

多源异构海量海洋数据综合管理平台构建研究

张明华, 黄冬梅, 熊中敏, 郑小罗, 何盛琪

(上海海洋大学 信息学院, 上海, 201306)

摘要: 为满足海洋应用对数据管理的需求, 针对海洋数据的特点, 采用分层的思想提出了一个海洋数据综合管理平台的体系结构, 并对海洋数据的共享, 构建海洋元数据和海洋数据安全性管理等关键问题分别给出解决方案。通过一个基于 B/S 的多源异构海量海洋数据综合管理平台的实现表明本文的方法是可行且有效的。

关键词: 海洋数据综合管理平台; 海洋元数据; 数据共享

中图分类号: TP319 文献标识码: A 文章编号: 1000-3096(2012)02-0110-06

地球上 70% 的面积被海洋所覆盖, 海洋不仅对人类生存的自然环境起着十分重要的调节作用, 而且是社会经济发展新的增长点, 世界上经济文化发达的地区大多分布在海岸线附近, 海洋这一巨大的天然资源开始受到世界各国越来越多的关注。我国海岸线漫长, 海洋资源丰富。为更透彻地了解海洋, 合理利用海洋, 促进海洋经济的健康、可持续性发展, 国家近年来相继开展了一系列重大的海洋专项调查和研究活动, 在此过程中获得许多宝贵的海洋数据。如何有效地分析、展示、应用和管理海洋数据是海洋信息化技术中的重要研究方向, 也是构建海洋应用系统的基础。

文献[1-4]分别从海洋数据应用模式、可视化、分析和组织等多个方面研究了海洋数据的管理和利用。文献[5]给出了南海海洋科学数据库数据管理和发布系统的设计实现方法。这些研究大多是针对某种海洋数据, 实现的海洋数据管理系统功能较为单一, 对海洋数据的管理缺乏系统性。本文在分析海洋数据特点的基础上, 面向海洋应用, 设计并实现一个海洋数据综合管理平台, 能够支持对海洋数据的存储、管理、访问和应用, 并给出了其中的关键问题的解决方法。

1 需求分析

首先分析海洋数据具有的特点, 然后归纳了海洋应用对海洋数据管理平台的的需求, 从而为海洋数据管理平台的设计奠定基础。

1.1 海洋数据的特点

在设计适合海洋应用的数据管理平台之前, 需

要全面分析和了解海洋数据的特点, 它们是构建海洋数据综合管理平台的基础。海洋数据的特点表现在: (1)海洋数据具有多元、多源性。海洋数据内容非常广泛, 涉及物理、化学、生物、环境等多个领域和学科, 且各类数据之间相互关联。此外, 还有多种比例和要素的海洋基础地理数据, 海洋遥感影像数据以及海洋专题信息产品数据, 数据类型和格式多样。随着海洋经济时代的到来, 海洋数据体系中又增加了关于海洋科技、海洋执法、海洋社会经济等方面的数据, 扩大了海洋数据的来源, 也给海洋数据服务提出更高的要求。(2)海洋数据具有时序性和海量性。与陆地很大的不同是, 海洋时刻处于一个动态变化的过程中, 它和大气、陆地密切相关, 因此海洋环境基础数据常表现为时间序列数据, 数据量特别巨大, 而且正以指数级速度增长。(3)海洋数据的异构性。由于海洋数据的多源性、海量性, 数据一般采用分级管理, 集中和分布相结合的存储方式。各地经济、科技、管理水平发展的不平衡性, 造成了海洋数据的异构性。不同的单位会根据海洋数据的来源和数据类型选用不同的数据存储形式, 如传统的关系数据库(RDB)、Excel 文件、Word 文件或者文本文件。这给原始观测数据的使用和交换带来了一定的困难。(4)海洋数据的标准化和机密性。由于海洋数据的多元性, 数据的测量方法、仪器设备也多种多样,

收稿日期: 2010-03-31; 修回日期: 2010-08-12

基金项目: 上海高校选拔培养优秀青年教师科研专项基金(ssc09015); 海洋公益性行业科研专项(200905014-06)

