

电网污区图中智能化技术应用

姚楠¹, 王晓武²

(1. 江苏省电力试验研究院有限公司, 江苏 南京 210036;

2. 国投新集能源股份有限公司, 安徽 淮南 232007)

摘要: 为了实现对电网污区图的智能化管理, 以某电力系统为研究背景, 在对原有的数据源进行分析、处理和重新组织的基础上, 构建电网污区图数据仓库, 并建立多维雪花模式的数据立方体。运用在线联机分析处理 OLAP(On-Line Analytical Processing)和数据挖掘技术, 从多角度、多层次快速地分析和查询数据仓库的数据, 实现电网污区图管理的科学化, 并为污区图的实时滚动修订提供了技术基础。说明了 OLAP 数据仓库能够为电网管理人员提供有效的决策信息。

关键词: 在线联机分析处理; 数据仓库; 多维数据立方体; 电网污区图

中图分类号: TM 711; TP 311.131 **文献标识码:** B **文章编号:** 1006-6047(2007)08-0086-03

0 引言

伴随着工业的迅速发展和环境污染的日益严重, 各地区的污秽程度越来越高, 使得电力设备上的污秽物越来越多, 可能会导致设备污闪事故的发生, 这种现象在工业发达以及沿海等污染源较多的地方尤其严重^[1]。为了方便、迅速地掌握电网的线路、设备所在地区的污秽情况以及线路、设备的反污情况, 以便保持设备的健康状况, 减少事故的发生, 产生了污区图管理系统。经过多年的运行, 该系统积累了大量数据, 但是却没有有效的手段利用这些数据为决策支持服务。数据仓库(data warehouse)和联机分析处理 OLAP(On-Line Analytical Processing)由此应运而生, 它们能将来自异种源的数据统一(如聚集和汇总), 产生高质量的、纯净的和集成的数据, 自动地挖掘数据间潜在的模式, 并为决策服务^[2]。

在 OLAP 数据仓库设计与实现过程中, 将某电力系统日积月累的大量数据重新组织, 去粗取精, 经过提取、转换、集成、加载和汇总等预处理后, 存储到面向电网污区主题的数据仓库中, 并通过 OLAP 快速、交互地访问各种所需的信息视图, 进行高效多维分析, 从中挖掘出的知识能为电网生产管理以及污区、污源、盐密监测点的管理服务, 使得电网设备与污区及污秽等级之间的关系清晰明了, 盐密值统计、分析等工作变得简单容易, 大幅度提高了污区管理的水平, 快捷地得出有价值的电网决策信息, 更好地保证电力系统的安全、优质、经济运行^[3-4]。

1 OLAP 数据仓库系统设计

为了提高分析和决策的效率和有效性, 必须把分析型数据从事务性处理的数据库环境中提取出来, 按照决策支持处理的需要进行重新组织, 并进行再加工, 形成一个综合的、面向分析处理的体系化环

境。OLAP 数据仓库正是为了构建新的在线分析处理环境而出现的一种储存和组织技术^[5]。

1.1 总体结构

数据仓库就是一个用以更好地支持企业或组织决策分析, 面向主题的、集成的、不可更新的并随时间不断变化的数据集^[6]。所述数据仓库系统的体系结构采用客户机/服务器体系结构。该系统能对来自数据仓库的操作数据进行多维化或综合预处理, 它不同于传统联机事务处理 OLTP(On-Line Transaction Processing)软件的 2 层客户机/服务器结构, 而是 3 层客户机/服务器结构, 如图 1 所示。

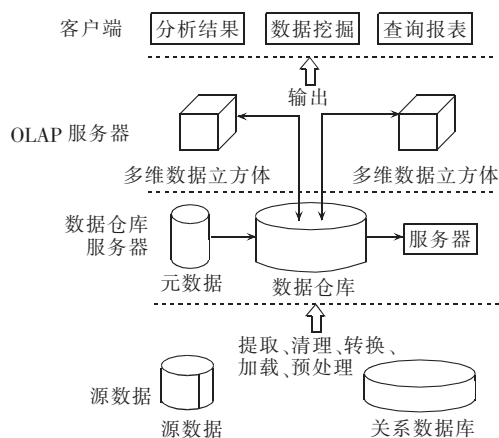


图 1 3 层数据仓库总体结构图

Fig.1 Structure of three-layer data warehouse

a. 底层是数据仓库服务器, 它从关系数据库系统及外部数据源中提取所需要的数据, 再经过预处理后存储到数据仓库。

b. 中间层是 OLAP 分析服务器, 它的主要功能有: 从数据仓库中提取数据、创建和处理数据立方体、管理大量历史数据、提供汇总和聚集机制, 并在不同的粒度级别上存储和管理信息; 根据关系数据库或数据立方体的内容建立数据挖掘模型, 将进行聚类、归纳、预测等处理后的信息一起存入知识库。

