

doi:10.3799/dqkx.2013.000

国家级地质环境数据仓库的设计与实现

张鸣之¹, 喻孟良¹, 王 勇¹, 李振华², 梅红波^{3*}, 吴湘宁²,
肖敦辉³, 胡光道³, 谭照华⁴

1. 中国地质环境监测院信息室, 北京 100081

2. 中国地质大学计算机学院, 湖北武汉 430074

3. 中国地质大学资源学院, 湖北武汉 430074

4. 三峡库区地质灾害防治工作指挥部, 湖北宜昌 443000

摘要: 目前由中国地质环境监测院具体实施的国家级地质环境信息化建设工作, 以构建国家级地质环境数据中心, 实现全国地质环境信息的大综合、大协作和大集成为总体目标。其数据来源多样、数据类型复杂, 且体量巨大、结构性和非结构性数据并存。为实现快速、充分地挖掘提炼数据中蕴含的价值, 基于大数据思想, 运用数据仓库技术, 将各类操作型数据面向业务分析整合, 形成协同有序的数据中心, 按业务进一步划分为地质灾害防治、地下水监测与保护、矿山地质环境监测评价等4个数据集市, 每个集市下又包含若干主题或二级主题。应用该数据仓库横向打通了各业务系统的数据, 实现了不同维、不同粒度、不同侧面查询和观察数据的功能, 为业务分析和决策支持提供了数据保障, 为捕获地质环境业务数据价值提出了新模式。

关键词: 大数据; 数据仓库; 数据集市; 数据挖掘; 地下水; 地质环境评价; 地质灾害。

中图分类号: P641.8

文章编号: 1000-2383(2013)06-1347-09

收稿日期: 2013-02-26

Designing and Building the National Geo-Environment Monitoring Data Warehouse

ZHANG Ming-zhi¹, YU Meng-liang¹, WANG Yong¹, LI Zhen-hua², MEI Hong-bo^{3*},
WU Xiang-ning², XIAO Dun-hui³, HU Guang-dao³, TAN Zhao-hua⁴

1. Information Center of China Geological Environmental Monitoring Institute, Beijing 100081, China

2. School of Computer Science, China University of Geosciences, Wuhan 430074, China

3. Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China

4. Headquarters of Geological Hazard Control in Area of the Three Gorges Reservoir, Yichang 443000, China

Abstract: China Institute of Geo-Environment Monitoring is building a national-level geological environment information platform, as an important part of a national data center aiming to collect, integrate and analyze the national geological environment data characterized by its huge size, multi-source, both structured and unstructured storage. In order to extract valuable information from the data center rapidly, a data warehouse based on the concept of big data is applied to integrate operational data into data center for professional analysis. The system running results show that the data warehouse, which consists of 8 data marts (geo-hazard prevention, underground water monitoring and protection, mining geology environment and etc) and each of them can be divided successively into various subjects and sub-subjects by loading data into star schema in multiple dimensions, different levels of granularities and various perspectives, can provide multi-dimensional data analysis and support decision-making, and therefore presents a new strategy on discovering valuable information from geo-environmental data.

Key words: big data; data warehouses; data mining; groundwater; geo-environment assessment; geological hazard.

大数据时代正在来临(White, 2009; Mahoney, 2012), 数据仓库作为一种比较成熟的数据集

基金项目: 中国地质调查局“环境地质调查数据集成与服务系统建设”项目(No. 1212011220353); 中国地质环境监测院“国家级地环节点系统地质环境数据仓库及数据仓库管理系统研究”项目(No. 2013194013)。

作者简介: 张鸣之(1981—), 工程师, 主要从事水工环信息化、通信与信息系统研究。E-mail: zhangmz@mail.cigem.gov.cn

* 通讯作者: 梅红波, E-mail: hbmei@cug.edu.cn

成和分析手段将在大数据的利用中得到广泛应用. 地学数据是一个典型的大数据实例, 并且比一般的商业数据远为复杂, 因此, 开展地学数据仓库的研究, 对于促进地学大数据的研究与发展具有重要意义.

自 1993 年 Inmon(1992)系统地提出了数据仓库的定义和设计方式后, 数据仓库在工商领域得到了长足的发展, 在金融、电信、保险等主要传统数据处理密集型行业的应用尤为广泛. 到 1996 年, 美国联邦地理数据委员会开启了框架示范工程计划(Framework Demonstration Projects Program, FD-PP), 成为美国空间数据基础设施的组成部分, 自此, 数据仓库开始应用于地理领域, 如 1997 年美国科技信息服务局完成了一项空间数据仓库的示范项目.