作者简介: 张明华(1977-), 女, 河南郑州人, 讲师, 博士, 主要从事海洋信息技术研究, 电话: 021-61900622, E-mail: mhzhang@shou.edu.cn

从而存在许多数据格式、比例尺和坐标系统。为实现数据共享和交换，海洋数据的标准化工作显得尤为重要。另一方面，海洋资源是国家的一种战略资源，而数据是获取海洋资源和信息的基础和依据，在提供海洋数据公共服务的同时，更要注重海洋数据的机密性和完整性保护。

1.2 功能需求

由于海洋信息采集分散且工作量大、数据量庞大、分布式存放管理且管理成本高、共享使用成本低，为了使应用系统能有效地访问所需的海洋数据，数据综合管理平台应以实现海洋信息服务为目标，提供如下主要功能：(1)一体化组织和管理功能。目前，我国海洋信息相对分散，一体化组织和管理水平较低，标准一致的数据及数据产品较为匮乏，无法满足快速发展的海洋信息化对海洋基础信息的需求，为形成统一信息平台，需要对分散的信息进行整合及标准化改造；同时，海洋信息状况与各部门业务应用的需求存在差别，分散的、不规范的信息难以满足业务需求。(2)数据整合、集成处理服务功能。由于涉及的信息内容广泛，数据内容及形式复杂多样，因而综合管理平台需要具有针对不同类型的信息进行整合处理的能力，在现有信息资源的基础上形成可以共享使用的标准化、规范化信息，对结构化的规范数据和非结构化的规范数据提供查询、统计、分析功能，以及插

入、更新、删除等管理功能。(3)共享与交换功能。数据综合管理平台的数据需要与应用系统进行内部信息交换，为各个应用系统的正常运行提供数据支撑，同时将系统中产生的需要保存的分析数据和成果数据通过数据综合管理平台提供的数据共享和交换功能保存到相应的数据库中。(4)安全保密功能。必须采取各种有效措施加强海洋基础数据的安全保障体系，使海洋基础信息领域里的各种机密数据及其相应地理空间信息得到高效、灵敏和强有力的保护。

2 海洋数据综合管理平台系统设计

在需求分析的基础上，提出一个海洋数据综合管理平台的设计方案，详细给出了该管理平台的逻辑结构设计和功能设计。

2.1 逻辑结构设计

针对海洋数据的特点，结合海洋专题应用对海洋数据访问，数据服务需求不断提高的要求，本文给出了海洋综合数据管理平台的逻辑结构，如图 1 所示。整个平台的体系结构采用分层的思想，由下至上设计为数据准备层、数据基础层、数据管理层、数据服务层和数据应用层等 5 层，数据安全和数据标准化则贯穿于所有层次。