c. 顶层是客户分析层, 是面向用户的数据需求

的前端服务,能够方便地浏览数据仓库中的数据立方体,支持各种 OLAP 操作,如切片、切块、旋转、趋势分析、比较等处理。在客户端,PivotTable 服务程序直接用于服务器端 OLAP 通信,并提供了一个可以访问的接口。所有的应用程序都必须通过 PivotTable 服务才能访问 OLAP 服务器和数据挖掘所得到的数据。

这种结构的优点在于将应用逻辑、图形用户界面 GUI(Graphical User Interface)及数据库管理系统 DBMS(DataBase Management System)严格区分开来,复杂的应用逻辑不是分布于网络中的客户机上,而是集中存放在分析服务器上,由服务器提供高效的数据存取和分析处理。数据仓库系统将各种应用系统集成在一起,为历史数据统一联机分析处理提供坚实的平台。数据仓库侧重于存储和管理面向主题的数据,而 OLAP 则侧重于数据仓库中的数据分析和,并将其转换成决策信息。

1.2 源数据

源数据是从关系数据库和外部数据源为数据仓库提供运作的数据库,也是系统主要输入的数据源,包括污区、环境污染、输变电设备外绝缘、电网发展状况、变电站盐密和灰密、线路盐密和灰密、自然地理特征、经济发展状况和气象信息等^①。

1.3 数据预处理

各类源数据的来源不同,输变电设备外绝缘、盐密和灰密等数据来自于电力运行、检修、试验数据库,气象信息来自于气象台每天的定时发送,环境污染数据来自于环保局的月报,经济发展状况等来自于统计局的年报。这些原始数据很多不完整,因此需要对数据进行提取、转换、清理和集成。根据电网污区分析的需求,通过数据转换工具将不同的数据按照一定的规则,纠正、排序、聚合、综合、集中后加载到数据仓库中,从而保证数据的完整性,达到充分利用各种数据源的目的。

1.4 数据仓库模型

目前,最流行的数据仓库多维数据模型有 3 种^[7]: 星系模型、雪花模型和事实星座模型。此处的电网污区图数据仓库设计采用雪花模型。雪花模型是面向联机分析的一种多维化的数据组织形式,模型图形类似于雪花的形状^[8]。将多维数据库中的多维结构划分为 2 类表:一类是事实表,用来存储事实的度量值和各个维的码值;另一类是维表,对每一个维而言,至少有一个表用来保存该维的描述信息,包括维的层次和成员类别等。这种模型易于维护,减少冗余,并节省存储空间。

数据模型是建立数据仓库的关键。这里采用多维数据雪花模型,它是数据的一种多维结构,由维表和事实表定义,围绕电网污区各种主题组织数据,主题由事实表确定,而数据模型由维度和度量值的集合进行定义。维度是关于一个组织所需记录的透视或实体,而度量值用多维数据集事实表中的列创建。

电网污区图数据仓库经过分析和预处理后,确定有几个主题:污闪故障、污源点、盐密度、灰密度、

输变电设备外绝缘等,如图 2 所示。在图 2 中以污闪故障(事实表)为核心,与之相关的时间、地点、污闪设备、天气(维表)。每个维表有自己的属性,维表和事实表通过维关键字相关联^②。