但由于地学数据远比地理数据复杂, 且空间特性更为明显, 数据仓库的成功应用还比较少, 并且其对地学数据仓库的认识还有待加强. 目前国际上比较著名的有 2 个数据仓库: (1) 加拿大水文地理局的海洋深度数据仓库(Forbes, 1999), 采用 Oracle 关系数据库及其空间数据管理技术, 面向 TB 级的数据存储与管理; (2) 美国国家水质评价(NAWQA)(Cohen, 1999; Rosen and Lapham, 2008)数据仓库, 它从 1991 年开始收集全美境内的地质水质数据, 并对数据的动态趋势进行评价. 国内著名的数据仓库是大型数字海洋和地理信息系统项目“数字南海”, 它的仓库包括 50 多种专题数据源, 几十种数据格式, 涉及到南海数据资料收集、集成、存储等各个环节, 目的在于实现南海基础地理信息数据和海洋专题数据的统一管理, 为海域使用管理、海洋功能区规划、海洋资源调查与海洋环境监测等提供可靠的信息(温伟, 2007). 从上述 3 个国内外的已建成的地学数据仓库来看, 它们沿用商业数据仓库的设计思路较多, 数据仓库功能在商业数据库基础上拓展不多, 考虑地学数据自身的特点较少, 特别是数据类别比较单一.

作者在全国地质环境信息化体系建设中负责数据中心总体设计和多级网络体系构建等核心内容, 认识到地学数据的复杂性与特殊性, 并认为地学数据仓库应摒弃商业数据库中基于时间的数据组织形式, 改用基于空间的数据组织形式, 用比例尺作为粒度级别代替传统的时间粒度级别, 这在项目组负责的“十五”国家攻关项目“金属矿产资源快速评价预测系统研究”中有过一定的研究基础. 基于此认识, 作者提出了地学数据仓库的初步框架(胡光道,

1999; 李振华等, 1999); 2002—2004 年, 在理论上继续探索了地学数据仓库的发展历史、现状和存在问题(Li *et al.*, 2003), 在实践上以云南省数据为例开发了一个地学数据仓库原型系统, 该系统面向矿产资源预测, 涵盖了地质、物探、化探、遥感、地理、矿产、钻孔(井)、成果等几乎全部地质数据, 兼容现有的所有地质数据的国家标准, 是一个提供多种数据格式的输入、输出、查询、管理等功能的地质信息系统(胡光道和李振华, 2002; 李振华等, 2002). 作者在 2009 年承担了国土资源部项目“三峡库区地质灾害预警指挥系统(GHPACS)数据仓库及管理系统建设”, 在此项目进行过程中, 地学数据仓库的设计和 实践得到了进一步的完善和体现, 所涉及的数据类型更加多样, 数据量也激增(胡光道等, 2011).

2012 年作者开始了“环境地质调查数据集成与服务系统建设”及“国家级地质环境数据仓库及数据仓库管理子系统研究”项目的研发, 项目要求从国家地质环境监测的角度, 整合和重组各省市的地质环境监测数据, 构建多维数据模型及数据仓库, 利用数据快速检索查询、联机分析处理及数据挖掘等工具, 为地质灾害预报预警、地质环境评估等需求提供决策支持服务.

1 系统需求与设计目标

1.1 系统需求

建立数据仓库的目的是为了将现有各系统各自为政的操作型数据进行面向分析的整合, 形成一个集成的一致性的数据中心, 直接为国家级地质环境信息平台决策分析服务. 它可以定义为整个决策分析系统的数据“调度中心”(或“中转站”).

不同于传统的商业数据仓库及通常意义上的地理数据仓库, 国家级地质环境信息平台的数据来源及应用背景更为复杂, 其难点在于集成多源、多格式(异构)、多类别、多尺度海量数据, 并提供给多个差异性较大的分析系统使用.

该数据仓库拟采取分区存储的数据组织形式, 实现数据的分门别类的分布式存储. 空间维以比例尺或行政区划作为维层次, 时间上以日、月、年等时间维层次, 属性上以属性隶属关系为层次, 实现多层次的数据存储; 不同维的层次对应不同的数据粒度. 多维数据的组织形式便于从多个观察角度对当前以及历史数据进行多角度、多层次灵活组合及分析, 采用下钻、上钻、旋转、切片操作, 供用户直观地

(5)发布相关信息为社会公众服务.考虑到各省市信息化建设差异,对于一些基础条件较差,建设较为滞后的省级节点,可由国家级地环节点提供相关的宏观性的信息服务.