数据准备层从网络传输、软硬件资源方面为数据的收集做好准备，主要包括网络通信与传输系统，

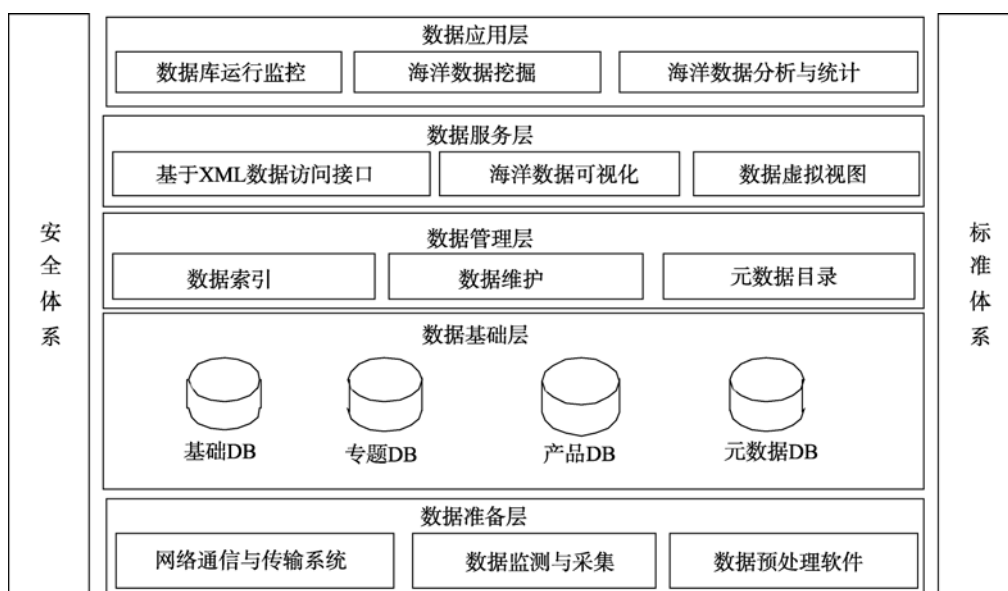


图 1 海洋数据综合管理平台体系结构

Fig. 1 Architecture of an integrated ocean data management platform

数据监测与采集系统, 数据预处理软件系统。

数据基础层是数据管理平台的数据源, 负责海洋基础数据库、海洋专题数据库和海洋信息产品数据库的建库和数据入库。由于海洋数据种类多样, 涉及的调查单位多而分散, 数据库也是依实际情况分级部署, 既有集中式管理的数据库, 又有在各地的分数据库。为了使应用系统有效地访问分布的各类数据, 数据基础层还负责从各数据库中提取相关元数据, 构建一个集中的海洋元数据库。

数据管理层是数据综合管理平台的核心, 主要实现数据的高效组织和管理, 保障数据的完整性和一致性。通过建立数据索引, 提高数据检索速度。实现海洋数据的自动化导入、导出、更新和数据质量控制。构建海洋元数据目录, 提供元数据分类浏览和查询功能。

数据服务层面向上层应用和外部用户的需求, 将与数据相关的公共功能以服务的形式发布, 并支持对数据服务的注册、发布和查找。数据服务层使

得综合管理平台可以方便地添加其他的服 务, 增强了数据管理平台的可扩展性。基本的数据服务包括, 为应用提供统一的数据视图, 用户不需要知道实际数据分布在什么位置。提供基于 XML 的数据访问接口。实现海洋数据的二、三维展示等。

数据应用层主要提供一些代表性的海洋数据应用, 如海洋数据挖掘, 海洋数据分析与统计以及海洋数据库运行实时监控等。

2.2 功能设计

海洋数据综合管理平台的功能如图 2 所示, 分为数据管理, 数据分析, 共享与交换, 安全管理和系统管理 5 个模块。

数据管理模块以树形结构组织被管理的各种海洋数据, 包括基础地理数据、综合专题数据、应用数据、海洋产品信息数据和元数据, 支持对数据的浏览、多条件模糊查询和增、删、改操作, 还可以对整个平台的数据库运行情况 进行实时动态监控。

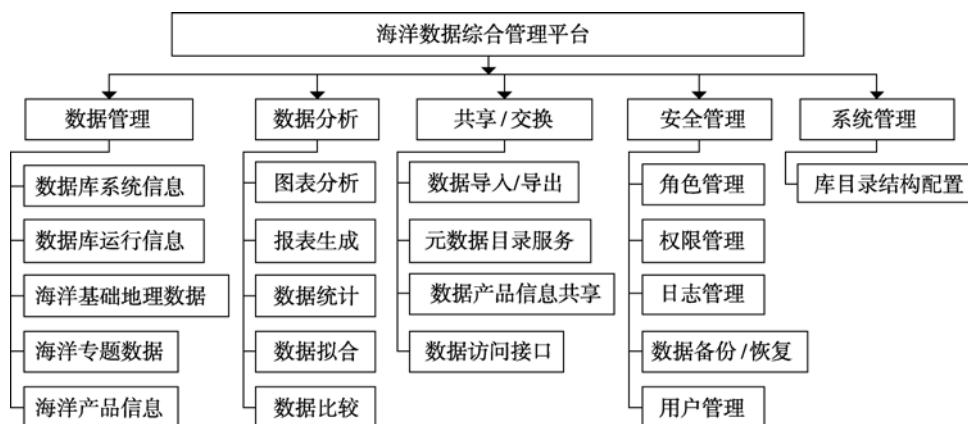


图 2 海洋数据综合管理平台功能结构

Fig. 2 Functions provided by the integrated ocean data management platform

数据分析模块提供了灵活的用户参数设置和图形化的数据展示。用户选择想要查看的数据, 例如数据名称, 数据获得的年月或观测台站等, 系统根据这些信息访问数据库获得数据并绘制成图。数据比较使用户可以对两个不同观测日的数据做曲线拟合后在同一坐标系下观察对比, 分析数据的变化情况。

