

数据仓库技术在地区电力调度系统中应用的探讨

杜俊红, 滕欢, 代小翔

(四川大学电气信息学院电力系, 四川 成都 610065)

摘要: 分析了传统电力调度数据库系统在决策分析应用等方面的不足, 并提出了在现有数据库系统的基础上建立数据仓库的思想。文中详细介绍了该数据仓库的设计模型、结构以及实施原则, 并对其应用前景做了展望。

关键词: 地区电力调度; 数据仓库; 数据挖掘; 联机分析处理

中图分类号: TM73 **文献标识码:** A **文章编号:** 1003-4897(2004)07-0061-04

0 引言

地区电力调度同时担负着所在城市及城市所辖市县电网的调度运行工作和所在城市的配电网管理工作。由于历史的原因, 电网调度和配电网管理彼此独立, 并由不同职责部门负责。针对地区电网调度的 EMS (能量管理系统) 和配电网管理的 DMS (配电管理系统) 是彼此独立的两个系统, 多数软件由不同的生产厂家提供, 各种功能相对独立, 数据不一致问题突出, 相互之间的数据共享困难。而电力调度部门在进行决策和分析时需要分析大量的数据, 既要有输电网的数据, 又要有配电网的数据, 并要求能挖掘出数据之间的相互关联, 很显然这种传统的 EMS 和 DMS 数据模式不适合决策分析的需求。

随着电网规模的不断扩大和电力市场竞争机制的引入, EMS 和 DMS 各自积累了海量的数据, 如何更好地利用和管理这些日益庞大的同构和异构数据库, 并挖掘出数据之间的潜在联系, 帮助企业更好地分析和决策, 已成为地区供电企业日益紧迫的需求。数据仓库技术可以把企业内、外部数据进行有效的集成, 主要应用于分析型处理, 基于此, 本文提出了建立地区电力调度数据仓库的思想和应用模型。

1 数据仓库技术

1.1 数据仓库技术概要

数据仓库以改进后的数据库技术作为存储数据和管理资源的基本手段, 以统计分析技术作为分析数据和提取信息的有效方法, 通过人工智能、神经网络、知识推理等数据挖掘方法来发现数据背后隐藏的规律, 实现从“数据 信息 知识”的过程, 从而为企业经营管理阶层提供各种层次的支持”。

1.2 数据仓库在电力调度的应用特点

1) 面向主题: 调度部门的 EMS 和 DMS 是以优化事务处理的方式来构造数据结构的, 对于某个主题的数据常常分布在不同的数据库中, 这意味着访问某个主题的数据实际上需要访问分布在不同数据库中的数据集合。而数据仓库将这些数据集中在一个地方, 对应于某个主题的全部数据被存放在一张数据表中, 这样调度部门可以非常方便地在数据仓库中的一个位置检索包含某个主题的所有数据。

2) 数据集成: 地区调度部门的数据库系统中的数据是分散而非集成的, 由于应用软件提供的厂家不同, 分别建有各自的数据库, 导致了现有的数据属性存在很多同名异义、异义同名, 及数据格式、测量单位和数据代码含义不一致的现象。因此必须对进入数据仓库的数据进行清洗和转化, 保持数据的一致性。

3) 数据的稳定性: 在数据仓库中, 数据一旦写入就不再变化了, 数据仓库可以看成是一个虚拟的只读数据库系统, 用户以只读方式访问数据仓库。数据仓库在数据存储方面是分批进行, 定期执行提取过程为数据仓库增加记录, 但是这些记录一旦加入, 就不再从系统中删除。

4) 随时间变化: 数据仓库记录系统的各个瞬态, 从而在数据分析的时候再现系统运动的全过程。数据追加实际上只是增加在上次数据输入后数据库变化了的数据。

2 地区调度数据仓库模型

建立地区调度数据仓库的基本思想就是在现有 EMS 和 DMS 基础上, 把 EMS 和 DMS 中的海量数据进行抽取和转化后存入数据仓库, 并针对不同的主题在数据仓库中建立数据集市, 然后利用 OL TP

和数据挖掘软件对不同主题的数据进行分析和处理,从而帮助调度人员做出决策。现建立如图 1 所示的地区电力调度数据仓库模型,并对每个功能模块分析如下:

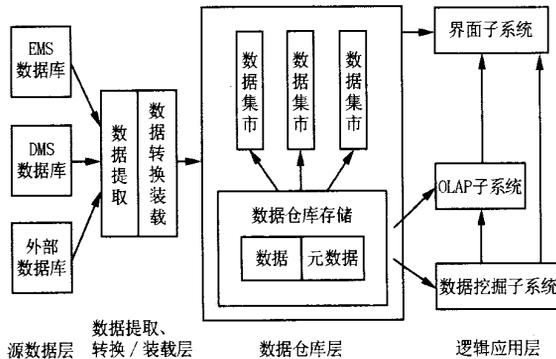


图 1 地区调度数据仓库体系结构

Fig. 1 Structure of data warehouse system for district electric power dispatching

2.1 源数据层

源数据层主要是指地区电力调度现在所拥有的数据库系统,即 EMS 和 DMS。另外,由于电力调度在分析和决策时需要用外部数据,如气象资料,省域网的部分资料数据等,因此,需要使用的外部数据也包括在源数据层内。

2.2 数据提取、转换/装载层

地区电力调度系统数据库中的数据量非常巨大,并不是所有数据都是分析决策所必须的,因此,只需用专用软件提取分析决策所必须的 EMS、DMS 数据和外部数据。另外针对原数据库系统中数据不一致的情况,必须对不一致的数据进行清洗和转换,使载入数据仓库中的数据和数据格式能够保持一致,供分析决策使用。

2.3 数据仓库层

2.3.1 数据仓库存储

数据仓库中存储了数据和元数据,其中数据的存储方式主要有虚拟存储方式、关系表存储方式和多维结构存储方式。由于虚拟存储方式效率差,而关系数据库的使用比较普遍,故采用关系表存储方式。使用 ORACLE 作为数据仓库设施,将数据存储于 ORACLE 的表结构中,并按星型结构来组织这些关系表。现具体介绍数据仓库中的数据、元数据、数据的组织结构和对数据的处理。

1) 数据

地区电力调度数据仓库中的数据指的是从 EMS、DMS 或外部数据库中提取,并经过清洗和转

化的数据。由于数据仓库主要用于 OLAP 分析和数据挖掘,因此需在原始数据的基础上增加冗余信息,进行预运算,建立多维数据库,以迅速转换数据。

2) 元数据

元数据是描述数据的数据,它描述了数据仓库中的数据和环境,遍及数据仓库的所有方面。它包括两种,一种是操作型环境向数据仓库环境转换而建立的元数据,包括所有源数据项的名称、属性及其在提取仓库中的变化;第二种元数据是数据仓库中用来与最终用户的多维商业模型和前端工具之间建立映射的,这种数据称为决策支持系统(DDS)元数据。在数据仓库中建立专用的元数据库来存放和管理元数据。

3) 电网数据组织结构

由于 EMS 和 DMS 中数据量非常庞大,因此有必要对数据进行综合。在数据仓库中,数据被分成 4 种级别,分别是高度综合级、轻度综合级、当前细节级和早期细节级。数据总是首先进入当前细节级,然后根据应用的需求,通过预运算将数据聚合成轻度综合级和高度综合级。若系统中的一些细节数据随着时间的推移已经老化,很少会被使用,可以将这些数据导出备份到设备上。数据的组织结构如图 2 所示。

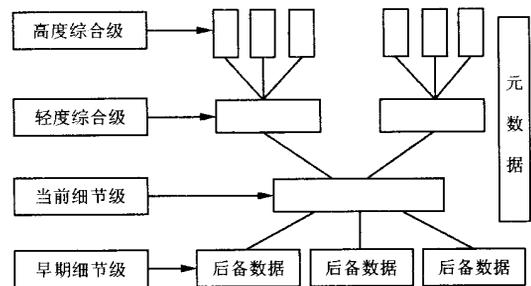


图 2 数据仓库的数据组织结构

Fig. 2 Structure of data organization for data warehouse

4) 数据的处理

数据仓库中一般存放 5 至 10 年的数据,若将全部数据放在一张表内,由于数据量太大,会降低数据访问效率,因此必须对数据表进行合理的分割。可按时间对表进行分割,在表中增加时间字段,去除与分析主题无关的纯操作型数据。

在数据仓库中,有些数据更新的较为频繁,如实时数据,而有些数据更新的时间较长,如设备信息等。因此,有必要按数据更新的频率对表进行划分,将不同变化频度的字段放在不同的表中,各表之间

