

商务智能在电网调度控制系统数据分析中的应用

宋鑫¹, 郭骏², 尹寿垚¹, 张勇², 张哲², 王茂海²

(1. 国电南瑞科技股份有限公司, 江苏省南京市 211106; 2. 国家电网公司华北分部, 北京市 100053)

摘要: 为了准确定位电网调度控制运行的异常信息, 需要从电网调度控制系统的大量历史信息中获取关键信息, 文中从商务智能(BI)的角度提出了海量数据统计分析的解决方案。首先阐述了 BI 的概念及在电网调度控制系统中推广应用困难的原因, 接着介绍了 BI 整合到电网调度控制系统中所必备的核心软件模块, 最后结合智能电网调度控制系统(简称“D5000 系统”)的调控业务, 以实例展示如何将 BI 融合到 D5000 系统中。

关键词: 智能电网调度控制系统; 商务智能; 星形模型; 抽取-转换-加载; DashBoard

0 引言

近年来, 电网调度控制系统建设快速发展, 智能电网调度控制系统(简称“D5000 系统”)已成为国内主流系统。随着智能电网建设的快速推进, 调度控制系统数据量呈几何级数上升, 给调度和自动化运行维护带来极大的压力; 保护信息、雷电、水电调度、广域测量系统(WAMS)等集成在 D5000 系统中, 为多业务数据关联分析提出了更高的要求; 系统中积累的大量历史数据描述了电力系统长期的运行状况, 蕴涵着电力系统运行的历史特性, 可以从中总结出丰富的经验及规则, 用于指导电网的运行^[1-3]。

如何使电网调控运行维护人员快速、准确地从“数据风暴”中获取关键信息, 已成为电网调度控制系统亟待解决的问题。本文结合国家电网公司华北电力调控分中心科技项目“BI 多维联动分析及钻取技术在调控一体化系统中的研究与应用”, 从商务智能(BI)的角度, 将 BI 技术与 D5000 系统有机地结合起来。首先介绍了 BI 的概念及其核心模块功能, 然后结合 D5000 系统现有的告警及历史数据, 结合多维数据建模理论, 设计并实现了灵活的数据展现方式, 可协助调控运行维护人员及时发现并定位异常数据, 为电网调控运行提供了有力支撑。

1 BI 概念

BI 的概念最早是 Gartner Group 于 1996 年提出来的, 当时将 BI 定义为一类由数据仓库、查询报

表、数据分析、数据挖掘、数据备份和恢复等部分组成的, 以帮助企业决策为目的的技术及其应用。BI 是对商务信息的搜集、管理和分析过程, 目的是使企业的各级决策者获得知识或洞察力, 促使其作出对企业更有利的决策^[4-6]。

目前数据挖掘在电网调度控制领域的应用往往通过对特定应用场景或模式匹配, 开发特定算法来进行分析^[7], 数据挖掘领域的经典算法, 如分类、聚类、决策树等算法, 在电网调度控制领域的实际应用尚处于研究阶段, 本文将重点放在常规数据统计分析处理上。传统的电网调度控制系统数据分析功能往往是开发特定的应用程序, 实现特定的分析展示功能, 而 BI 则通过数据抽取-转换-加载(ETL)、多维建模、通用用户界面(UI)等手段, 为各种数据分析主题提供了解决方案, 从而避免重复性的硬编码工作。

过去几年, BI 在银行、电信等行业有了实际应用, 但在电力系统领域, 仅在信通专业有部分应用, 主要是使用第三方 BI 软件进行相关数据分析及展示, 在电网调度控制领域几乎没有应用。究其原因, 虽然市场上有多种 BI 应用软件, 但是如果将其看成一种软件, 就会陷入“懂业务的不懂软件, 懂软件的不懂业务”的怪圈。笔者认为, 这是 BI 在电力系统中应用面狭窄的主要原因。如果换一种思路, 将 BI 看成一种思想, 即在多维模型的基础上, 通过数据多维观察、深度钻取等方法, 快速发现并定位异常数据, 从而为决策者提供判断依据, 则可以采用轻量级的方式, 研发并整合 BI 的部分核心软件模块, 并将其嵌入业务系统, 在电网调度控制领域快速推广 BI 的应用。