共享与交换模块包括数据导入导出, 元数据目录服务, 海洋数据产品信息共享, 数据访问接口等功能。通过元数据目录服务可以有效地查找到需要

的数据。基于 Web Service 技术的数据访问接口为海洋应用程序提供一个安全、统一的数据访问和共享方式。

安全管理模块通过用户管理, 角色管理, 权限管理, 日志管理, 数据的备份/恢复等功能来保证系统的安全性。角色、权限管理的设计是基于 RBAC 的层次管理模型, 采用用户-角色-权限-模块 4 级管理模型, 提供集成操作权限、系统表权限、栏目功能权限和数据表权限的管理, 以满足多层次组织中权限的集中控制。

系统管理模块提供对数据库目录树结构的维护,包括数据表维护,数据表层次维护,数据表关联等操作。

3 关键技术

讨论海洋数据综合管理平台实现中的几个关键问题,包括海洋数据共享、海洋元数据的设计和海洋数据的安全管理。

3.1 基于存储资源代理的数据网关

海洋数据多源、异构,从地理上看通常分布在多个不同的位置,采用不同的数据库管理系统进行管理,如 SQL Server、Access、Oracle 等。数据管理平台的服务层基于这种现状,通过数据服务的形式,使外部应用系统在访问数据时不需要考虑上述不同,能够以统一的接口和方式获取或访问数据。本文采

用基于存储资源代理 (Storage Resource Broker, SRB)^[6]的数据网关来实现服务层的上述功能。数据网关的结构如图 3 所示。

存储资源代理 SRB 是一个基于 C/S 模式的中间件,它为用户提供一个访问文件系统、档案系统、数据库系统等多种异构存储系统的统一接口,屏蔽了存储系统的异构特性。本文基于 SRB 的数据网关可以看成是一种轻量级的存储资源代理系统。SRB 完成的主要功能有:元数据目录的维护;数据访问请求的解析;基于元数据目录的数据源映射;数据检索和获取;结果数据的封装。元数据目录服务主要提供元数据的查询、浏览、数据视图映射等功能,便于用户查询、检索与快速定位(数据源映射)所需查找的海洋数据。用户可以通过浏览器浏览或查询元数据目录,获取数据集。应用系统可以通过 Web 服务与 SRB 交互,通过元数据目录提供的接口访问各种海洋数据库的数据。

