第28卷 第3期
2006年9月

西北地震学报
NORTHWESTERN SEISMOLOGICAL JOURNAL

Vol. 28 No. 3
Sept., 2006

短文

数据仓库技术在地震信息系统中的应用

石 伟

(黑龙江省地震局,黑龙江 哈尔滨 150090)

摘 要:数据仓库是近年来信息领域中迅速兴起的一种技术。本文对此技术和应用进行了讨论,提出了适用于地震信息系统数据仓库的方案设计思想。

关键词:数据仓库;联机分析处理;数据挖掘

中图分类号:P315.63;TP311.13 **文献标识码**:A **文章编号**:1000-0844(2006)04-0285-02

Application of Data Warehouse to Seismic Information System

SHI Wei

(*Earthquake Administration of Heilongjiang Province, Harbin* 150090, *China*)

Abstract: Data warehouse is a new technology in information field which develops rapidly in recent years. In this paper the method and application of Data Warehouse are discussed, and a resolve program for appling this technology to seismic information system is put forward.

Key words: **Data warehouse; On-line analytical processing; Data mining**

0 引言

近年来随着地震行业计算机应用的不断深入,已经投入大量的时间和资源建立了庞大而复杂的信息系统。"十五"期间中国数字地震观测网络项目建设完成,地震信息系统存放了大量各种测震和前兆数据,庞大的数据库系统给数据的查询和应用带来了困难,迫切需要有一个强有力的分析工具能从这些海量数据中充分挖掘有意义的信息。而近年来在数据库基础上产生的数据仓库(Data Warehouse)能够满足决策分析所需要的数据环境,它被普遍认为是数据库技术未来发展的方向,著名的数据库厂商如Informix、IBM、Oracle、NCR、Sybase等都提供了各自的数据仓库解决方案。数据仓库的目的是建立一种体系化的数据存储环境,将分析决策所需要的大量数据从传统的操作环境分离出来,使分散、不一致的操作系统转成集成、统一的信息,进而支持决策。完整的数据仓库包括三个方面的技术内容:数据仓库技术、联机分析处理技术和数据挖掘技术。本文对数据仓库技术及其决策支持工具进行介绍,并提出适用于地震行业的方案设计思想。

1 数据仓库及其决策支持工具

根据W. H. Inmon的定义,数据仓库是面向主题的、集成的、稳定的、随时间变化的数据集合,用于支持管理决策。数据仓库作为一个集成的数据库,把数据从各个信息源中提取出来,按照数据仓库所用的公共数据模型进行相应变换,并与仓库中现有数据集成在一起。在数据仓库中,由于数据模型和语法等方面的差异已被消除,数据库可直接被访问,因此查询和分析处理都很快。数据仓库创建以后,当用户使用数据仓库时,可以通过联机分析处理、数据挖掘等数据仓库的应用工具对数据仓库进行决策查询分析或知识挖掘。

1.1 联机分析处理

联机分析处理(On Line Analysis Processing,OLAP)是一个得到广泛应用的数据仓库使用技术,它是一种基于数据仓库的快速查询分析技术,侧重于对决策人员和高层管理人员的决策支持,专门用于复杂的分析查询,可以将各个层次的数据进行综合处理,获取决策分析所用的关键信息。OLAP技术主要有两个特点:一是在线性,表现为对用户请求的快速响应和交互式操作,其实现是由客户机/服务器体系结构完成的;二是多维分析,这也是OLAP技术核心所在。

1.2 数据挖掘

数据挖掘(Data Mining,DM)是近几年随着数据仓库技术的人工智能技术发展起来的一门新兴的数据库技术,就是从大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中的事先不知道的,但又是潜在有用的信息和知识的过程。

2 地震信息系统数据仓库建议方案

2.1 系统概述

随着各种类型的地震信息数据不断出现,对地震数据分析需求呈现出多样化的要求,例如:地震科学研究、地震分析预报、地震监测、地震应急指挥、防震减灾工程、电子政务等。如果为每个部门、每个应用分别建立数据库并编制相应的应

收稿日期:2006-08-31

作者简介:石 伟(1971-),男(汉族),黑龙江哈尔滨人,硕士,工程师,主要从事地震信息处理等工作.

用软件，工作量将会非常巨大，而且难以实现。而在数据仓库中建立完整的数据视图，在此基础上根据不同的需求建立不同的数据集市，将有助于这一问题的解决。应用数据仓库提供的即时查询、数据挖掘、联机分析等工具可以满足不同部门的需要，根据数据仓库的内容为用户生成实用的各种报表和图形。

2.2 系统总体架构

数据仓库技术的关键是从许多来自不同的信息系统的数据中提取出有用的数据并进行清理，以保证数据的正确性，然后经过抽取(Extraction)、转换(Transformation)和装载(Load)，即ETL过程，合并到一个整体级的数据仓库里，从而得到数据的一个全局视图。在此基础上利用合适的查询和分析工具、DM工具、OLAP工具等对其进行分析和处理(这时信息变为辅助决策的知识)，最后将知识呈现给使用者。

