技术·方法

漫谈数据仓库技术在地勘行业内矿产资源信息共享的应用前景

黄　晶
(中国冶金地质勘查工程总局矿产资源信息中心,燕郊　065201)

[摘　要]将冶金地质系统 50 年来积累的地矿资源数据集中起来,便于管理和合理使用有用信息,建立矿产资源数据仓库是很重要的一种手段,数据仓库技术能将集成的数据置入多维空间,供地质人员从不同工作视角获取信息,综合分析数据,以指导找矿工作,在此,简要介绍了如何建立矿产资源数据仓库的过程。

[关键词]数据仓库　数据库　计算机信息化　矿产资源
[中图分类号]P628　[文献标识码]A　[文章编号]0495 - 5331(2002)04 - 0067 - 04

0　前 言

50 多年来冶金地质工作积累了丰富的地质矿产资料。通过计算机技术实现对这些信息共享是矿业和地质工作者多年来一直追求的目标。近年来,人们在这方面做了许多工作,利用传统的数据库技术已建立或正在建立各种矿产资源数据库。也就是说矿产资源的原始数据正在快速地积累和膨胀,如何保存并利用好这些珍贵的资源,将其中蕴藏的信息转化为地质生产力,将成为业内人士所关注的又一个技术热点。在此,笔者漫谈一下最近几年才快速发展起来的数据仓库技术以及基于此技术的情报咨询智能体系,利用该技术深层次、多角度地挖掘,分析当前和历史的地质勘探数据以及相关自然环境、经济条件等多种数据,发现其内在的规律,从而得到宝贵的找矿支持信息。

1　数据仓库及特点

1998 年才发展起来的数据仓库(Data Warehouse)技术对许多人来说也许还很陌生,而它在国外已是如火如荼了,数据仓库之父 W H inmon 将其定义为"数据仓库是支持管理决策过程的,面向主题的、集成的、随时间而变的、持久的数据集合。"

数据仓库与传统的数据库系统相比有着本质的区别,数据库是一种通用平台,建立于严格的数学模型之上,用来管理一个行业或部门或企业的数据,进行事务处理,完成相关业务。而数据仓库没有严格的数学理论,更偏向于工程,它不是花钱就可购买到的成品,而是行业部门一个日积月累的建立过程,它的应用对象是不同层次的工作人员,它的数据源是多种数据源,库中数据无须修改删除,主要是大规模查询和分析,因此要求有大量的历史数据和汇总数据。

2　数据仓库的建立与管理

这是数据生成环节,需要在充分了解地质找矿人员需求的前提下按数据仓库方法设计数据仓库结构。数据仓库将数据仓库体系结构的设计环境分 4 层:即操作层、数据仓库细节层、部门层和个体层。根据这 4 层管理要求分步设计多级数据环境:

1) 设计操作型数据环境。数据操作层只保存某一主题的原始数据,而且一般是当年的非集成的原始数据。这些数据可以是在数据仓库服务器端手工录入,图形扫描,也可以在野外利用 GPS 数据采集、卫星接收等方式收集,还可从其它类型数据库中转换而来。它的管理方法同传统的数据库管理方式相类似,基本是插入、删除、修改等操作。

2) 将操作型数据转入数据仓库中。从数据操作层导出的数据就移入数据仓库细节层(复制、抽取、清洗…),这些都是逐年累积的原始数据。如果操作

[收稿日期]2002 - 04 - 23;[修订日期]2002 - 06 - 03;[责任编辑]曲丽莉。
[作者简介]黄　晶(1959 年 -),女,现主要从事计算机软件开发工作。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

层数据有了修改,只需定时在数据仓库的细节层中追加新记录,再加上时间期限进行跟踪即可。数据转换的时间间隔可以定时段进行一次。存储年限通常是 10～20 年。

数据仓库的细节数据要向小规模数据集市的数据复制和分布,也就是说从数据仓库细节层中按某一指标作轻度综合统计,形成部门数据层,再在此基础上对某一指标进行高度综合统计,形成数据个体层,这一层中的数据用以完成大多数启发式分析。这种分层存储的管理结构有一个最重要的目的,就是根据不同的查询要求向地质工作人员提供不同的细节。