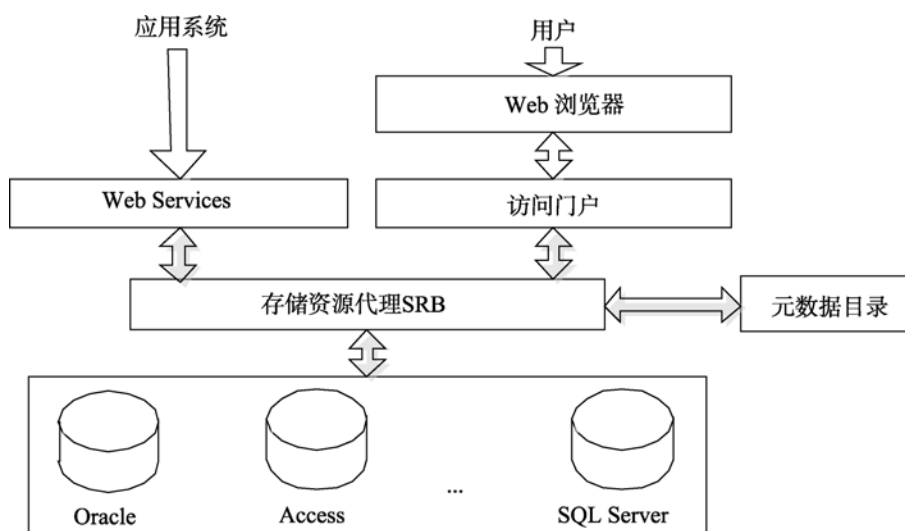


图 3 数据网关的结构

Fig. 3 Architecture of data gateway

3.2 海洋元数据设计方案

元数据是关于数据的数据,其目的在于提供一个中间级别的描述,使得人们据此就可以做出选择,确定哪些是想要浏览或检索的信息包,而无需检索大量不相关的数据。海洋信息学科众多,种类复杂,海洋元数据的建立对于加强信息的共享和交换,异构、多平台数据的访问,统一数据存储交换格式,提高信息的有效利用率具有重要的作用,是海洋数据管理平台底层建设中的一个重要组成部分。

目前,海洋信息还没有统一的海洋数据的元数据标准。本文在研究相关元数据标准,如国际标准 ISO 19115^[7],国家标准 GB/T 19710-2005 地理信息数据,中国科学院科学数据库核心元数据标准^[8],海洋环境数据目录 MEDI^[9]等的基础上,结合海洋数据的特点和海洋元数据应用的要求,将海洋信息元数据划分为两个层次:(1)是数据库层次的元数据,以 26 类基础数据库及 13 类信息产品库为对象描述的元数据,即数据管理平台中总的元数据目录,例如包括海洋基础空间地理数据库、海洋环境基础数据库、

海洋综合管理数据库和信息产品库；(2)是数据集层次的元数据，它以数据库为组织单元，对该数据库内数据集进行描述。例如海洋环境基础数据库中包括水文、气象、台站等数据集。

元数据的内容分为空间数据的元数据和属性数据的元数据两部分，包括对数据集的描述；对数据

集中各数据项、数据来源、数据所有者及数据序列(数据生产历史)等的说明；对数据质量的描述，如数据精度、数据的逻辑一致性、数据完整性、分辨率、元数据的比例尺等；对数据处理信息的说明，如量纲的转换等；对数据转换方法的描述；对数据库的更新、集成方法等的说明。如图 4 所示。

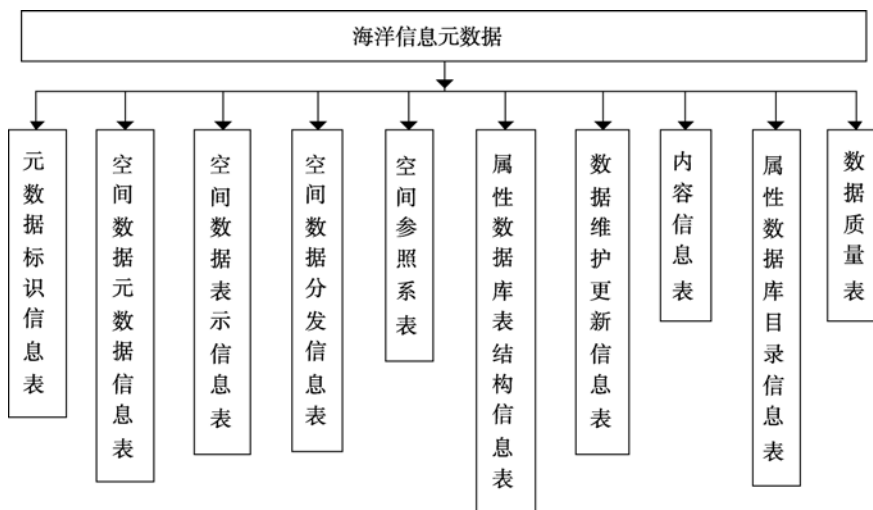


图 4 海洋信息元数据

Fig. 4 Metadata of ocean information

3.3 海洋数据的安全管理

海洋信息数据量大，数据的获取和预处理需花费大量的人力、物力和财力，而且这些数据蕴含了我国海洋资源的重要信息和知识，因此具有很高的密级，海洋数据的安全性保障也是海洋数据管理平台建设中的一项关键性工作。本文在一个安全数据库原型^[10]的基础上，结合用户对数据的安全性层次需求，将较高密级的数据存入安全数据库，其余数据存放在成熟的大型数据库 Oracle 中，从而兼顾到数据的安全保护和存取访问效率。通过对数据库系统的数据存储、传输和访问控制的全过程改造，即采用强制访问控制、敏感数据的加密存储、基于 SSL 协议的数据传输、三权分立的权限管理、数据库审计，实现全方位的安全控制。安全数据库中的数据访问采用基于 Oracle 虚拟视图的方式，为用户呈现一致的接口，提高系统的易用性。

