器主要提供空间查询、空间分析、空间数据挖掘以及空间可视化等工具.空间查询工具主要实现对空间数据仓库中的内容(基于“事实”表和空间维)以及分析结果的查询.空间分析工具主要实现对空间数据仓库中的多维数据模型进行分析和综合,常用的分析工具如上卷、下钻、交叉、过滤、切片等^[19].空间分析的结果通常为空间数据挖掘服务.空间数据挖掘(也称为空间知识发现)工具,主要从空间数据仓库中寻找空间对象或模式的潜在关联和趋势,从而进行评价和预

测,在地球物理数据处理和解释工作中有较广泛的应用前景^[20, 21].常用的空间数据挖掘方法主要有分类发现、聚类、关联、人工神经网络、模糊算法等.空间可视化工具主要将查询、分析和挖掘的结果通过可视化的方式进行清晰、直观地表达,便于信息的理解和发现.海洋地球物理数据可视化的表达方式主要有三维空间立体图^[22~24]、二维空间平面图^[25]、柱状图、饼状图、曲线图^[26]、统计表格等.

2 系统实现

基于上述的空间数据仓库设计方法,本文基于 ORACLE 数据库管理系统和 ArcSDE 空间数据引擎对海洋地球物理空间数据仓库系统进行了开发和建设.针对用户不同的可视化查询方式,分别基于点空间要素类型和线空间要素类型对海洋地球物理空间数据进行提取.经过提炼和转换,将数据格式和空间标准进行统一,基于星型模型在空间数据仓库中进行组织和管理.另外,基于 ArcEngine 设计和开发应用服务器,针对应用需求,实现面向空间数据仓库的空间分析、空间统计和可视化等操作.图 4 是海洋地球物理空间数据仓库的地图查询界面,用户可以直接在地图上浏览空间数据仓库中各种专业要素类

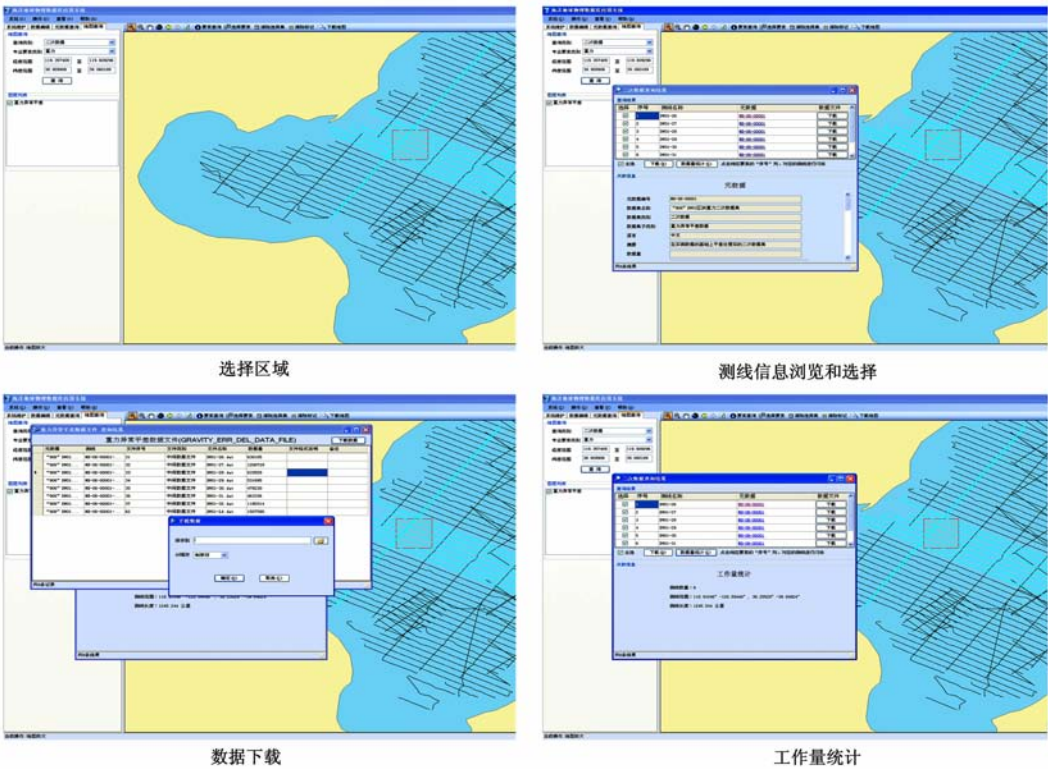


图 4 海洋地球物理空间数据仓库应用界面

Fig. 4 Application of marine geophysical spatial data warehouse

别的海洋地球物理测线分布情况,然后选择目标区域,获取所在目标区域的测线基本信息,用户可以选择感兴趣的测线下载相关的数据,同时还可以对选中测线的工作量进行统计.由于海洋地球物理空间数据仓库已经基于空间测线对海洋地球物理数据进行了重新组织和管理,建立了空间索引机制,因此,上述空间查询和空间统计操作的执行速度较快,而且代码编写的工作量也比传统的业务化数据库减少很多.