转入数据仓库里的数据已完全不同于传统的数据库里的数据了,传统的数据库是二维数据结构,而数据仓库里的数据结构则是一种多维视图,这是因为人们对事物的思考是在时间与空间中对一个实体的综合思考,从而掌握详细定义一种工作环境下实体之间关系的方法。用来管理这种多维结构的模型被称为"星形连接"。图 1 给出星型连接的一个简单例子。由图 1 可知,事实表位于星型连接的中央,它是被大量载入数据的实体度量,以数值形式出现。在其周围分别是"矿区地理特征表"、"地质工作程度表"、"矿床地质表"、"矿区地质特征表"、"矿体地质特征表"等等,是用以进一步阐述中心主题的辅助论题,被称为维表,它们均以字符形式表示。事实表包含了一个用于索引查询的主键,例如"矿种编号",以及若干指向其周围的表——维表的外键,例如"矿区编号"、"矿体编号"等,由这些外键放射性地同其它维表的主键相关联,事实表可有任意多个外键与任意多个维表相关。采取这种联接方式主要是因为数据仓库是一个装载—访问环境,它包括很多历史数据,且有大量的数据要管理,因此,星型连接的数据结构对于大多数数据仓库环境的设计来讲,是非常有用的。利用这种结构模型用户可从事实表入手,随意查阅其它相关维表的任意数据。

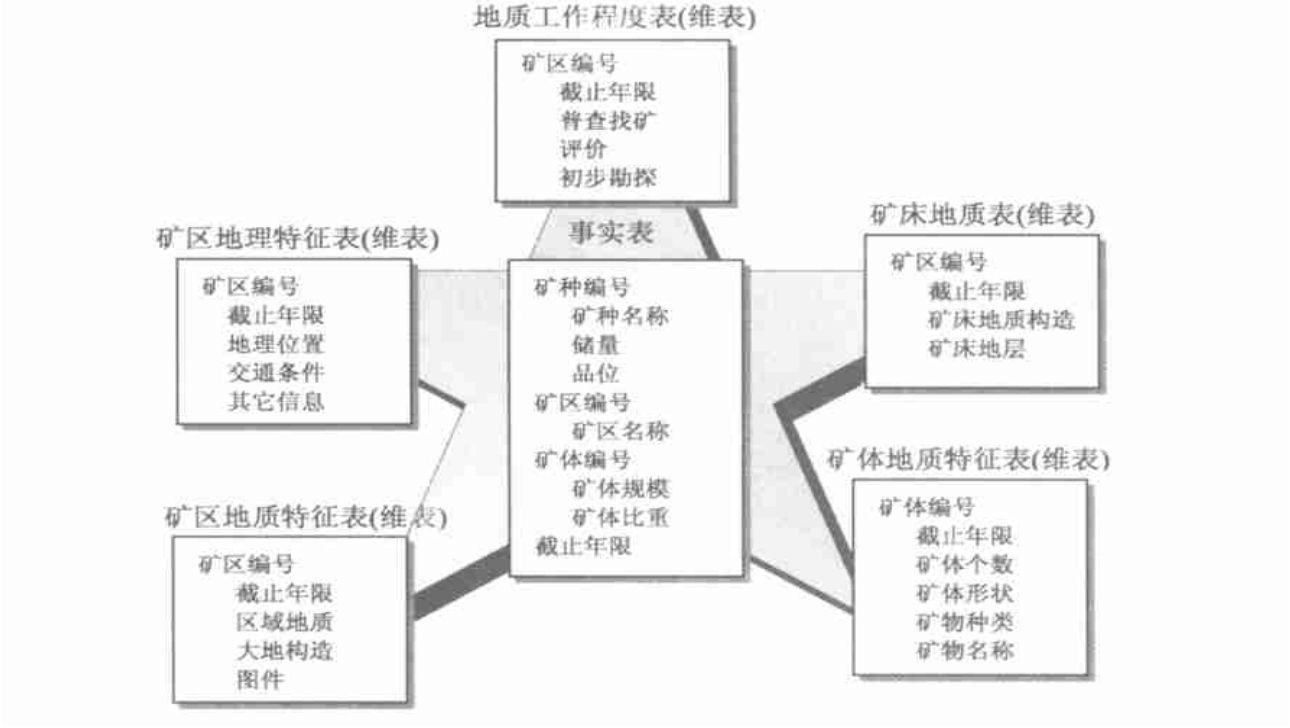

图 1 一个简单的星型连接模型

3 数据仓库利用

建成的数据仓库可针对不同的地质工作需求以不同的风格展现分析查询数据的方式。具体有以下几种方法:

1) 普通的数据查询。它可以分组查询、分类查询、综合关联查询。数据仓库里的数据是多维立方体结构,可以从多个被索引键码联接起来的维表中抽取,也可以从一个维表中抽取需要分组查询的维,围绕事实表为中心议题作任意组合,以便于综合分析。现举一简单例子说明对一种相关主题可进行几种分类组合查询。将以储量表为主脉,从其它维表中抽取相关的维信息,查找 1981 年至 1982 年的北京、广西、湖南 3 省金矿和铁矿的储量,查出的结果