4 系统实现

本文在上述分析和研究的基础上，选择

Microsoft Visual.net 为开发平台，C#为开发语言，Oracle 为后台数据库管理系统，以 Microsoft IIS 为应用服务器，实现了一个基于 B/S 的多源异构海量海洋数据综合管理平台，为各类海洋应用系统的建设提供数据支撑。图 5 是海洋数据综合管理平台登陆后的主界面。

5 结束语

海洋里蕴藏着大量丰富的资源，在陆地资源日益减少的情况下海洋资源显得更加珍贵，如何对海洋进行可持续地开发、利用和保护受到越来越多国家的重视，而海洋信息化是发展海洋的有效途径。本文对海洋信息化的基础，即多源、异构、海量海洋数据的管理进行研究，提出了一个适应海洋数据特点和海洋应用需求的海洋数据综合管理平台的体系结构，对其中的海洋数据共享，海洋元数据构建和海洋数据的安全性等关键问题进行探讨，实现了一个多源异构海量海洋数据综合管理平台。



图 5 海洋数据综合管理平台的主界面

Fig. 5 Main interface of the integrated ocean data management platform

参考文献:

[1] 王晓民, 张新, 池天河. “数字海洋”的数据处理与应用模式研究[J]. 计算机应用, 2008, 28(z1): 358-359.

[2] 高存彬, 秦勃, 洪峰. 基于 k-means 聚类的水团划分可视化算法[J]. 计算机应用, 2008, 28(z1): 360-363.

[3] 覃如府, 叶娜, 许惠平, 等. GIS 系统中多维海洋数据可视化研究[J]. 同济大学学报(自然科学版), 2009, 37(2): 272-276.

[4] 薛惠芬, 周燕遐. 数据仓库技术在海洋环境信息管理中的应用研究[J]. 海洋通报, 2005, 24(3): 66-72.

[5] 黄林丛, 李莎. 南海海洋科学数据库数据管理和发布系统的设计与实现[J]. 热带海洋学报, 2006, 25(1): 71-75.

[6] 王意洁, 肖依, 任浩, 等. 数据网格及其关键技术研究[J]. 计算机研究与发展, 2002, 39(8): 943-947.

[7] ISO/FDIS 19115, Geographic information – metadata[S].

[8] 中国科学院计算机网络信息中心科学数据库中心. 中国科学院科学数据库生态研究元数据标准[M]. 北京: 中国科学院计算机网络信息中心, 2003.

[9] 薛惠芬. 国际上几种海洋元数据内容剖析[J]. 海洋信息, 2004, 3: 25-28.

[10] 应乐年. 安全数据库多级安全模型策略与实现研究[D]. 上海: 复旦大学, 2008.

Construction of an integrated management platform for multi-dimension heterogeneous and massive ocean data

ZHANG Ming-hua, HUANG Dong-mei, XIONG Zhong-min, ZHENG Xiao-luo, HE Sheng-qi

(Information College, Shanghai Ocean University, Shanghai, 201306, China)

Received: Mar., 31, 2010

Key words: integrated management platform for ocean data; ocean metadata; data sharing

Abstract: To meet the requirements of data management in ocean applications, an integrated data management platform with a layered structure is proposed to accommodate the feature of ocean data, solutions to some key problems including the sharing of ocean data, construction of ocean metadata, and the ocean data security management has been given. An implementation of the B/S based management platform for multi-dimension heterogeneous and massive ocean data has demonstrated that the method is feasible and efficient.

(本文编辑: 刘珊珊)