地震信息共享数据仓库的体系结构分为五个部分：源数据、ETL过程、综合层数据仓库、数据集市和最终的应用程序(图1)，下面分别加以说明：

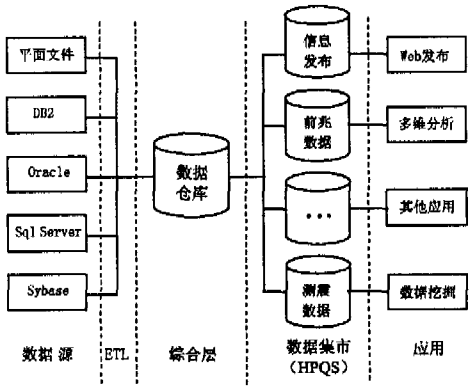

图1 地震信息系统数据仓库体系结构图

Fig. 1 System structure of data warehouse in seismic information system.

数据源是数据仓库系统的基础，通常包括内部信息和外部信息。内部信息包括存放于RDBMS中的各种地震业务处理数据和各类文档数据；外部信息包括各类法律法规、防震减灾政策信息等。数据源可以有多种类型，如各类关系数据库和平面文件等。

ETL过程包括了数据抽取、数据清洗、数据转换、数据加载等环节。数据的抽取是从数据源，也就是各种业务系统将用于分析的数据抽取出来，一般可以通过ODBC数据源或是文本文件进行。通过ODBC方式抽取数据的优点是可简化抽取过程，缺点是抽取速度较慢；通过文本导出数据文件的方式优点是速度较快，而缺点是中间过程较多，不利于自动化处理。因为数据的抽取过程可能会对业务系统造成一定的影响，所以这部分工作大部分是在业务系统不繁忙的晚上进行处理。由于业务的变化、误操作、输入错误等原因，会造成业务系统数据的不规范，不一致，甚至错误。为了保证数据仓库中用于分析的数据的正确性，必须在数据进入数据仓库之前对其进行清洗和规范。数据加载就是将经过清洗、规范的数据加载到目标数据仓库中来，可分为两种：A增量加载(将增量数据加载到数据仓库的表中)；B更新加载(将物理表完全更新进行加载)。

综合层数据仓库是数据的一个全局视图，它的结构是否合理对以后的应用是否成功至关重要。因此综合层数据仓库结构的设计是整个数据仓库实施中的重中之重。仓库中存在着不同的综合级别，一般称之为“粒度”。粒度越大、表示细节程度越低、综合程度越高。级别的划分是根据粒度进行的。

数据集市(Data Mart)或者叫做“小数据仓库”。如果说数据仓库是建立在地震行业级的数据模型之上的话。那么数据集市就是行业级数据仓库的一个子集，他主要面向部门级业务，并且只面向某个特定的主题。数据集市可以在一定程度上缓解访问数据仓库的瓶颈。数据集存储了由数据仓库来的，经过裁剪和归整的数据，这些数据针对某个业务部门或某种业务分析应用而建立。它一般都对数据进行了各种层次的汇总，并建立多维分析的模型，同时也包括了数据采样。其存储主要有关系数据库和多维数据库。其中多维数据库存放多维分析数据，而关系数据库则存储星型模式。另外数据集市一般都被构建成逆规范化的高性能查询结构(High Performance Query Structure，HPQS)，以提高查询效率。

最终的用户程序即呈现(Presentation)和分析(Analysis)工具。从最终用户的观点来看，展示层是最重要的部件，包括呈现和分析工具。不同的用户类型需要不同的前端工具，但所有的用户都能访问相同的数据仓库结构。同样，不同级别需要对结果进行不同程度的可视化处理。例如，图像用于高层次的展示，而表格用于进一步的分析。前端工具主要包括各种报表工具、查询工具、数据分析工具、数据挖掘工具以及各种基于数据仓库或数据集市的应用开发工具。其中数据分析工具主要针对OLAP服务器，报表工具、数据挖掘工具主要针对数据仓库。

3 结束语

本文对数据仓库、OLAP、DM等几个概念做了介绍，在此基础上对数据仓库和OLAP解决方案在地震系统中的应用进行了较为深入的分析，给出了地震信息系统数据仓库的总体结构，初步满足了目前地震系统管理的需要。随着“十五”中国数字地震网络项目的建设，该系统会逐渐引入OLAP、DM等技术，以便于在大量的数据中提取有价值的信息，为地震预报跟踪和科学研究提供保证，为科研人员的决策分析提供及时准确的信息。

[参考文献]

[1] 王珊. 数据仓库技术与联机分析处理[M]. 科学出版社. 1999.

[2] 项军，雷英杰. 数据仓库技术与应用[J]. 计算机与现代化. 2004，11(11)：86－91.

[3] 马征. 数据仓库技术的研究与应用[J]. 丹东纺专学报. 2004，11(1)：5－7.

[4] 柳莺，赵艳红，钱旭，等. 数据仓库技术研究与应用探讨[J]. 计算机应用. 2004，21(2)：46－48.