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用二维数据表序列的形式表示,如表 1 所示。用三　　维数据立方体表示如图 2。

表 1　矿产储量数据

	北京				广西				湖南			
	1981 年		1982 年		1981 年		1982 年		1981 年		1982 年	
矿种	探明储量	保有储量	探明储量	保有储量	探明储量	保有储量	探明储量	保有储量	探明储量	保有储量	探明储量	保有储量
金(t)	200	300	100	500	500	300	200	100	600	400	200	300
铁(万 t)	56	80	90	100	300	400	600	200	300	700	600	200

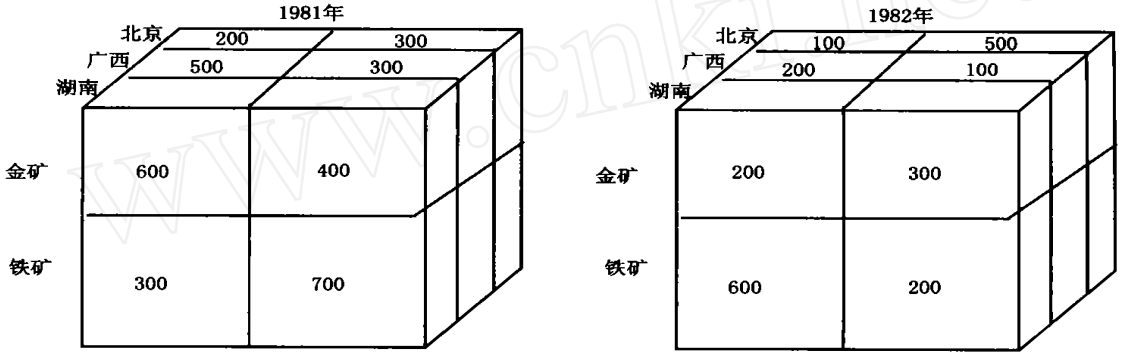

图 2　数据的 3 - D 数据立方体表示
(维是矿区、矿种和储量)

对于 n 维数据立方体,可能产生的分组总数是:

$$T = \prod_{i=1}^{n}(L_i + 1)$$

其中,n 是可分组查询的维数,L_i 是维 i (除去顶层不分组全部查询)的层次数。该公式基于这样一个事实,如果数据立方体有 4 个查询维:矿区名称维、矿种维、储量维、年代维,每个维只有一个层次再加上无分组全部查询这一层,则可能产生供查询的分组总数是 $2^4 = 16$。

2) 关键指标对比分析。地质人员从数据仓库中提取有用资料可向用户显示成多维表格并辅以簇状柱形图进行比较分析,既可用于对区域性的多个矿床进行综合对比分析,还可针对某个矿床的某几个特征进行综合对比分析。例如对比分析区域矿产地分布特征;对比分析矿床中矿体特征及分布情况;对比分析区域内矿床的地质特征;对比分析各矿床、矿体中矿物属性;区域内各矿床或矿床内矿石质量分布及对比;全国或某大区域某矿产储量的分析对比;区域内矿产分布及开发利用程度的分析和对比,等等。帮助地质人员根据工作需要对区域内成矿条件、矿床(点)分布及其它地质特征开展研究,以指导进一步的找矿工作。

3) 趋势对比分析。从查询数据中还可生成趋势分析图表,据此地质人员对某一矿区的情况在时间轴线上作趋势分析,利用趋势变化图可以直观地确定该地区是否有工业利用和再利用的前景。

4) 向下探察分析。从一个汇总数据开始,将该汇总数据逐次地分解成一系列细致的汇总数据。先从高度综合数据层提取信息,再分解到轻度综合数据层,最后追究到数据仓库的细节层。图 3 显示了一个向下探察分析的简单实例。

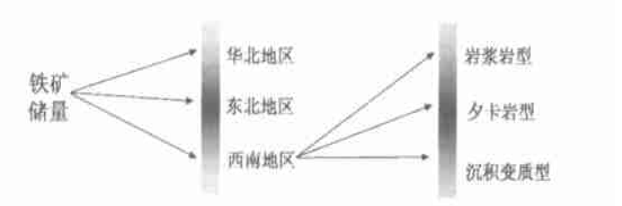

图 3　向下探察分析

如图 3 所示,地质人员已经看过了某年的铁矿储量结果并想进一步探讨。地质人员观察构成汇总的各个地区的铁矿总储量,要分析的是华北地区、东北地区、西南地区的数据。在观察这几个地区数据的过程中,决定仔细查看一下西南地区岩浆岩型、夕卡岩型、沉积变质型的储量情况,另外还可以再进一步查找其它各类型的地质特征等等,这样逐层深入观察,可以详细地了解该区的找矿重点,圈定该地区的找矿规划。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