收稿日期: 2014-08-11; 修回日期: 2014-11-21。

国家电网公司华北分部科技项目“BI 多维联动分析及钻取技术在调控一体化系统中的研究与应用”。

2 BI 核心模块

2.1 多维模型设计

BI 与数据仓库理论是紧密结合的^[8],究其原因,主要是由于业务系统模型一般使用实体-关系(E-R)方法设计,符合数据库设计第三范式标准(3NF),具备高灵活性,但这种数据库建模方式呈网状结构,不利于大批量分析查询,且关系模型过于灵活,导致需要进行定制化开发不同的 UI 以适应不同的模型。

为解决上述问题,可以将关系型数据模型转换为以星形模型为代表的多维数据模型。星形模型是以一个事实表为中心,通过外键与一组维度表连接在一起的建模方式^[9]。星形模型是以查询为中心的建模方式,具备多维观察、数据钻取以及快速查询等关系型模型所不具备的特点^[10]。同时,由于星形模型有着固定的建模方式,可以研发通用的前台 UI,适应不同主题的星形模型。典型星形模型如图 1 所示。

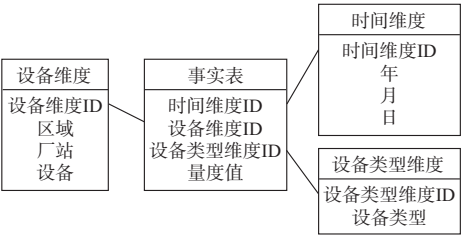


图 1 典型星形模型
Fig.1 Typical star model

如何在不影响系统运行的前提下,快速将 D5000 系统的关系型模型无缝转换为星形模型,并在此基础上研发通用的 BI 仪表盘展示工具,是本文支撑项目的研发重点。

2.2 ETL 软件设计

以往调度自动化系统多采用专用数据接口程序与外部系统进行数据交互,这种方式难以对数据交换流程进行监视,且在交互数据格式发生变化时,需要修改代码以匹配新的数据格式,为日常运行维护带来很大压力。为解决上述问题,本项目自主研发了 ETL 软件,基于开源 Spoon 软件架构,开发了适应 D5000 系统的 KeyID、告警、采样及统计等数据处理功能,结合开源常规数据处理 ETL 控件,能够在不开发任何接口程序的前提下,将 D5000 系统的模型、数据及告警无缝转换为多维数据模型。

本文支撑项目中研发的 ETL 软件具备如下特点。

1)支持多种数据源。能够以开放数据库互

连(ODBC)、Java 数据库连接(JDBC)等标准接口读写金仓、达梦及 Oracle 等数据库,能够处理 E 格式、Text 和 Excel 等常规文件。

2)支持多种控件。支持各种读、写、转换控件,特别提供能够处理 D5000 系统特殊数据结构的控件,实现关系型模型数据到多维数据模型的转换。

3)流程化管理。支持图形方式创建并浏览 ETL 过程,能够直观地监视 ETL 全过程,并能快速定位发生错误的位置。

4)任务调度。能够指定时间或周期,执行定制任务,并记录任务执行情况信息。

基于自主研发的 ETL 软件,可在不影响系统运行的前提下,无需进行代码级开发,实现关系型数据模型到多维数据模型的无缝转换,并能指定时间周期对模型和数据进行增量更新。需要说明的是,面对不同的分析主题,只需要针对特定的数据模型进行转换,并非针对整个系统的数据模型。例如,本文 3.1 节越限分析统计这个主题,所涉及的 D5000 系统数据仅仅涉及区域表、厂站表、母线表、交流线段表、变压器表、断面信息表及越限告警信息表。

2.3 DashBoard 软件设计

DashBoard 是 BI 仪表盘的简称,是将多个仪表、图表、报表等内容整合在一个可交互的可视化页面上的工具。在电网调度控制系统中,往往需要从多个观察角度的灵活组合来观察数据,从而发现运行数据的内在规律,这就是多维观察;在分析过程中,可能需要在统计结果的基础上,将数据进一步细化,定位异常发生的具体设备,这就是数据钻取。DashBoard 软件的主要功能是实现多维观察和数据钻取。