2.2 系统逻辑模型

整个系统在逻辑上分为 3 层:原始数据层、中间逻辑层(业务逻辑层、WEB 服务层、安全服务层)和应用层,如图 2 所示.原始数据层:以统一规范的方式存储数据.中间逻辑层:解析应用层的业务逻辑,使应用层和原始数据相互独立,提高应用层系统(程序)的可扩展性、可移植性.应用层:面向最终用户,提供友好、简洁、方便的用户界面,具有良好的业务无关性.

2.3 项目设计边界

根据业务流程和数据流程,本项目具有上、下两个边界,从下边界提取数据,向上边界输出数据.下边界从操作数据库和空间数据库抽取数据.从操作数据库抽取属性数据并存入数据仓库;从空间数据

库提取空间信息模型,但不抽取具体空间数据,数据仓库及其上层应用系统通过数据仓库中的空间信息模型调取空间数据.上边界向决策分析系统提供数据支持.支持的数据主要有 2 种形式:通过联机分析形式提供多维数据模型,方便用户以不同粒度、不同层次、不同侧面对数据进行可视化的观测;通过数据挖掘方式,找寻数据规律,并通过历史数据的综合分析对现有模型和方法进行评价.

3 数据仓库建设设计

数据仓库建设部分主要描述各个集市以及主题的建设情况.此部分基本按照总体设计的要求进行设计,根据应用的实际需求,划分为地质环境决策、地质灾害防治、地下水监测与保护、地面沉降监测、矿山地质环境监测评价、地质遗迹保护、地质环境综合评价及资源环境承载力评估等 8 个集市(或称为

