

●微机开发和应用

数据仓库技术原理及其应用

李社宗, 赵海青, 马青荣, 胡俊霞, 杨光仙

(郑州市气象局, 河南 郑州 450005)

摘 要: 分析了传统数据库与数据仓库的实现技术, 讨论了数据仓库的主要特征和存储数据的层次结构。**关键词:** 数据仓库; 数据特征; 数据结构**中图分类号:** TP311.131**文献标识码:** A**文章编号:** 1004-6372(2003)03-0037-02

引 言

在 20 世纪 90 年代初, W. H. Inmon 提出了“数据仓库”的概念, 它的目标是借助于数据仓库全面、大量的数据存储, 依靠数据挖掘技术与数据分析工具, 达到高效的决策支持。由于数据仓库是数据库、数据挖掘和联机分析处理等多类技术的结合体, 故在实现过程中仍有许多问题需要解决。

1 数据仓库与数据库技术的异同

数据仓库是近几年发展起来的将大量的、决策分析所必需的、历史的、分散的、详细的数据加以转换、清洗、综合, 适应测算利润、分析风险、分析市场以及加强客户服务与营销活动的一种新技术。数据仓库的建立并不是要取代数据库。相反, 它进行复杂分析时, 需要尽可能全面的数据。数据仓库实质上仍是计算机存储数据的系统, 由于使用目的不同, 其存储的数据在量上和质上都与操作型数据库有所不同。数据仓库是以最终用户的观点组织和管理数据的, 而传统的数据库则是以提高查询数据的效率为目的来设计库结构的^[1]。

2 数据仓库的主要特征

数据仓库实质上是对大量的、多种类的数据的集成策略^[2], 是一种管理技术, 目的是使人们更好地利用已有信息进行有效的决策分析。面向主题、集成、稳定和随时间变化是数据仓库的最主要特征。

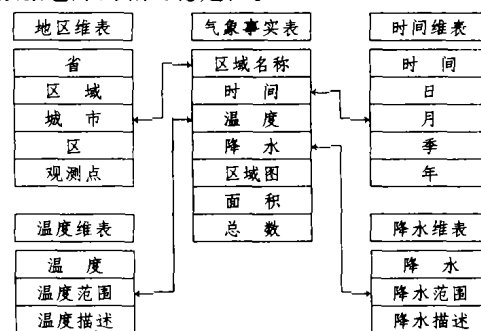
2.1 主题性

与传统面向应用的数据库不同, 数据仓库是面向主题的。主题是一个在较高层次将数据归类的标准, 每一个主题基本对应一个宏观的分析领域。基于主题组织的数据被划分为各自独立的领域, 每个领域有自己的逻辑内涵, 互不交叉; 而基于应用的数据则只是为了处理具体的应用而组织在一起的。在数据仓库中, 每个主题可以包含若干维, 每个维都与一张维表相关联, 可根据数据的综合程度和所属时间段不同来划分。附图是一个面向天气预报的空间数据仓库。

2.2 稳定性

数据仓库存储面向分析的宏观的历史数据, 这就要求数据仓库是相对稳定的。对于决策分析来说, 更关心的是支持

趋势分析的某个时间段内宏观的历史数据, 而不是某一时刻的细节数据。数据仓库中存在的数据一经确定后变化不大, 保证了数据仓库的相对稳定性。



附图 气象空间数据仓库的星形模式示例

2.3 集成性

数据仓库从各种不同的原始数据源中定期抽取、转化数据并最终集成为面向决策分析的数据。由于数据来源、数据类型和存储方式的不一致, 所以数据从传统数据库进入到数据仓库之前, 必须要进行加工与集成。如现在的气象资料, 有来自于报表的 EXCEL 格式, 有从中国气象中心接收过来的文本格式等。在把它们转入数据仓库之前, 必须对它们用相应的转化程序转化成一致的数据格式, 并进行错值、缺值等处理, 还要进行一定的加工和集成。

2.4 时效性

为了适应决策支持系统进行趋势分析的需要, 仓库数据的时限远远长于操作型环境数据时限, 一般在 5~10 年。在实现上, 数据的码键都包含时间项, 它标明数据的历史时期, 因此时间是数据仓库的一个重要属性和组织方式。在面向天气预报的数据仓库中, 历史资料样本长度常常在一二十年甚至三四十一年, 资料的年、月、日是必不可少的属性, 离开了它们, 任何资料都失去了意义, 也无从寻找天气变化规律。

3 数据仓库的层次结构

数据仓库中, 由元数据的时间机制生成的历史数据和由元数据的综合机制生成的综合数据, 形成了决策分析的数据基础。根据对数据的不同粒度处理, 可将数据仓库中存储的数据划分不同的层次, 不仅有与原始数据相对应的细节型数据, 也有在细节数据上综合提炼的综合数据, 还包括数据仓库的结构信息, 即元数据。

