

## 技术 方法

# 一个地学数据仓库的初步设计与实现

李振华, 胡光道, 王淑华

(中国地质大学, 武汉 430074)

[摘要]初步设计与实现了一个地学数据仓库:一个涵盖了地质、物探、化探、遥感、地理、矿产、钻孔(井)、成果等几乎全部地质数据的、兼容现有的所有地质数据的国家标准的、提供多种数据格式的输入、输出、查询、管理等功能的、基于网络运行的地学信息系统。

[关键词]数据仓库 地学数据 地学信息管理系统

[中图分类号]P628 [文献标识码]A [文章编号]0495-5331(2002)05-0067-04

数字化时代的来临,使得地学信息的标准化、数字化、集成化、网络化等工作变得尤为迫切,数据库的建设也成为地学研究的重点之一,但由于地质数据所固有的复杂性,这些数据库基本上都属行业数据库。如近年来在地质与勘探行业做的工作比较多的数据库有:地质矿产点源数据库<sup>[1]</sup>和全国主要大中型金铜多金属矿区域地球化学数据库<sup>[2]</sup>,它们很难涵盖地学的所有方面。国外的情况也是如此,如美国的国家水质评价数据仓库<sup>[3]</sup>和加拿大的海底地形数据仓库<sup>[4]</sup>,它们所涉及的也只是一个行业的数据。

能涵盖地学所学方面的数据并为地学的分析和决策服务,正是地学数据仓库的设计目的。胡光道<sup>[5]</sup>、李振华等<sup>[6]</sup>将地学数据仓库定义为:“是一个用以更好地支持地学决策分析处理的、面向主题的、集成的、不常更新的、能存储空间数据的、随空间不断变化的地学数据集”。同时,认为地学数据仓库相比于一般数据仓库(Data warehouse 下称 DW)而言,有如下 3 个特点:1)一般 DW 中的数据表现为时间属性,而地质数据表现为空间属性(有些变动较快的地理数据还具有时空四维的特征);2)一般 DW 中的数据是不可更新的,而地质数据是可更新的;3)一般 DW 中的数据类型比较简单,而地质数据的数据类型比较复杂。基于以上认识,初步设计并实现了一个地学数据仓库。

## 1 系统实现目标与特点

所设计的系统是一个涵盖了地质、物探、化探、

遥感、地理、矿产、钻孔(井)、成果等几乎全部地质数据的、兼容现有的所有地质数据的国家标准的、提供多种数据格式的输入、输出、查询、管理等功能的、基于网络运行的地学信息系统。

设计特点如下:

1) 整个仓库分为安全库、管理库、数据库 3 部分。安全库提供 3 种身份的不同权限管理;管理库存放元数据信息,对整个仓库的结构进行了说明和规定;数据库包括地质、物探、化探、遥感、地理、矿产、钻孔(井)、成果等 8 个主题,每个主题又分为若干子类,每子类由若干表组成。

2) 设计中尽量参考现有的数据标准,参考了国家标准九项,行业及部门标准二项。

3) 提供多种形式的数据库管理形式,且支持 TXT、DBF 等文件的格式的输入及 TXT、DBF、EXCEL、HTML 等格式的输出,并通过此种形式可与所有的 GIS 软件交换数据。

4) 对所有数据实现了查重并自动删除的功能,还特别针对矿产地主题数据库实现了查不一致、查重、重新分号等数据预处理功能。

5) 系统采用客户机/服务器网络结构,服务器端在商用大型数据库系统—SYBASE 上实现,并可易于扩展到其他商用数据库系统。

## 2 总体设计

### 2.1 数据仓库结构

如图 1 所示,整个仓库分为安全库、管理库、数据库 3 部分。

[收稿日期]2001-09-27;[修订日期]2002-02-04;[责任编辑]余大良。

[基金项目]国家“九五”科技攻关项目(编号:96-914-05)与国土资源部矿产资源定量预测与勘查评价开放研究实验室基金联合资助。

[第一作者简介]李振华(1969年-),男,1994年毕业于中国地质大学研究生院,现主要从事计算机软件及其地学应用的教学与科研工作。

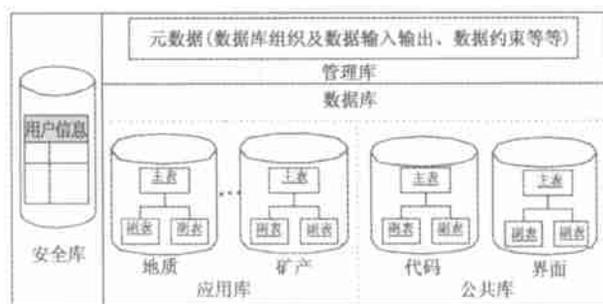


图 1 数据仓库结构简图

安全库存放用户安全信息。

管理库存放元数据信息,其整个仓库的结构进行了说明和规定。

数据库部分按各部分用途又分应用库、公共库。应用库存放各主题地质数据。公共库存放各主题地质数据均需涉及的公用信息,如代码库存放各类地质代码信息、界面库中存放各主题地质数据的界面显示信息,通过此种方式,可实现用户界面的自动生成,如能更进一步结合应用程序的话,就更能充分体现“数据驱动”的设计思想。