使用相同的“标识号”进行关联,以节省存储空间。

2.3.2 数据集市

由于数据仓库中的数据量很大,若每次访问都要在海量数据中进行数据检索,会降低数据处理效率,故此可建立数据集市。数据集市是面向某一特定主题的、从数据仓库中逻辑上或物理上划分出来的数据子集。使用数据集市的主要目的是减少数据处理量。它具有面向部门、有特定的应用、规模小和实现速度快的优点。在地区电力调度中可建立的数据集市有:

1) 负荷预测

负荷预测是地区电力调度工作的重要环节,可分为系统负荷预测和母线负荷预测两类;而系统负荷预测按周期又分为超短期负荷预测、短期负荷预测、中期负荷预测和长期负荷预测。在地区电力调度中,EMS和DMS数据库系统中的历史数据的保存期限大多为一年,超过一年的数据则被导出另外存放,不利于在线进行长期和中期负荷预测。数据仓库能保存5至10年的历史数据,可运用数据挖掘技术对电力负荷数据进行深度挖掘,获得影响负荷的主要因素并对负荷进行有效的分类,然后再利用得到的知识建立负荷模型,从而提高负荷预测的精度。

2) 系统安全稳定性评估

电力系统有五种运行状态,分别为正常运行状态、告警状态、紧急状态、危急状态和恢复状态。可利用挖掘技术挖掘海量的电力系统运行数据,获悉电力系统在何种条件下处于正常运行状态、告警状态、紧急状态、危急状态和恢复状态,并对系统的安全稳定性做出评估,从而辅助调度人员做出决策。

3) 电力系统故障分析

地区电力调度部门已经积累了大量的故障数据。系统故障的发生既有偶然的一面,又有规律性可遵循。运用数据挖掘技术对电网故障数据进行挖掘分析,获得发生故障与气候变化和负荷变化之间的联系,从而辅助调度人员进行决策,合理安排检修计划,保证电网安全可靠运行。

4) 电力系统规划设计

电网规划是地区供电企业规划活动的基本环节,在规划过程中,需要处理大量的信息,可利用聚类、分类、关联、总结等挖掘工具挖掘模型和数据间的关系,为辅助决策系统增加约束条件,从而更合理地规划电网。

5) 电力用户特征分析

随着电力市场的发展和竞争的加剧,地区供电企业需要制定出有竞争力的实时电价,因此,必须对电力用户进行准确分析。电力系统的用户具有多样性和随机性,可利用OLAP多维分析的特性,对用户的行为模式及负荷要求进行分门别类,在保证系统安全和稳定性的前提下,对不同类别的用户,采用不同的营销策略,制定出合理的电价收费表,以提高市场竞争力。

2.4 逻辑应用层

逻辑应用层主要包括OLAP、数据挖掘和界面子系统。现具体分析如下:

2.4.1 OLAP子系统

在EMS和DMS中,调度人员通过联机事务处理(OLTP)和SQL可对数据库进行简单查询。随着时间的推移,电网规模不断扩大,EMS和DMS中的数据量也急剧增加,调度人员需要从多个维度来观察调度系统的运营情况,从而辅助决策。很显然,传统的简单查询模式已不能满足电力调度人员的需求。OLAP是针对某个特定主题进行联机数据访问、处理和分析,通过直观的方式从多个维度、多种数据综合程度将系统的运营情况展现给使用者。它通过对多维组织后的数据进行切片、切块、聚合、钻取、旋转等分析动作,剖析数据使调度人员能从多种维度、多个侧面、多种数据综合程度查看数据,了解数据背后的规律,从而辅助决策。

2.4.2 数据挖掘子系统

随着计算机技术在电力系统的广泛应用,地区电力调度部门已经积累了大量的运行数据和非运行数据,这些数据记录了地区供电企业多年的运行状况,电力调度人员急需对这些历史数据进行深入分析,以获得有价值的信息,辅助调度员进行决策,提高电网运行的可靠性。

数据挖掘是一个利用各种分析工具在大量数据中发现模型和数据间关系的过程。数据挖掘从大量数据中提取隐含在其中的人们事先未知、但又是潜在有用的信息,并以人们可以理解的模式展现。它的基本过程是首先对数据进行预处理,然后选择算法,提取规则,并对结果进行评价。

