

数据仓库在森林资源决策支持系统中的应用

杨卫民, 梁红旗, 谭骏珊

(中南林学院 电子与信息工程学院, 湖南 长沙 410004)

摘要: 数据仓库技术是目前数据库技术研究的热点问题之一。介绍了数据仓库的设计和联机分析处理, 提出了建立全国性的分布式森林资源数据仓库的方案。讨论了如何在 Microsoft SQL Server 2000 中开发基于数据仓库的森林资源决策支持系统。

关键词: 数据仓库; 联机分析处理; 决策支持系统

中图分类号: TP392

文献标识码: A

文章编号: 1005-3751(2006)04-0187-03

Application of Data Warehouse in Forest Resource Decision Support System

YANG Wei-min, LIANG Hong-qi, TAN Jun-shan

(Coll. of Electronics and Info. Eng. of Central South Forest Univ., Changsha 410004, China)

Abstract: Data warehouse technology is one of hot problems of database technology. Here introduces design of data warehouse and online analytical processing. It also presented to be establish distributing forest resource data warehouse based whole country. How to develop forest resource decision support system based on data warehouse in Microsoft SQL Server 2000 is discussed.

Key words: data warehouse; online analytical processing; decision support system

0 引言

决策支持系统是在管理信息系统的基础上发展起来的, 是管理信息系统概念的深化。它主要用来解决结构化和半结构化的问题。它综合运用了信息论、人工智能、管理科学和行为科学等学科的理论、方法和技术。传统的决策支持系统以模型库为主体, 通过定量分析进行辅助决策。数据仓库技术是决策支持的新技术, 利用它可进行统计分析和预测并且能从大量的历史数据中发现隐藏在数据背后的规律。

各级林业部门经过长期的森林调查获得了大量数据, 面对如此多的海量历史数据如何从中得到有价值的信息进行科学决策是林业管理部门急需解决的问题。数据仓库技术引入林业信息系统中将给林业决策支持系统注入新的活力。

1 数据仓库

数据仓库是一个面向主题的、集成的、稳定的、不同时间的数据集, 用以支持经营管理中的决策制订过程。它

以多维表或多维数据库结构形式来组织数据, 提供决策分析服务。同时, 数据仓库能将分布在网络中不同站点的数据集成到一起^[1]。

数据仓库包含了早期细节级、当前细节级、轻度综合级的数据, 它通过 OLAP 分析工具和数据挖掘技术实现数据分析。数据仓库的开发方法不同于 OLTP 系统, 其开发过程是一个数据驱动的过程。

森林资源数据来源于各林场的二类调查资料或森林普查, 各林场一般建立了自己相对独立和完整的事务处理系统, 也需建立自己的局部数据仓库, 以支持经营决策活动。各村、乡、县、地区、省以至全国要对整个的经营活动进行监控, 做出战略决策, 也需要建立全局数据仓库。若采用集中式的数据仓库方案, 会给数据仓库的使用和维护带来很多不便。建立分布式数据仓库是最优方案, 它能较好地满足全国性的森林资源决策的需要, 主要表现在: 各林场根据自己的业务需要建立数据仓库便于数据分析, 数据维护也比较方便; 全局数据仓库按照整体分析需求, 从各局部数据仓库中抽取数据不与各林场的业务系统发生联系, 转换效率高^[2]。

各林场通过数据的抽取、转换和加载进入自己局部的数据仓库。数据是各森林经营单位为掌握森林资源现状及动态、分析和评价经营活动而进行的资源清查中获取的数据。中心数据仓库是通过 Internet 网络从各个林场的局部数据仓库中抽取的为全局战略决策分析需要的数据。

收稿日期: 2005-07-25

基金项目: 中南林学院青年科学基金项目(05038B); 湖南省科技厅项目(03JZY3023)

作者简介: 杨卫民(1969-), 男, 湖南湘乡人, 副教授, 硕士, 从事数据仓库、数据挖掘等方向的研究。

通过数据挖掘和 OLAP 分析对森林资源进行统计分析和知识获取,然后以可视化的形式表达给中高层管理人员。用户分析、报表、查询工具为用户进行分析决策使用的工具,因此其所有操作要非常简单,但提供的功能却要非常强大。

2 联机分析处理

数据仓库是存储用于数据分析的数据场所,而联机分析处理是以查询、分析为特征的。OLAP 系统可以提供给用户强大的统计、分析和报表处理功能及进行趋势预测的能力。OLAP 是将数据仓库的数据转移到多维结构中,在进行多维分析时,分析的目标数据称为事实数据,如森林资源中的面积、蓄积等,这些数据量比较大。事实数据的属性称为维,如森林资源数据中的时间、林地、林种等。维是有层次的,例如时间维可由年、月组成维层次。在多维结构中,维即观察的角度,它用超立方体或多立方体结构来描述。在关系型的数据仓库中,数据的多维视图主要是通过星型模式和雪花模式等来实现。多维分析对多维形式组织起来的数据采取切片、切块和旋转等各种分析动作剖析数据,最终用户能从多个角度、多个侧面观察数据库中的数据。