3 结 论

海洋地球物理调查和研究的对象具有空间特征,因此面向空间数据的组织、管理、查询、分析和表达等操作必不可少,空间数据仓库正是针对这一应用需求提出并发展起来的.本文将空间数据仓库技术引入到海洋地球物理空间数据的管理和应用中,提出了一套建设海洋地球物理空间数据仓库系统的设计方法,并且对系统进行了开发和实现.通过应用和实践,相比较传统的数据库而言,海洋地球物理空间仓库系统在空间数据的管理、检索和分析等方面,其开发效率和执行效率都有明显的提高,适应今后日益增长的海洋地球物理数据的空间应用需求.因此,应该进一步将空间数据仓库技术在包括海洋地球物理在内的海洋科学研究领域进行推广和应用,更好地为海洋科研和管理工作提供空间统计分析以及辅助决策支持服务.

当然,本文设计和开发的海洋地球物理空间数据仓库系统只是空间数据仓库技术与海洋科学研究相结合的初步探索和实践,还有许多关键技术问题有待解决,需要我们在今后的应用实践中进行更加系统深入的探索和研究.

参 考 文 献 (References):

- [1] Brackett M H. The data warehouse challenge: taming data chaos [M]. New York: John Wiley & Sons Inc, 1996. 579.
- [2] Inmon W H, Hackathorn R D. Using the Data Warehouse [M]. New York: John Wiley & Sons Inc, 1994. 285.
- [3] 邹逸江. 空间数据仓库研究综述[J]. 测绘学院学报, 2002, 19(4): 287~289.
Zou Y J. The outline on research of spatial data warehouse [J]. Journal of Institute of Surveying and Mapping. (in Chinese), 2002, 19(4): 287~289.
- [4] Bédard Y, Han J W. Fundamentals of spatial data warehousing for geographic knowledge discovery in Geographic Data Mining and Knowledge Discovery, Second Edition, edited by Harvey J. Miller and Jiawei Han [M]. Florida, US: Taylor & Francis Group, 2009. 45~68.
- [5] 周炎坤,李满春. 大型空间数据仓库初探[J]. 测绘通报, 2000, 8:22~23.
Zhou Y K, Li M C. A study of large spatial data warehouse [J]. Bulletin of Surveying and Mapping. (in Chinese), 2000, 8:22~23.
- [6] 杨 群, 闫国年, 陈钟明. 地理信息数据仓库的技术研究[J]. 中国图像图形学报, 1998, 4A(8): 621~626.
Yang Q, Lv G N, Chen Z M. Studies on the technology of geographical information data warehouse [J]. Journal of Image and Graphics. (in Chinese), 1998, 4A(8): 621~626.
- [7] 邹逸江. 多维空间分析的关键技术——空间数据立方体[J]. 地理与地理信息科学, 2006, 22(1):12~16.
Zou Y J. The key technology of multi-dimension spatial analyses: spatial data cube [J]. Geography and Geo-Information Science. (in Chinese), 2006, 22(1):12~16.
- [8] 田扬戈, 边馥苓. 空间数据仓库的 ETL 研究[J]. 武汉大学学报:信息科学版, 2007, 32(4):362~365.
Tian Y G, Bian F L. ETL of spatial data warehouse [J]. Geomatics and Information Science of Wuhan University. (in Chinese), 2007, 32(4):362~365.
- [9] 李振华, 胡光道, 王淑华. 一个地学数据仓库的初步设计与实现[J]. 地质与勘探, 2002, 38(5):67~70.
Li Z H, Hu G D, Wang S H. Preliminary design and development of a geological data warehouse [J]. Geology and Prospecting. (in Chinese), 2002, 38(5):67~70.
- [10] 刘展, 李津, 杨作升, 等. 适于“数字海底”的 GIS 数据仓库模型初探[J]. 西安石油大学学报(自然科学版), 2004, 19(1): 77~81.
Liu Z, Li J, Yang Z S, *et al.* Model of GIS data warehouse for “digital seabed” [J]. Journal of Xi'an Shiyou University (Natural Science Edition). (in Chinese), 2004, 19(1):77~81.
- [11] 于焕菊, 谢传节, 李云岭等. 中国华北地区地震空间数据仓库的构建与分析[J]. 地球信息科学, 2006, 8(3):88~93.
Yu H J, Xie C J, Li Y L, *et al.* Study of building the seismic spatial data warehouse of North China [J]. Geo-Information Science. (in Chinese), 2006, 8(3):88~93.
- [12] 宋国玺, 张晓华, 肖克炎, 等. 基于 ArcIMS 和 ArcSDE 的全国固体矿产资源远景区带数据共享平台建设[J]. 地球物理学进展, 2009, 24(3):1069~1074.
Song G X, Zhang X H, Xiao K Y, *et al.* Distributing on the web database of Chinese mineralogenetic belt of mineral resource based on ArcIMS and ArcSDE [J]. Progress in Geophys. (in Chinese), 2009, 24(3): 1069~1074.
- [13] Morrison R, Low B, Kucera H, *et al.* A Spatial Data Warehouse for NFIS [A]. In: Proceedings of the 13th Annual Conference on Geographic Information Systems (GIS 99)[C], Vancouver, BC, Canada, 1999. 242~247.
- [14] Pecar-Ilic J, Ruzic I. Application of GIS and Web technologies for Danube waterway data management in Croatia [J]. Environmental Modelling & Software, 2006, 21