### 2.2 主题数据库数据组织

以比例尺为度量参数进行数据组织,如图 2 所示,共分为 4 层,从底至上比例尺依次减小,数据综合程度也依次增加。理论上上层数据可由下层数据生成。不同层次的数据用途不同,如最上层的小比例尺数据回答诸如远景、规划等方面的问题,最底层的大比例尺数据回答诸如定位、开采方面的细节问题。

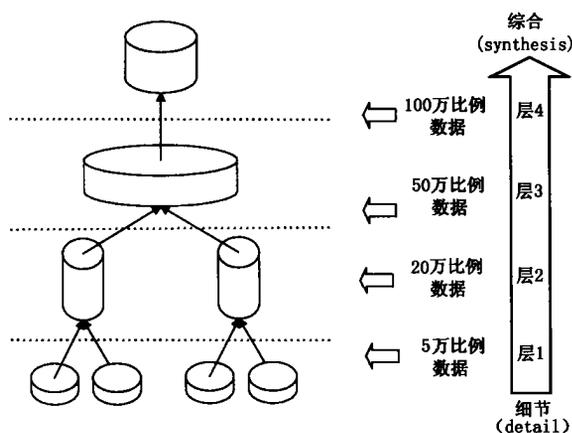


图 2 主题数据库数据组织示意图

### 2.3 不同主题数据库之间的数据叠合

不同的主题数据库之间(亦即不同类型的地质数据之间)的数据类型互不相同,数据结构也不尽一致,为使它们统一起来作为一个逻辑上的整体,借鉴 GIS 中的图层概念,依据空间关系对这些不同类型的数据进行了初步叠合。

如图 3 所示,整个数据仓库在逻辑上呈金字塔结构,自上而上按比例尺大小分为 4 层。每一层又分为多个图层,每个图层代表某一地质类型的数据,多个图层依照空间关系进行叠合。在叠合过程中,空间控制点的选择尤为重要。以此结构为基础,理论上能实现任意类型、任意区块、甚至任意比例尺的数据输出。

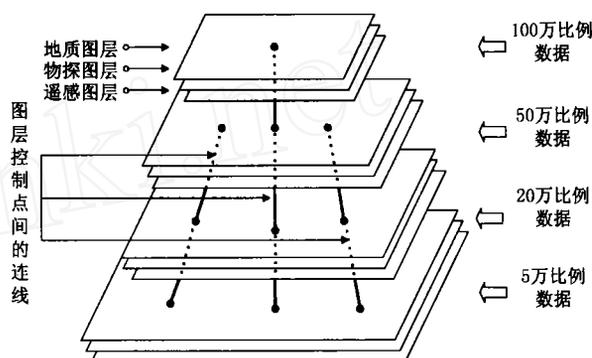


图 3 不同类型数据叠合示意图

### 2.4 网络结构

本系统采用 Client/ Server 架构(图 4),一般来说,在这种模式下,服务器只集中管理数据,而计算任务分散在客户机上,客户机和服务器之间通过网络协议来进行通讯。这种模式充分利用了客户机的性能,使计算能力大大提高;另外,由于客户机和服务器之间通讯是通过网络协议进行的,是一种逻辑的联系,因此物理上在客户机和服务器两端是易于扩充的。同时,由于数据的集中管理,也有效地控制了数据的不一致性。