3 数据仓库的设计

数据仓库的设计也可采用三级建模方式^[3],即:信息包图的设计、逻辑模型设计和物理模型设计。下面以林种面积、蓄积统计分析主题为例进行讨论。

3.1 概念模型设计

概念模型设计就是通常所说的需求分析,通过对现实世界要处理的对象(组织、部门企业等)进行详细调查,在与用户交流的过程中,确定数据仓库所需要访问的信息,然后根据需求分析结果和企业数据流程图建立信息包图。信息包图提供了一个在多维空间建立用户信息模型的方法。信息包图拥有 3 个重要的对象:指标、维度和类别。指标表明在维度空间衡量商务信息的一种方法,而类别是在一个维度内为了提供详细分类而定义的,其中的成员是为了辨别和区分特定数据而设。图 1 为林种统计分析信息包图。

3.2 逻辑模型设计

数据仓库的需求分析阶段得到了信息包图,采用星型图建模技术可以为数据仓库建立完善的逻辑模型。同信息包图中的 3 个对象相对应,星型图包括 3 个逻辑实体:维度、指标和详细类别实体。指标实体位于星型图的中心,对应信息包图中指标对象是用户最关心的基本实体和查询活动的中心。一个指标实体代表一系列相关的事实,它仅仅与每个相关维度的一个对应。位于星型图角上的是维度实体,对应信息包图中的维度对象。详细类别实体对应信息包图中的类别对象,这些实体与对应的事务数据库结构产生映射。

信息包图转换成星型逻辑模型关键是定义好三大实体,即指标实体、维度实体和详细类别实体。维度实体还要向外延伸至详细类别实体。在星型图中,用户通过维度实体得到指标实体的数据。

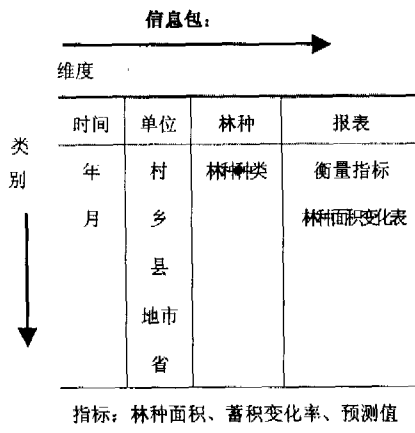


图 1 林种统计分析信息包图

3.3 物理模型设计

在物理设计阶段需确定如下内容:

(1)定义数据标准:在定义物理实体、关系和字段之前,首先应该明确命名约定,包括数据类型、约束条件、设备、索引、缺省等。

(2)定义实体:星型图可以很方便地确定面向主题的数据仓库共享实体,完整定义其属性,包括主键标、可选键标、外部键标、非键标数据、空值等。

(3)确定数据容量和更新频率:要对每一个数据仓库实体进行容量和更新频率的评估,容量包括实体预期的行和模式增加的容量。

(4)确定实体特征:完全识别实体特征很重要,这包括键标属性、值的有效范围完整性约束条件、类型和长度等。

4 应用实例

为了实现林种面积、蓄积统计分析,整个系统开发环境采用如下工具:

- (1)服务器操作系统为 Microsoft Windows Server 2000;
- (2)数据库管理系统为 Microsoft SQL Server 2000;
- (3)前端开发工具为 Visual basic 6.0, Microsoft Excel 2000;
- (4)OLAP 分析服务器为 Microsoft SQL Server 2000 Analysis Services。

4.1 森林资源数据仓库

森林资源数据仓库采用星型模式架构。它由一个事实表 and 一组维表来构成。事实表的主要特点是包含数字数据(事实),而这些数字数据可以汇总,以提供有关单位运作的历史数据。每个事实表还包括一个或多个部分组成的索引,该索引包含作为外键的相关性维度表的主键。而维度表包含事实记录的特性。事实数据表不包含描述

性信息。事实表的主码是组合码,维表的主码是简单码,每一张维表中的简单码与事实表组合码中的一个组成部分相对应。建立如图2的林种面积、蓄积统计分析星型模式并按此在SQL Server 2000中创建事实表和维表对应的关系表^[4]。通过使用SQL Server 2000中的DTS数据包将业务系统的数据导入数据仓库中。