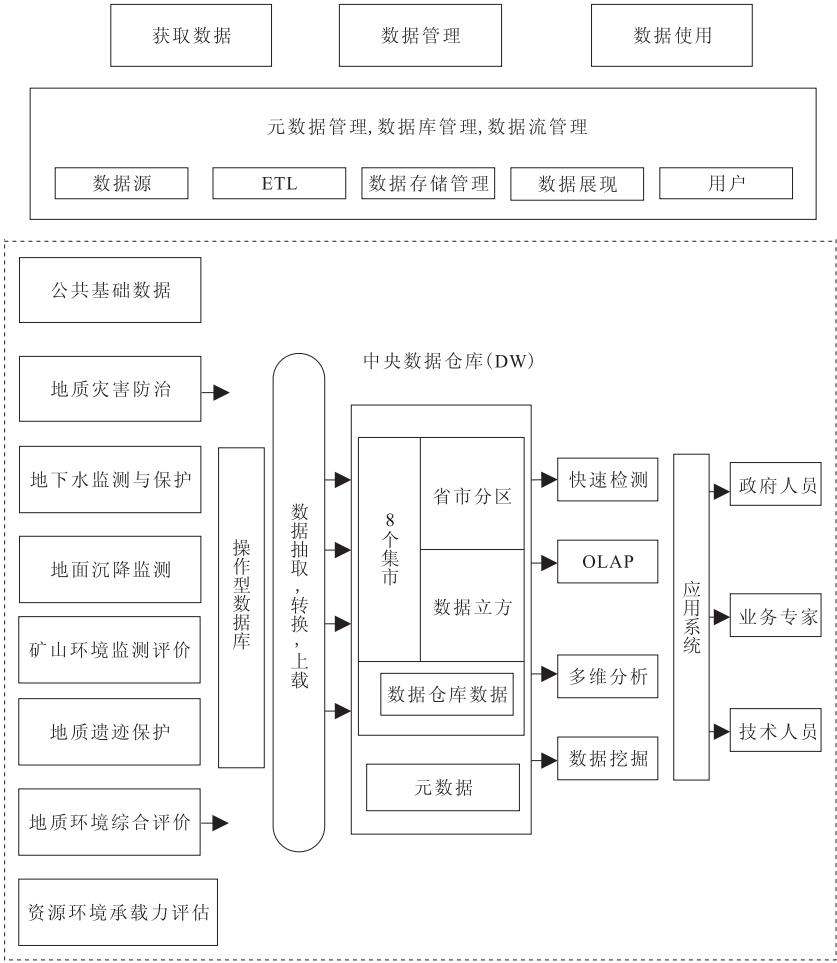


图 2 系统逻辑模型

Fig. 2 Logic model of the national geo-environment monitoring data warehouse

领域),每个集市下又包含若干主题或二级主题。

3.1 地质环境决策公共集市的建设

本集市存放公共数据,供各领域集市调用;本集市由基础地理底图、基础地质、水文地质、工程地质、环境地质等主题组成。(1)基础地理底图是指基础的地理空间框架数据,包括地形地貌、行政区划、居民地、交通、河流湖泊、数字高程模型等。(2)基础地质数据主要包括地层分区、地层界线、地质构造(褶皱、断层、节理构造)、地层产状等。(3)水文地质是研究地下水的分布和形成规律,地下水的物理性质和化学成分,地下水资源及其合理利用,地下水对工程建设和矿山开采的不利影响及其防治等的科学,其数据主要包括地下水类型、地下水富水性、地下水径流模数、地下水水质、水文地质特征点(界线)、地下水利用规划等。(4)工程地质是调查、研究、解决与人类活动及各类工程建筑有关的地质问题的科学,其数据主要包括:岩土组分、组织结构、物理、化学与力学性质、地下水运动、地表地质作用数据、建筑工程及其附属设施等。(5)环境地质是研究人类技术—经济活动与地质环境相互作用、影响的学科,其数据主要包括灾害地质、矿山与地下工程灾害、土地退化(水土流失、土地沙化、沙漠化、土地盐渍化、土地沼泽化)、水动力灾害(海水入侵、海岸线变迁、水库渗漏、水库淤积、河流淤积)、地下水位下降、地下水污染、特殊土灾害、地方病(地氟病、克山病、大骨节病、地甲病)等。

3.2 地下水监测与保护集市的建设

本集市由以下主题组成:全国地下水分区评价、主要盆地(平原)地下水动态评价、典型流域(城市)地下水开发与管理、特殊水文地质问题区水文地质研究等。(1)全国地下水分区评价。其按研究对象分为大盆地、平原、断陷盆地等小水文地质单元。按以下研究内容的相关参数建立数据立方:区域水文地质条件、含水层系统分布、地下水补给、径流、排泄特征、地下水位、含水层厚度等。(2)主要盆地(平原)地下水动态评价。按研究对象分类:地下水开发的流域、平原、小水文地质单元。按以下研究内容的相关参数建立数据立方:地下水位动态变化、超采、含水层厚度多年变化、资源潜力变化趋势等。(3)典型流域(城市)地下水开发与管理研究。按研究对象分类:地下水开发盆地、城市。按以下研究内容的相关参数建立数据立方:基础水文地质条件、资源(潜力)评价、开发利用规划、地下水管理模型等。(4)特殊水文地质问题区水文地质研究。按研究对象分类:地方

病、污染、海水入侵、石漠化地区。

3.3 地质灾害防治集市的建设

本集市由以下主题组成:稳定性评价、预测预报、气象预警、预警决策支持与应急指挥。(1)稳定性评价。其二级主题为:滑坡稳定性评价、危岩稳定性评价、高切坡稳定性评价、塌岸预测。各二级主题将直接采用“稳定性评价系统”产生的结果数据,按工况、预测方法、预测结果等建立相关立方。(2)预测预报。其二级主题为:地质灾害预测预报、单体地质灾害预测预报、涌浪预测、地质灾害灾情预评估。地质灾害预测预报二级主题选择适合地质灾害易发性预测的指标和参数:地质灾害易发性分区图、降雨、库水位、江河水位、变形监测数据、滑坡性质等建立相应数据立方;单体地质灾害预测预报二级主题选择灾害体基础地质数据、物理力学参数、监测数据、预测破坏概率、威胁范围、影响人口、威胁财产等,拟建立时间预报、稳定性评价、典型案例等相关立方;涌浪预测二级主题选择如下指标:滑坡和河道剖面以及相关岩土物理参数、入江方量、初始涌浪高度、距滑坡点距离、传播浪高度、爬坡浪高度等建立相关立方;地质灾害灾情预评估二级主题选择对地质灾害直接影响及灾害发生后涌浪破坏范围内可能影响的居民(户、人口)、土地、房屋、单位、公路、桥梁、高压线、通讯等各类重大设施具体情况及影响程度、可能造成的财产损失等建立数据立方,可能的话,辅助编制地质灾害危害范围及可能受损重大设施分布图。(3)气象预警。本主题的目的是将气象预报信息(指降水信息)与地质灾害点(体)分布信息进行综合分析,根据灾害点(体)降雨模型,绘制相关分析图件,建立地质灾害气象预警分析显示机制。本主题拟采用区域维、时间维、地灾属性维建立气象预报阈值、变形预测、预警级别等数据立方。(4)预警决策支持与应急指挥。本主题的目的是对出现险情(或灾情)的地质灾害及时作出响应,提出科学决策意见,支持地质灾害发现、鉴别及处置。本主题直接调用其他主题的数据结果,如基础地理、基础地质、人文经济等地质环境背景数据、气象预报数据、地质灾害调查及监测数据、地质灾害现场数据等。