- (11):1562~1571.
- [15] Nilakanta S, Scheibe K, Rai A. Dimensional issues in agricultural data warehouse designs [J]. Computers and Electronics in Agriculture, 2008, 60(2):263~278.
- [16] Zhou W, Chen G, Li H, *et al.* GIS application in mineral resource analysis—A case study of offshore marine placer gold at Nome, Alaska [J]. Computers & Geosciences, 2007, 33(6):773~788.
- [17] Rai A, Dubey V, Chaturvedi K K, *et al.* Design and development of data mart for animal resources [J]. Computers and Electronics in Agriculture, 2008, 64(2):111~119.
- [18] Shekhar S, Chawla S 著, 谢昆青, 马修军和杨冬青等译. 空间数据库[M]. 北京:机械工业出版社, 2004.
- Shekhar S, Chawla S. Spatial Databases: A Tour [M]. Beijing: China Machine Press, 2004.
- [19] Han J W, Kamber M. Data Mining: Concepts and Techniques, 2nd edition [M]. San Francisco, CA, USA: Morgan Kaufmann Publishers, 2006. 105~115.
- [20] 孟照旭, 李洪奇, 李雄炎, 等. 利用数据挖掘技术识别深层火山岩气层. 地球物理学进展, 2009, 24(6):2208~2214.
- Meng Z X, Li H Q, Li X Y, *et al.* Applying data mining techniques to identify deep volcanic gas reservoir. Progress in Geophysics. (in Chinese), 2009, 24(6):2208~2214.
- [21] 李雄炎, 李洪奇. 数据挖掘技术在石油天然气勘探领域的应用探索. 地球物理学进展, 2009, 24(5):1807~1813.
- Li X Y, Li H Q. Application of the data mining techniques in exploring the oil and natural gas. Progress in Geophysics. (in Chinese), 2009, 24(5):1807~1813.
- [22] 陈军, 葛瑛芳, 阎洪军. 复杂地层三维可视化显示的实现[J]. 地球物理学进展, 2009, 24(1):321~325.
- Chen J, Ge Y F, Yan H J. Realization of displaying 3D visualisation/or complicated strata [J]. Progress in Geophysics. (in Chinese), 2009, 24(1): 321~325.
- [23] 祁民, 张宝林, 梁光河. 高密度电法的三维数据场可视化[J]. 地球物理学进展, 2006, 21(3):981~986.
- Qi Min, Zhang B L, Liang G H. The visualization of three dimension data fields of high-density electrical technique [J]. Progress in Geophysics. (in Chinese), 2006, 21(3):981~986.
- [24] 汤井田, 公劲酷. 基于 GiD 的有限元前处理和后处理可视化 [J]. 地球物理学进展, 2010, 25(1):186~195.
- Tang J T, Gong J Z. Visualization of the pre-and post-processing in finite element analysis based on GiD [J]. Progress in Geophysics. (in Chinese), 2010, 25(1):186~195.
- [25] 宋国玺, 张晓华, 肖克炎, 等. 基于 ArcIMS 和 ArcSDE 的全国固体矿产资源远景区带数据共享平台建设[J]. 地球物理学进展, 2009, 24(3):1069~1074.
- Song G X, Zhang X H, Xiao K Y, *et al.* Distributing on the web Database of Chinese mineralogenetic belt of mineral resource based on ArcIMS and ArcSDE [J]. Progress in Geophysics. (in Chinese), 2009, 24(3):1069~1074.
- [26] 林家勇, 汤井田, 丁茂斌, 等. 面向对象技术在地球物理电磁成像可视化中的应用[J]. 地球物理学进展, 2009, 24(6):2233~2241.
- Lin J Y, Tang J T, Ding M B, *et al.* Application of object-oriented technology in geophysical electromagnetic visualization [J]. Progress in Geophysics. (in Chinese), 2009, 24(6):2233~2241.