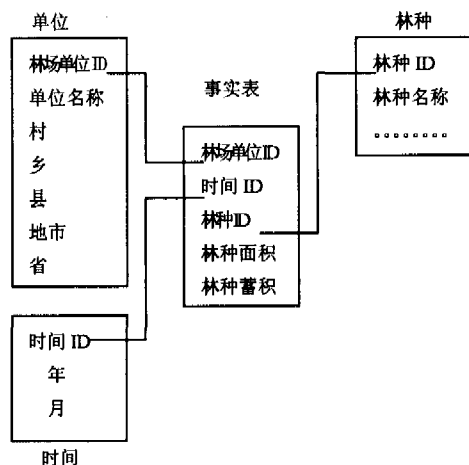


图2 林种面积、蓄积星型图

4.2 多维数据集的建立

SQL Server 2000中的森林资源数据仓库是多维数据集的数据源,由于它采用的是星型模型,它已经为多维数据集的创建做好了准备^[5]。在分析服务器中新建多维数据库,设置好与数据仓库forest数据源的连接。建立维度并定义好维层次(如图3所示)及建立多维数据集(如图4所示)。

4.3 用数据透视表服务实现Web的多维数据分析

在frontpage 2000或ACCESS 2000中建立Web页,插入数据透视表服务。利用数据透视表属性工具箱设置连接远程OLAP数据集。在Web页视图区加入各维及度量值。调整在浏览器中的布局,保存成Web页文件并把它分布到Web服务器中。

在浏览器中运行的页面用户可以通过属性工具箱设

置显示格式,也可以通过字段列表对话框进行多维分析。

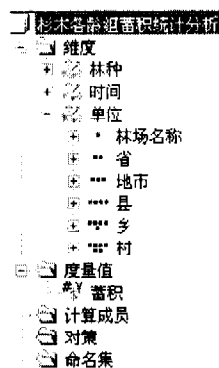


图3 维层次

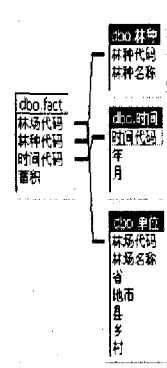


图4 多维数据集

5 结束语

森林资源是实现林业可持续发展的物质基础。如何经营管理好现有森林,实现森林可持续经营,进而实现林业可持续发展具有重要的理论和实践意义。数据仓库技术在林业行业的应用是林业行业应用新技术来提高信息管理水平和为林业发展战略规划提供科学依据和信息的有效方式。

参考文献:

- [1] 杨卫民,谭骏珊,汪斌.数据仓库和数据挖掘技术在DSS中的应用研究[J].计算机工程与设计,2004,25(1):1695-1697.
- [2] 韩览山,邵贝恩.面向集团型企业的混合型分布式数据仓库构建[J].计算机集成制造系统-CIMS,2003,9(1):80-83.
- [3] 张玉芳,熊忠阳.数据仓库数据模型的设计[J].计算机应用,1999,19(9):10-12.
- [4] 陈昌鹏,吴保国.林业数据仓库设计[J].农业网络信息,2004(4):30-32.
- [5] 罗运模.SQL Server 2000数据仓库应用与开发[M].北京:人民邮电出版社,2001.37-50.

(上接第186页)

(4)IP电话移动问题、地址管理比较困难,IP电话移动时是靠IP地址和MAC地址来定位的。

(5)如何向下一代IP网络(IPv6)平稳迁移的问题。

5 前景展望

作为NGN系列协议的重要组成部分,SIP体现了IP电话向商业电话模式演进的趋势,可以向用户提供强大的功能。SIP协议凭借其简单、易于扩展和便于实现等诸多优点,正越来越受到业界的青睐,逐步成为了NGN和3G多媒体子系统中的重要协议^[5]。尽管H.323仍占据着相当一部分VoIP的市场份额,但市场上将出现越来越多支持SIP的客户端软件和智能多媒体终端,以及用SIP协议

实现的服务器和软交换设备。尽管SIP目前仍处在发展和完善阶段,但其前景是光明的,它必将成为NGN的关键技术,基于SIP的电话系统也将得到广泛的应用。

参考文献:

- [1] 糜正琨.IP网络电话技术[M].北京:人民邮电出版社,2000.
- [2] IETF RFC3261-2002,SIP;Session Initiation Protocol[S].
- [3] IETF RFC3428-2002,SDP;Session Description Protocol[S].
- [4] 张智江,张云勇,刘韵洁.SIP协议及其应用[M].北京:电子工业出版社,2005.
- [5] 陈建亚,余浩.软交换与下一代网络[M].北京:北京邮电大学出版社,2003.