3.4 地面沉降监测集市的建设

本集市由以下主题组成:地面沉降成因与发育规律、全国地下水分区评价、地面沉降监测技术方法、地面沉降地裂缝防治技术试验、地面沉降易发区划、地面沉降控制管理等。(1)地面沉降成因与发育规律研究。以长江三角洲、华北平原、汾渭盆地、其他

地面沉降地区作为研究对象,采用调查、遥感解译、钻探、物探、抽水试验、土工试验等多种方法,查明地面沉降历史、现状和发展趋势,查明与地面沉降有关的地层结构、粘性土的物理力学性质、砂类土密实度、地下水动态、孔隙水压力变化等条件。根据以上研究成果,综合收集地面沉降的形成原因、机理、发育规律相关信息建立数据立方。(2)地面沉降监测技术方法。以水准、GPS、InSAR、自动化等监测技术作为研究对象,依托已建立的区域地面沉降监测网络,系统评价、比较和验证水准、GPS、InSAR、自动化等监测技术在不同地区的监测适用性和可行性。以监测技术为二级主题,以监测适用性和可行性的相关参数建立数据立方。(3)地面沉降地裂缝防治技术试验研究。开展地下水回灌试验研究,地下水合理利用调配研究,地裂缝对工程建筑的原位破坏试验研究等,为地面沉降和地裂缝防治提供理论依据和决策支持。收集相关试验结果数据,建立相应数据立方。(4)地面沉降易发区划研究。以长江三角洲、华北平原、汾渭盆地、其他地面沉降地区作为研究对象,开展地面沉降分层标、基岩标及控制性剖面测量,掌握区域地面沉降发展动态,在此基础上开展易发区划研究,为研究制定地面沉降分区控制目标和管理措施提供依据。收集相关结果数据,建立相应数据立方。(5)地面沉降控制管理研究。以长江三角洲、华北平原、汾渭盆地、其他地面沉降地区作为研究对象,结合地面沉降风险的大小和性质,确定控制风险的途径并付诸实施。具体来说,依据地面沉降风险评价成果,进行控制管理区划;针对地面沉降发育现状和地区社会经济发展趋势,以地面沉降控制管理区划为基础,结合地区实际和地面沉降防治水平,研究制定不同管理区的地面沉降控制目标,并有针对性的对控制管理区进行具体工作部署。

3.5 矿山地质环境监测评价集市的建设

本集市由以下主题组成:全国矿产资源集中开采区的矿山地质环境调查、全国矿山地质环境监测体系、矿山地质环境领域科学研究等。(1)全国矿产资源集中开采区的矿山地质环境调查。将全国矿产资源集中开采区的矿山地质环境调查数据,按区域维、时间维和调查结果维,将有关指标建立相应数据立方;(2)全国矿山地质环境监测体系。加快国家级矿山地质环境监测示范区建设,建立 10 个国家级矿山地质环境动态监测示范区,探讨矿山地质环境监测技术方法,制定矿山地质环境监测技术规程,初步建成以矿山企业自主监测为主,定期监测与应急监

测相结合的国家、地方和矿山企业三级监测体系。以国家、地方和矿山企业三级区划为区域维层次,以监测参数为度量建立相应数据立方;(3)矿山地质环境领域科学研究。该研究的主要内容为:(1)矿山地质环境基础理论研究;(2)矿山地质环境调查、监测及预测预报技术研究;(3)矿产资源勘查开发全程矿山地质环境保护及修复技术研究;(4)矿山地质环境评价指标体系和方法研究,探索建立基于不同需求的多级矿山地质环境评价指标体系和方法;(5)矿山地质环境治理技术标准研究,研制矿山地质环境治理恢复技术规范和矿山地质环境治理恢复工程验收标准;(6)矿山地质环境多源信息综合分析与处理研究,依托多种信息源研制基于多源信息的综合统计与分析功能,将相应的研究结果建立数据立方。

3.6 地质遗迹保护集市的建设

本集市由以下主题组成:地质遗迹调查、地质遗迹保护程度评价等。(1)地质遗迹调查。地质遗迹调查的方法有地质遗迹资料登录、遥感解译、野外调查和资料整理与综合研究等。地质遗迹调查程度包括 3 类:基础类地质遗迹调查程度、地貌类地质遗迹调查程度和灾害类地质遗迹调查程度。每一类分 4 个方面:迹点数、调查面积、遗迹类型和遗迹价值。以地质遗迹调查的方法、地质遗迹调查程度及其类别、地质遗迹调查结果内容为内容建立数据立方。(2)地质遗迹保护程度评价。它是对地质遗迹的科学价值、观赏价值进行客观评价,确定其保护等级,为地质遗迹保护管理和开发利用提供科学依据。对地质遗迹保护程度进行评价包括以下内容:①地质公园的建设程度,包括揭牌开园公园数量、已批准资格公园数量以及公园级别;②自然保护区的建设程度,包括保护区面积、遗迹类型;③重要遗迹点的保护程度,包括重要遗迹点数量、遗迹类型。以全国区划为空间维,以自然时间为时间维,以地质遗迹保护程度指标为属性维,建立相应数据立方。