4 展望未来

若要有效地运用数据仓库里的数据,还必须将数据仓库和万维网(简称 Web)技术在若干层面上结合起来,才能改善数据的流通渠道、促进信息在广域范围内进行共享。本作者粗略地勾勒了一个在 Web 机制下,数据集中存储,信息方便地传输及共享的轮廓图(图 4)。

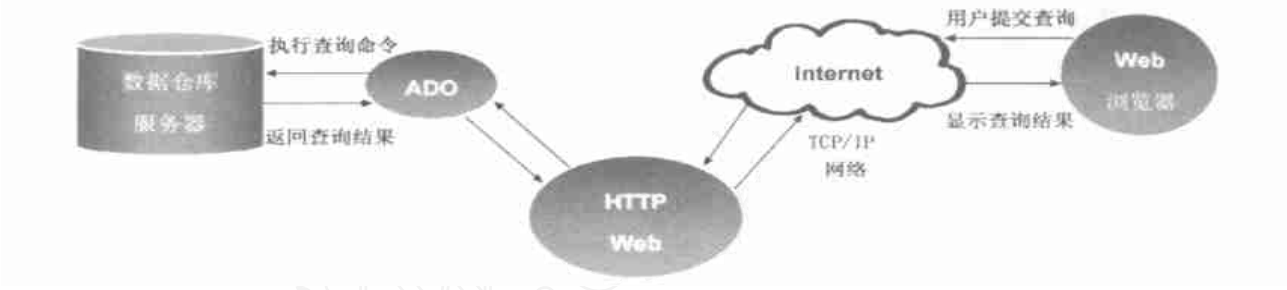

图 4 在 Internet 网上可查询数据仓库的数据

图 4 的工作原理是这样的:Web 浏览器是图形用户界面的客户机,它以作为表达工具,用于解释和表达以 HTML 格式存储于 HTTP Web 服务器上的内容,它由嵌入的图形、超文本链接和 Java 程序组成。通过单击热区,可以激活隐藏的寻址机制。假若某山东局的一位地质人员想查询位于北京的中央数据仓库中的一条信息,他可在本地的 Web 浏览器的客户端上发送他的查询请求,机器便以超文本传输协议(HTTP)进行通信,通过 TCP/IP 网络,传送到 HTTP Web 服务器中,在那里,查询请求被 Web 后端的数据访问层分析解释,运用的工具是普通网关接口(CGI)程序,HTTP Web 服务器通过 CGI 程序分析这是一个数据仓库访问请求,就调用 ADO,这是一个 web 通用的数据访问方法,由 ADO 来响应那位山东局地质人员的请求,访问中央数据仓库,从数据仓库里取得结果后加载到嵌在 HTML 文档里的控件中。再从 HTTP Web 服务器返回到客户浏览器上,显示查询结果。上述完成的工作流程是在排除了网络安全问题的前提下完成。

这就是未来建立和利用地质矿产仓库的美好前景,数据建设一个完整的矿产信息智囊仓库,任何相距遥远的合作者不用出远门就可共享某一共同项目的观点和关于子项目的所有内容。

[参考文献]

[1] (美) William A. Giovinazzo. 面向对象数据仓库设计[M]. 北京:人民邮电出版社,2000.

[2] (美) W. H. Inmon. 数据仓库[M]. 北京:机械工业出版社,2000.

[3] (美) Lou Agosta. 数据仓库技术指南[M]. 北京:人民邮电出版社,2000.

[4] (加) Jiawei Han, Micheline Kamber. 数据挖掘概念与技术[M]. 北京:机械工业出版社,2001.

[5] 袁鹏飞. SQL Server 7.0 数据库系统管理与应用开发[M]. 北京:人民邮电出版社,1999.

APPLICATIONS OF DATA WAREHOUSE IN MINERAL RESOURCE COMPUTERIZATION

HUANG Jing

(*Mineral Resource Information Certer, China Exploration & Engineering Bureau, Yanjiao* 065201)

Abstract: Since last 50 years, more and more mineral data resources have been collected. In order to manage, reuse and extract information from them, computerized data warehouse is one of the most important methods. The data warehouse can integrate the data into multi - dimensional space so that geologists can search, analysis and obtain more useful information from different viewpoint in more efficient way. In this article, the creation and processing of data warehouse have been presented.

Key words: data - warehouse, database, computerization, mineral resources

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net