2.4.3 用户界面子系统

用户界面子系统主要是把OLAP和数据挖掘的分析结果通过友好的界面展现给最终用户,从而辅助用户进行决策。

目前数据仓库在电力系统的应用还处于探讨阶段,没有大规模的投入使用,因此在建设地区调度数

据仓库时,不应采用传统的瀑布式开发方法,而是采用螺旋式开发的方法,将一个庞大的任务划分成多个阶段,在每个阶段完成后,再开始新的开发,从而避免由于早期考虑不完善而造成巨大的经济损失。

3 结束语

由于数据仓库具有主要面向分析型应用,辅助企业进行分析决策的强大优势,数据仓库技术正在逐步应用于地区电力调度部门。在原有数据库基础上,建立地区电力调度数据仓库,利用 OLAP 和数据挖掘工具对数据进行深入的分析和挖掘,找出数据之间的潜在联系,从而在负荷预测、系统安全评估、电力系统故障分析、电网规划和电力用户特征分析等方面辅助调度人员进行决策,减轻调度人员的负担。然而数据仓库毕竟是一门新兴技术,其在地区电力部门的应用还处于初期阶段,很多功能还在探讨之中。随着数据仓库技术的进一步完善,数据仓库技术在地区电力调度系统中的应用也将进一步完善和发展。

参考文献:

- [1] 林宇,等(LIN Yu, et al). 数据仓库原理与实践(Theory and Application of Data Warehouse) [M]. 北京:人民邮电出版社(Beijing:People's Posts & Telecommunications Publishing House),2003.
- [2] 路广,张伯明,孙宏斌(LU Guang, ZHANG Bo-ming,

SUN Hong-bin). 数据仓库与数据挖掘技术在电力系统中的应用(Application of Data Warehouse and Data Mining Techniques to Power System) [J]. 电网技术(Power System Technology),2001,25(8):54-57.

- [3] 于之虹,郭志忠(YU Zhi-hong, GUO Zhi-zhong). 数据挖掘与电力系统(Data Mining and Power System) [J]. 电网技术(Power System Technology),2001,25(8):58-62.
- [4] 王军,吕震中,王培红(WANG Jun, Lü Zhen-zhong, WANG Pei-hong). 火力发电厂数据仓库的设计与实施(Design and Building of Data Warehouse for Thermal Power Plant) [J]. 中国电力(Electric Power),2001,34(3):57-60.
- [5] 廖志伟,孙雅明(LIAO Zhi-wei, SUN Ya-ming). 数据挖掘技术及其在电力系统中的应用(Data Mining Technology and Its Application on Power System) [J]. 电力系统自动化(Automation of Electric Power Systems),2001,25(11):62-66.

收稿日期: 2003-07-15; 修回日期: 2003-09-23

作者简介:

杜俊红(1975-),女,硕士研究生,从事电力系统调度自动化和软件应用研究;

滕欢(1964-),女,高级工程师,从事电力系统调度自动化和软件应用研究;

代小翔(1979-),男,硕士研究生,从事电力系统调度自动化和软件应用研究。

Application of data warehouse technology in district electric power dispatching system

DU Jun-hong, TENG Huan, DAI Xiao-xiang
(Sichuan University, Chengdu 610065, China)

Abstract: In this paper, the defects of the conventional electric power dispatching database system in application and decision making are analyzed, and a new idea of building data warehouse based on the existing database system is proposed. It illustrates the design model, structure and the implementing principle of data warehouse in detail. Finally the prospect of its application is predicted.

Key words: district electric power dispatching; data warehouse; data mining; online analysis processing (OLAP)

中国将成世界第五个建成新一代先进反应堆的国家

中国华能集团公司、中国核工业建设集团公司、清华大学日前签署建设高温气冷堆示范电站合作意向书。这标志着中国将成为继美、英、德、日后第五个掌握并建设这种国际公认的新一代先进反应堆的国家。

高温气冷堆效率高、安全性能好、可利用率高。目前国际上已建成 5 座高温气冷堆、2 座电功率为 300 MW 级的高温气冷堆示范电站。

中国第一座 10 MW 高温气冷实验堆由清华大学核能技术设计研究院自主设计、建设,于 2003 年 1 月建成并网发电成功,主要技术指标达到设计要求,现运行状况良好。中国核工业建设集团公司则承担了全部 6 座核电站、11 台核电机组反应堆工程的建设,已全面掌握了包括百万千瓦级核电站在内的多种堆型、多种规模的核电站建造技术。中国华能集团公司是由中央管理、截至去年底拥有全资和控股发电装机容量达 31360 MW 的国家骨干电力集团。