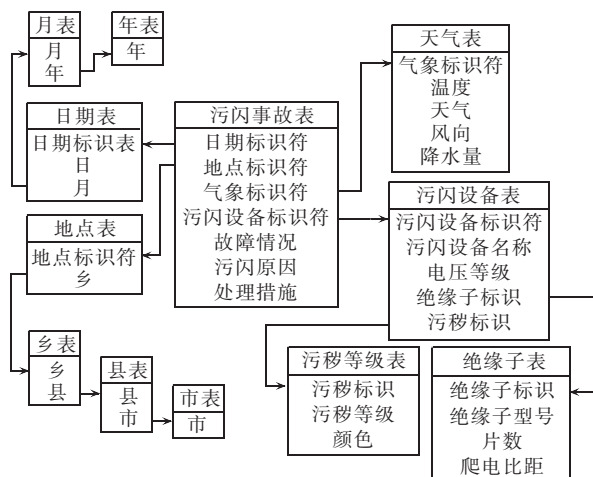


图 2 电网污闪事故主题数据仓库模型图

Fig.2 Data warehouse model of power grid pollution flashover faults

2 创建 OLAP 立方体

文中数据仓库和 OLAP 是基于多维数据模型。该模型将数据转化成数据立方体形式。数据立方体允许以多维对数据建模和观察。为了选择合适的维度和度量,创建数据立方体,以便从立方体中挖掘出决策支持信息。

2.1 定义立方体

用一种基于 SQL 的数据挖掘查询语言 DMQL 来定义立方体,对维和度量进行定义之后,便可以定义和创建与电网污闪事故类似的立方体。立方体的日期坐标轴来自日期维,天气坐标轴来自天气维,而污秽等级坐标轴来自污秽等级维。立方体上的某一个点度量是污闪事故数量。

数据立方体度量(measure)是一个数值函数,该函数可以对数据立方体的每一个点求值。通过对给定点的各维值对数据聚集,计算该点的度量值。度量用聚集函数表示,可以分成 3 类:1 类是分布聚集函数,如 count()、sum()、min()和 max();2 类是代数聚集函数,是若干个函数的代数运算,如 avg()、sum()、count()计算;3 类是整体聚集函数,整体函数的常见例子包括 Median()、mode()和 rank()。

OLAP 数据仓库存储方式有 3 种:多维 MOLAP(Multi OLAP)、关系 ROLAP(Relational OLAP)、混合 HOLAP(Hybrid OLAP),这里采用 MOLAP 存储方式。它将细节数值和合计数值都存放在立方体中,可将多维视图直接映射到数据立方体数组结构。这种存储方式的优点是对数据立方体的数据快速索引。

① 江苏省电力公司,江苏省电力试验研究院. 江苏电网污区分布图(2005 版)编制说明. 南京:江苏省电力公司,2005.

② 江苏省电力公司. 江苏电网污区分布图(2005 版)执行规定. 南京:江苏省电力公司,2005.

2.2 多层关联规则

由于多维数据空间数据的稀属性,在低层或原始层的数据项之间难以找到出强关联规则。在较高的概念层发现的强关联规则具有普遍意义的知识,为此,引入概念分层。一个概念分层指定数据仓库的一个维,它将低层概念映射到更高层概念,考虑时间的维概念分层,时间的节气值包括大寒、立春、大暑、立秋等 24 个值,每个节气可以映射到它所属的月份。例如,大寒可以映射到 1 月份,而立秋可以映射到 8 月份。概念分层是一种有用的知识形式,它使得原始数据在较高的抽象层上发现知识。使用概念分层可以进行数据挖掘。具有概念分层的关联规则挖掘产生的规则称为多层关联规则。

关联规则挖掘是寻找给定数据集中项之间的有趣联系。挖掘多层关联规则可以根据概念分层静态离散化,可以根据量化属性挖掘关联规则,量化特征规则可以由逻辑规则表示。如数据立方体可转换成逻辑规则形式:

$$\begin{aligned} \forall X, \text{Accident}(X) = 6 \Rightarrow \\ (\text{Time}(X) = \text{"July"})[t:w_1] \vee \\ (\text{Weather}(X) = \text{"Brume"})[t:w_2] \end{aligned}$$

该规则说明,如果 X 表示污闪事故数是 6,则 X 满足时间维 7 月份的可能性是 w_1 ,满足天气维大雾的可能性是 w_2 ,其中“ t -权”为兴趣度度量,“ $\vee$ ”表示对所有都成立。

2.3 OLAP 操作和分析机制

OLAP 的操作有几种:上卷(roll-up)是通过一个维的概念分层向上攀升或者通过维归约,在数据立方体上进行聚集。下钻(drill-down)是上卷的逆操作,它可以通过沿维的概念分层向下或引入新的维来实现。切片(slice)操作在给定的数据立方体的一个维上进行选择,导致一个子方体,固定天气轴,选定大雾天气。转轴(pivot)是一种目视操作,它转动数据的视角。