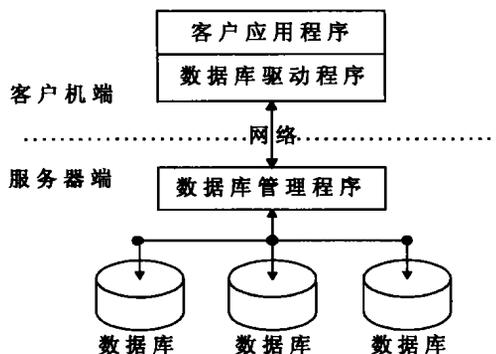


图 4 网络结构示意图

### 2.5 程序功能结构

如图 5 所示,本程序以数据库为核心,较为全面地完全成了数据输入、数据预处理、任意参数的数据

此部分功能在实际系统中仅初步实现。

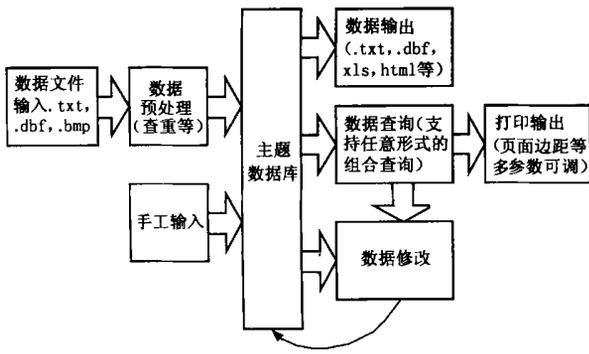


图 5 程序功能简图

查询以及与 GIS 兼容的多种格式的数据输出工作。

### 2.6 功能模块组成

功能模块组成如图 6 所示：

#### 2.6.1 系统安全性模块

1) 用户管理：该模块下分为两个子模块，用户增加和用户管理。前者负责增加系统的操作用户和一般用户，后者负责删除这些用户。

2) 口令修改：用于当前登录用户修改口令。

#### 2.6.2 系统管理及界面模块

数据表管理：存放有全句数据库的重要信息：系统的库组成，各库的表之间的相互关系，各库的文件在批输入和批输出时存放目录、存放文件名及文件格式的约定。同时，还存放有系统菜单及各主题数据库中按钮的生成信息。

基础库：存放图幅信息。其中图幅号是本系统管理的基础，是各主题数据库之间相互联系的纽带，通过它对库中所有数据进行了初步的匹配和叠合。

#### 2.6.3 公共库模块

代码：提供导航式的代码正查及反查功能。此功能帮助用户翻译主题数据库中存放的地质代码。

#### 2.6.4 功能性模块

- 1) 单库浏览：浏览和查询表中信息。
- 2) 单库管理：增加、删除、修改表中信息。
- 3) 文件输入：将 TXT、DBF 等格式的文件输入相应表中。
- 4) 文件输出：将表中数据存为多种格式的文件。
- 5) 清库：清除表中的所有信息。
- 6) 文件批输入：一次性的将多个文件的信息输入库中。
- 7) 文件批输出：一次性的将库中的信息存入多个文件中。
- 8) 数据输出：由用户据图幅选择所需数据，并指定存放目录、存放文件名或存放格式。

#### 2.6.5 数据处理模块

在任何数据操作后(文件批输入除外)，系统将自动进行空数据及重复数据的删除工作。

#### 2.6.6 矿产地数据预处理模块

查不一致：查找中转库中的不一致数据。

查重：查找与主库的重复的数据。

重新分号：自动为中转库中的数据重新分配矿产地编号。

### 3 结论及建议

借鉴传统数据仓库的设计思路，结合地质数据的固有特点，初步设计了一个地学数据仓库。该系统是一个涵盖了地质、物探、化探、遥感、地理、矿产、钻孔(井)、成果等几乎全部地质数据的、兼容现有的所有地质数据的国家标准的、提供多种数据格式的输入、输出、查询、管理等功能的、基于网络运行的地学信息系统。

本系统在云南省地矿局完全运行，至今已过一年，虽然云南省幅员广阔，地质数据种类复杂，数据量巨大，从而提供了很好的测试和运行的环境，但距离数据仓库的完全实现还比较遥远，其一，从数据量上看，还不能实现真正的实践意义上的数据仓库级的“海量”数据管理，其二，从技术上看，不同类型地学数据之间的叠合，以及更进一步的有目的的数据融合的工作做得还很少，其三，从使用上看，本系统还未与现有 GIS 软件实现完全的结合，这方面既有 GIS 软件对大型商用数据库产品的支持不够的现实问题，又有多维地学数据在二维 GIS 中进行转换和表达的技术问题。

致谢：感谢中国地质大学姚书振教授自始至终的指导和关心。感谢云南省地矿局谢蕴宏高工、李定平高工、何家学高工在系统实施过程中的大力支持。

#### [参考文献]

- [1] 吴冲龙,汪新庆,刘刚,等.地质矿产点源信息系统设计原理及应用[M].武汉:中国地质大学出版社,1996.
- [2] 史长义,张金华.全国主要大中型金铜多金属矿区域地球化学数据库[J].地质与勘探,2001(3):4144.
- [3] R W Bell. Data warehouse for the National Water - Quality Assessment Program of the U. S. Geological Survey[A]. The 34th annual meeting of Geological Society of America, South - Central Section[C]. U S A: Geological Society of America, 2000.
- [4] S R Forbes, R G Burke, H Varma. Designing and Building a CHS Bathymetric Data Warehouse[J]. International Hydrographic Review, 1999(2):111124.
- [5] 胡光道.地质数据仓库设计中的几个问题[J].地球科学-中国地质大学学报,1999(5):522524.