3.7 地质环境综合评价集市的建设

本集市充分利用全国地质环境数据资源和国土资源综合监测成果,研究和完善地质环境综合信息服务指标体系和国土资源综合监测指标体系,开展水工环地质工作战略研究,实施全国各区域地质灾害防治、地下水环境保护、矿山地质环境保护和地质遗迹保护等状况的综合评价,研制与修订水工环及海洋地质调查标准等。其由以下主题组成:地质环境综合监测成果、地质环境综合评价等。(1)地质环境综合监测成果。依托国家级地质环境监测与预报(监

测专项)项目,搜集整理地质灾害、地下水和矿山地质环境等地质环境综合监测成果,建立相应数据立方,为开展国土资源综合监测数据库建设和国土资源综合监测信息服务平台以及全国地质环境状况分析提供数据基础;(2)地质环境综合评价研究.充分利用全国地质环境数据资源,按照地质环境监测综合信息服务四级指标体系框架,实现全国各区域地质灾害防治、地下水环境保护、矿山地质环境保护和地质遗迹保护等状况的评价.以四级指标体系为维层次,以环评结果为主要内容,建立数据立主,为全国地质环境安全综合评价及基础地质环境分析与国土开发安全评价提供数据基础.

3.8 资源环境承载力评估集市的建设

根据土地资源、水资源、矿产资源、能源资源及地质环境、生态环境、气候环境、环境污染自然单元统计数据及各省、地、县社会经济统计数据进行分析,针对环境承载力评价指标的具体数值,采用统计学方法、系统动力学方法等对环境承载力进行综合分析,对区域环境承载力进行评估,为区域地质环境稳定性评价、水土保持以及区域经济可持续发展提供评价依据.本集市由以下主题组成:资源环境承载力评价指标、资源环境承载力评价模型等.主要数据类型如下:(1)土地资源、水资源、矿产资源、能源资源,按省、地、自然单元的现状统计数据及评价数据;(2)地质环境、生态环境、气候环境、环境污染环境等,按省、地、自然单元的现状统计数据及评价数据;(3)社会经济状况,按省、地、县为单元的统计数据;按区域、时间、评价方法、评价指标等建立相应数据

立方.可按全国、省或自然单元进行评估,输出评估报告及相关图表;输出评价成果图,主要有单项评价系列图(如生态脆弱程度图等)及综合评价系列图(如重要资源保障图等).

4 系统实现

系统开发遵循“中国地质环境监测院地质环境信息系统标准规范(2012.6)”,保证系统界面布局一致,系统风格、操作方式统一,其主界面见图 3. 系统主要包括数据仓库管理、元数据管理、数据快速查询、数据搜索引擎、数据挖掘、联机分析报表管理及联机分析处理等模块组成.

经过与各专业用户的充分交流,多维模型的数据组织和联机分析的多样化展示是地质环境数据仓库需要解决的重点问题.图 4 为北京市地下水水质情况多维模型;图 5 为北京市地下水水质情况多维模型的联机分析实例,可以对多维数据进行上钻、下钻、切片、切块、旋转等进行多方位分析,并可以生成曲线图、直方图、饼图等可视化图件,使用户从不同角度提取有关数据,便于使用者更直观地理解分析数据.