具体层次结构如下。

收稿日期: 2003-02-15

作者简介: 李社宗(1964-), 男, 陕西合阳人, 学士, 工程师, 从事网络管理工作。

河南气象 2003 年第 3 期

3.1 详细数据层

存储历史数据和最新数据,是细节型数据,供进一步的分析和预测所用。如每天的降水量、温度、涡度场、散度场等,均属详细数据层,是作进一步分析的基础。

3.2 不同程度的归纳总结信息层

包括多个层次的综合数据,根据所需分类和归纳的不同深度进行统计。如根据详细数据层总结出来的某一年郑州市区域性大雨、暴雨各有几场等信息。

3.3 结构信息层

反映数据仓库的内部结构信息的元数据,它包括各种信息在数据仓库中的位置分布和处理方式等,以便检索查询。

4 数据集市

由于数据仓库中的数据往往要满足某团体内所有部门的分析需求,而建立一个数据仓库的成本、周期常常是巨大的,于是就出现了更简化的、面向特定应用的易于快速实现的数据集市。这样,一个单位的决策支持系统可由一个中心数据仓库和几个部门的数据集市组成,或由几个数据集市组成。

用户可以通过网络访问其它站点的数据集市的数据,从而形成一个分布式的数据仓库系统。

区别于数据仓库,数据集市的特点为:就用户和数据库而言是小的;基于特定的应用,快速实现;由业务部门定义、设计、开发、管理和维护;可升级到完整的数据仓库。

5 结 语

从本质上讲,数据仓库是一个使用其自身数据库管理系统的数据库应用,但其实现还需要高性能的数据库服务器、并行数据库技术以及数据库的互操作技术的支持。随着对这一领域的深入研究,数据仓库技术会变得越来越完善,越来越易于被人们掌握和利用。

参考文献:

- [1] 时密林,胡 明,张 旻,等. 企业数据仓库与决策支持系统[J]. 吉林工学院学报,2001,22(2):35-38.
- [2] 曹 辉,蔡 颖,朱善君. 数据仓库技术在电力系统中的应用[J]. 华北电力技术,2001,(3):44-54.

The Principle and Application of Data Warehouse Technology

Li She-zong, ZHAO Hai-qing, MA Qing-rong, HU Jun-xia, YANG Guang-xian

(The Zhengzhou Meteorological Bureau, Zhengzhou 450005, China)

Abstract: Analyzing about the fulfillment technology of traditional database and Data Warehouse, discussing the latter's main features, layer-structure of data storing.

Key Words: Data warehouse; Data character; Data structure

宽带网——电视天气预报节目的最佳传递方式

王其英, 张利君, 喻谦花

(开封市气象局, 河南 开封 475004)

目前大多数地市级电视天气预报节目是由当地气象部门独立制作完成的,载体一般采用录像带或数据光盘,并以人工方式进行传递。由于受工作性质影响,电视台和气象部门的地理位置很少能够彼此靠近。电视天气预报节目从制作、传递到节目播出完成,需要耗费大量的时间、人力和财力。

气象节目制作到播出的程序为:气象部门准备好电视天气预报节目内容(其中包括图像、声音等文件);用编辑系统对天气预报节目内容进行编辑,然后合成视频文件;录制录像带或刻录数据光盘;对制作好的天气预报节目录像带或光盘进行审核;派专人将节目载体送至当地电视台进行播出。

用这种程序制作的节目缺点是画面质量不高,因为通过录像机录制的信号为模拟信号,其质量远不如数字信号,另外录像带长期反复使用受损,致使图像失真严重;需要专人专程送节目载体,浪费了人力和时间;还要考虑日常交通工具等,

浪费了资金;如果中间环节出现问题,将难以保证天气预报节目正常播出。

目前我国宽带业务已经得到广泛应用,理论上其传输速度已达到上行1M带宽和下行8M带宽的速度,终端用户被分配的是固定IP地址,因此用户间可实现点对点传输。电视台和气象部门只需各申请一条宽带专线,任何一方开放一种传输服务或借助第三方传输软件,就可实现点对点的传输。因此,电视天气预报节目通过宽带传输将成为可能。气象部门需要做的只是将天气预报节目合成为视频文件,在固定时段内通过宽带网络将其“发送”到对方即可。

其优点是:因无需载体,节目可以实现数字化,画面质量可得到质的飞跃;由于宽带网速度快,可大大缩短人工传递所耗费的时间;可以节省前期设备投入(载体制作设备);减少人力及相关的资金浪费;节目的安全性、可靠性得到保障。