OLAP 提供了分析建模机制,包括推导比率、变差等以及跨越多维计算度量的计算引擎。它能在每一粒度级和在所有维的交叉产生汇总、聚集和分层。OLAP 也支持预测、趋势分析和统计分析的函数模型。综上所述,OLAP 是一种强有力的数据分析工具。

在多维数据模型中,数据组织成多维,每维包含由概念分层定义的多个抽象层,这种组织可使用户灵活地从不同角度观察数据。有一些 OLAP 数据立方体操作形成不同视图,允许交互查询和分析数据。

3 结语

利用丰富的数据资源,将和电网污区相关的数据经过提取、转换、净化、加载预处理后存储到数据仓库中,并建立了数据仓库雪花模型和 OLAP 多维数据立方体,进行高效多维分析与集成。通过 OLAP 快速、交互地访问,可快捷地得出有价值的电网污区决策信息,即电网规划和管理人员可以进行跨区域和跨时间的多维分析,并及时掌握所需的各种信息。由此说明 OLAP 数据仓库的实现为电网防污闪工作的专业化管理,在电网规划中预先考虑污闪情况,通过正确的决策提高电网的安全、优质、经济运行提供了必要的辅助决策工具。

参考文献:

- [1] 郭浩,王汝英. 电网污区管理在地理信息系统中的实现[J]. 天津电力技术,2005(增刊):1-4.
GUO Hao, WANG Ru-ying. Realization of manage of power grid pollution areas in GIS[J]. Tianjin Power Technology, 2005(Supplement):1-4.
- [2] THEODORATOS D, SELLIS T. Designing data warehouses [J]. Data & Knowledge Engineering, 1999, 31(3):279-301.
- [3] 宿志一. 用饱和盐密确定污秽等级及绘制污区分布图的探讨[J]. 电网技术, 2004, 28(8):16-19.
SU Zhi-yi. Study on defining pollution classes and mapping pollution areas by saturated equal salt deposit density[J]. Power System Technology, 2004, 28(8):16-19.
- [4] 肖嵘,张宇. 用饱和盐密划分电网污区图研究的探讨[J]. 华东电力, 2006, 34(2):34-36.
XIAO Rong, ZHANG Yu. Research on mapping pollute areas for power grids by saturated salt concentration[J]. East China Electric Power, 2006, 34(2):34-36.
- [5] CHAUDURI S, DAYAL U. An overview of data warehousing and OLAP technology[J]. SIGMOD Record, 1997, 26(1):65-74.
- [6] 胡侃,夏绍玮. 基于大型数据仓库的数据采掘[J]. 软件学报, 1998, 9(1):53-63.
HU Kan, XIA Shao-wei. Data mine basing on large-scale database[J]. Journal Software, 1998, 9(1):53-63.
- [7] MALLACH E G. 决策支持与数据仓库系统[M]. 李昭智,李昭勇,译. 北京:北京电子工业出版社, 2001.
- [8] LEVENE M. Why is the snowflake schema a good data warehouse design[J]. Information Systems, 2003(28):225-240.

(责任编辑:康鲁豫)

作者简介:

姚楠(1976-),男,江苏镇江人,系统分析师,硕士,从事电力系统软件研究(E-mail:yaon@jsepc.com.cn)。

Application of intelligent technologies in pollution area in power grid

YAO Nan¹, WANG Xiao-wu²

(1. Jiangsu Electric Power Research Institute Corporation Limited, Nanjing 210036, China;

2. SDIC Xinji Energy Co., Ltd., Huainan 232007, China)

Abstract: With a real power system as the research background and based on the analysis, processing and reorganization of its original data sources, the data warehouse is formed and the multi-dimension data cubes in snowflake mode are established for the intelligent management of power grid pollution area chart. By using OLAP (On-Line Analytical Processing) and data mining technologies, the quick data inquiry and analysis from different users at different levels are realized for the scientific management and real-time updating of power grid pollution area chart. It shows that OLAP data warehouse may provide effective decision-making information for grid managers.

Key words: on-line analytical processing; data warehouse; multi-dimensional data cube; pollution area