[6] 李振华,胡光道,陈建国. 地质数据仓库的特点及其数据组织

[J]. 地球科学 - 中国地质大学学报,1999(5) :536538.

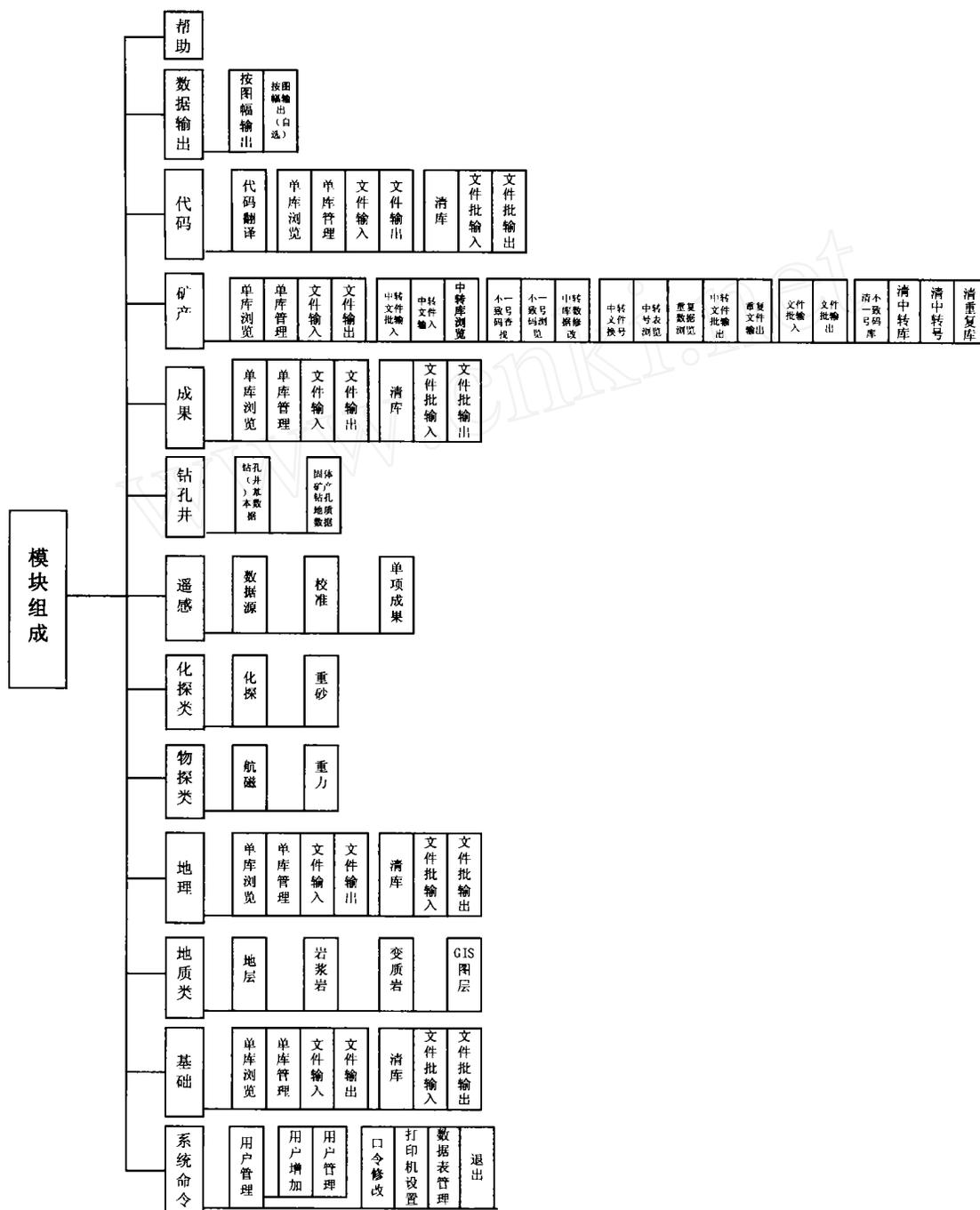


图 6 功能模块组成图

PRELIMINARY DESIGN AND DEVELOPMENT OF A GEOLOGICAL DATA WAREHOUSE

LI Zhen - hua ,HU Guang - dao , WANG Shu - hua  
 ( China University of Geosciences , Wuhan 430074)

**Abstract :** Preliminary design and develop a geological data warehouse : a geological information system with features of covering almost all geological data such as geology ,geophysics ,geochemistry ,remote sense ,geography ,mineral resource ,drill hole ,research result and so on ,compatibility to all available national standards ,supporting data input ,output ,query ,management ,etc ,having many data formats and running on network .

**Key words :** data warehouse ,geological data ,geological information system