5 结论

国家级地质环境信息平台地质环境数据仓库将中国地质环境监测院现有各系统各自为政的操作型数据进行面向分析的整合,形成一个集成的一致的

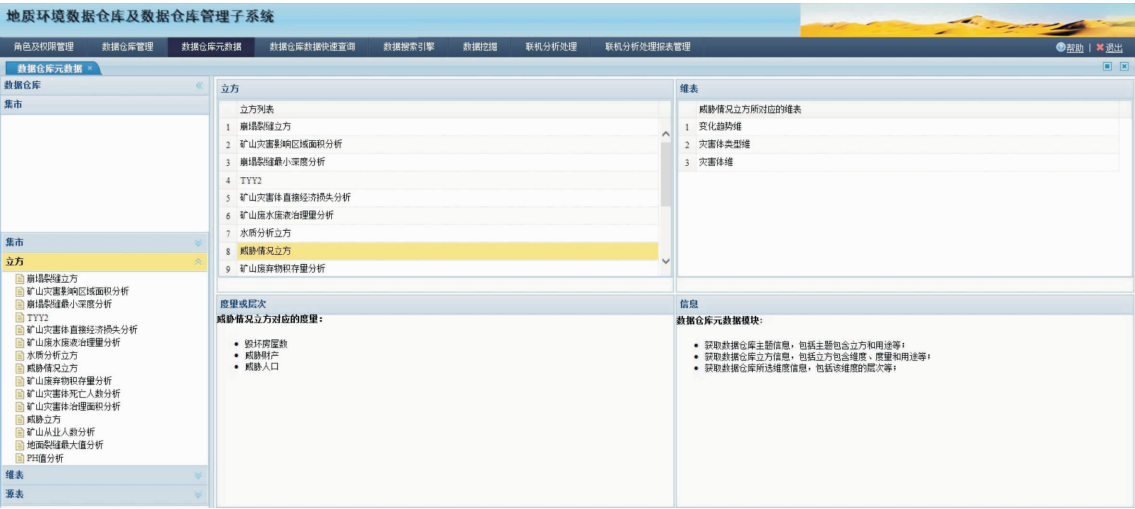


图 3 数据仓库系统主界面

Fig. 3 Interface of data warehouse system

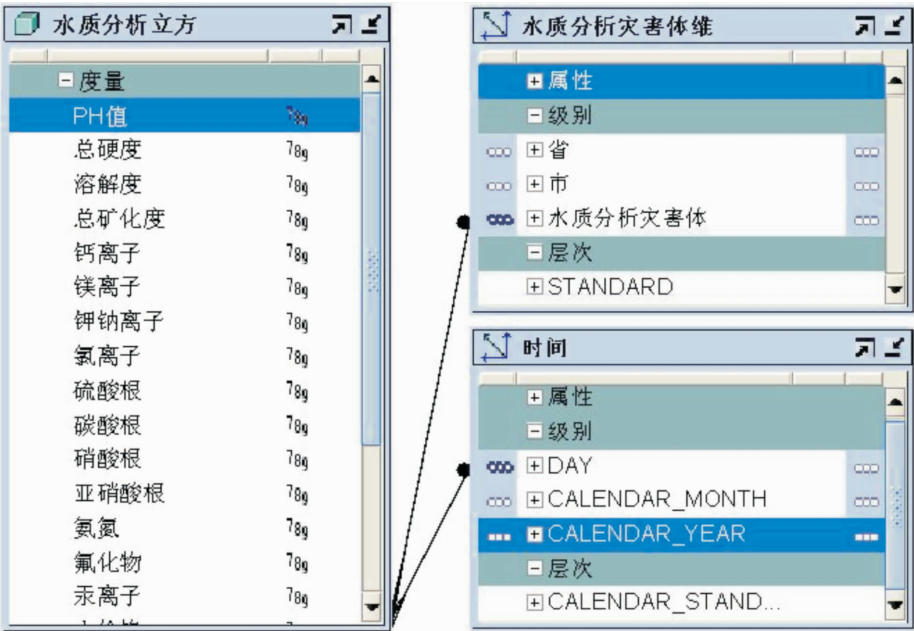


图 4 北京市地下水水质情况多维模型
Fig. 4 Multidimensional model of Beijing groundwater quality



图 5 北京市地下水水质情况多维模型联机分析
Fig. 5 Online analysis of Beijing groundwater quality

数据中心,直接为决策分析服务;突破了地质环境数据来源及应用背景复杂多源、多格式(异构)、多类别、多尺度海量数据处理的难题,可提供给多个差异性较大的分析系统使用。

根据应用的实际需求,数据仓库可分为地质环境决策、地质灾害防治、地下水监测与保护、地面沉降监测、矿山地质环境监测评价、地质遗迹保护、地质环境综合评价及资源环境承载力评估等 8 个集市(或称为领域),每个集市下又包含若干主题或二级主题。按不同集成的数据要求,定义了相应的区域维、时间维及各类数据属性维等;按不同主题和立方

所涉及的数据要求,进行了相应的粒度与度量的定义,从而建立了不同的多维数据模型。

所设计的数据仓库的主要功能和创新之处在于横向打通了各业务系统的数据,有效地把操作型数据集成到统一的环境中以提供决策型数据访问及分析,能适应不同维、不同粒度、不同侧面查询和观察数据的需要。

限于现有的数据条件和地质环境综合评价模型的研究水平,现有的数据仓库实现只是搭建出一个框架,实现了部分重要主题并具有一定的展示度,但离数据仓库的全面应用还有一定距离,随着后期数

据仓库的建设,其数据种类越丰富,数据质量越高,模型与方法越多,数据仓库就将越“聪明”,就能为科学决策提供更有效的支持,并成为地学大数据利用的一个重要方法。

References

- Cohen, S., 1999. NAWQA Results on the Web. *Ground Water Monitoring and Remediation*, 19(4): 47—49. doi: 10.1111/j.1745-6592.1999.tb00238.x
- Forbes, S. R., Burke, R. G., Varma, H., 1999. Designing and Building a CHS Bathymetric Data Warehouse. *International Hydrographic Review*, 76(2): 111—124.
- Hu, G. D., 1999. Several Issues in Design of Geological Data Warehouse. *Earth Science—Journal of China University of Geosciences*, 24(5): 522—524 (in Chinese with English abstract).
- Hu, G. D., Li, Z. H., 2002. Basic Platform of Land and Resources Information System Based on Data Center and Its Technical Questions. *Earth Science—Journal of China University of Geosciences*, 27(3): 306—310 (in Chinese with English abstract).
- Hu, G. D., Li, Z. H., Mei, H. B., et al., 2011. Geohazards Data Warehouse Design and Development of the Three Gorges Reservoir Area. *Earth Science—Journal of China University of Geosciences*, 36(2): 255—261 (in Chinese with English abstract).
- Inmon, W. H., 1992. Building the Data Warehouse. John Wiley & Sons, Inc, Chichester, 356.
- Jia, P. H., Zhu, D. K., Ma, J. S., et al., 2007. Spatial Data Warehouse Establishment for “Digital South China Sea”. *Remote Sensing Information*, 4: 66—70 (in Chinese with English abstract).
- Li, Z. H., Hu, G. D., Chen, J. G., 1999. Geological Data Warehouse Features and Data Organization. *Earth Science—Journal of China University of Geosciences*, 24(5): 536—538 (in Chinese with English abstract).
- Li, Z. H., Hu, G. D., Wang, S. H., 2002. Preliminary Design and Development of a Geological Data Warehouse. *Geology and Exploration*, 38(5): 67—70 (in Chinese with English abstract).
- Li, Z. H., Hu, G. D., Zhang, Z. F., 2003. Development of Geological Data Warehouse. *Journal of China University of Geosciences*, 14(3): 73—76.
- Mahoney, B., 2012. Big Earth—Big Data! In: 2012 Conference on Intelligent Data Understanding. IEEE, U. S.
- Rosen, M. R., Lapham, W. W., 2008. Introduction to the US Geological Survey National Water-Quality Assessment (NAWQA) of Ground-Water Quality Trends and Comparison to Other National Programs. *Journal of Environmental Quality*, 37(5): 190—198. doi: 10.2134/jeq2008.0049
- Tripathy, A., 2010. A Multi Dimensional Design Framework for Querying Spatial Data Using Concept Lattice. In: Mishra, L., Patra, P. K., eds., Advance Computing Conference (IACC), 2010 IEEE 2nd International. Browse Conference Publications, India. doi: 10.1109/IADCC.2010.5422922
- Wen, W., 2009. Building of Digital South China Sea Infrastructure Geographic Information Platform. *China Surveying and Mapping Newspaper*, 2009—12—17(2) (in Chinese).
- White, T., Gray, J., Stack, M., 2009. Hadoop—The Definitive Guide; Map Reduce for the Cloud. O'Reilly Media, Inc, USA.

附中文参考文献

- 胡光道, 1999. 地质数据仓库设计中的几个问题. *地球科学——中国地质大学学报*, 24(5): 522—524.
- 胡光道, 李振华, 2002. 基于数据中心的国土资源信息系统基础平台的构建及技术问题. *地球科学——中国地质大学学报*, 27(3): 306—310.
- 胡光道, 李振华, 梅红波, 等, 2011. 三峡库区地质灾害数据仓库的设计与实现. *地球科学——中国地质大学学报*, 36(2): 255—261.
- 贾培宏, 朱大奎, 马劲松, 2007. “数字南海”空间数据仓库构建技术研究. *遥感信息*, 4: 66—70.
- 李振华, 胡光道, 陈建国, 1999. 地质数据仓库的特点及其数据组织. *地球科学——中国地质大学学报*, 24(5): 536—538.
- 李振华, 胡光道, 王淑华, 2002. 一个地学数据仓库的初步设计与实现. *地质与勘探*, 38(5): 67—70.
- 温伟, 2009. 数字南海基础地理信息平台开建. *中国测绘报*, 2009—12—17(第2版).