传统的电网调度控制统计分析软件,往往需要针对特定的分析主题,定制化开发相应的界面,究其原因,是由于其后台数据模型大多采用关系型数据模型,而关系型模型过多强调灵活性,这就使得前台 UI 难以适应各种业务模型。本文支撑项目中所研发的基于浏览器/服务器(B/S)架构的 DashBoard 软件,以模式相对固定的星形模型作为数据基础,支持自定义页面布局、图元联动及数据钻取等功能,用户可以自行设计并实现数据分析及展示功能。

3 BI 在 D5000 系统中的应用

作为现阶段国内应用范围最广的调度控制系统,本项目在 D5000 系统中的应用具有普遍意义。D5000 系统的统计功能分为两种:①基于历史采样数据统计;②实时统计。基于历史采样数据统计的优点是可以直接利用常规采样数据,无需事先定义

统计数据点,缺点是只能进行最大值、最小值、平均值等常规统计,且统计耗费时间长;实时统计功能可以进行时长、次数的统计,但需要事先定义统计数据点,且程序常驻后台,消耗较多的系统资源,一般只针对少量关键数据进行实时统计,难以满足对所有测点统计的要求。

在电网调度控制系统运行数据的统计分析过程中,很重要的一点是能够直观地看出异常,并定位异常发生的详细信息。如果对全部数据点进行统计,其结果中绝大多数信息是调度运行维护人员所不关心的,真正有用的信息可能会淹没在信息海洋中,而且由于异常发生位置的不确定性,也难以事先定义数据点进行实时统计。鉴于上述原因,可以首先根据告警类型或告警数量确定异常发生时的告警信息,然后根据具体时间段和数据点,结合采样信息对其进行详细分析。

3.1 越限统计分析

本应用场景对母线电压、线路负载、主变负载、断面负载进行统计分析,能够对某一时间段内越限时长进行统计,并可以从“区域→厂站→设备/断面”的角度对越限时刻及持续时间进行数据钻取,以获得越限详细信息。

传统的实现方法有:①开发特定的应用程序实现统计及展示功能,这会带来较多的工作量;②对所有相关测点/断面的历史数据进行实时统计,这会消耗大量的系统资源,且因为数据量庞大而无法保证统计效率。如果采用 BI 的实现方法则无需进行代码级开发,首先,使用 ETL 软件的时间序列分析控件,根据告警信息确定越限设备及发生时刻;其次,有针对性地获取并统计每次越限的持续时间及越限时段的历史采样数据,这样可以极大地缩小统计范围,保证统计效率;然后,使用 ETL 软件,将分析结果自动形成星形数据模型;最后,使用通用的 DashBoard 软件设计并实现展示页面。此外,在设计模型时,同时支持母线电压、线路负载、主变负载、断面负载的越限分析,可以有效保证分析统计功能的通用性。

对告警进行分析时,使用基于自主研发的时间序列分析控件,对模拟量越限时长和发生时刻进行统计,最终形成了包含设备维度、时间维度及越限类型维度的模拟量越限统计模型。设备维度包含区域、厂站及具体设备 3 个自上而下的层级;时间维度包含年、月、日 3 个自上而下的层级;越限类型维度只有一个层级,共有主变、断面、线路及母线 4 种常规越限设备类型。

此模型支持从时间维度进行“年→月→日”钻

取,支持从设备维度进行“区域→厂站→设备”钻取,从越限类型维度可以对某一越限类型进行统计,或在多个维度上进行切片、切块操作。

基于此模型设计了母线电压越限数据展示页面,确定统计时段后,首先可以通过饼图概览指定时间段内各区域母线电压越限持续时间的分布情况;点击饼图中的指定区域后,可以查看该区域内各厂站母线电压越限分布情况;点击相关厂站后,可以看到具体母线的电压越限持续时间;点击母线后,可以看到母线电压越限发生时刻及持续时间。通过这种自上而下的数据钻取,可以迅速获取所需的信息。具体数据如附录 A 图 A1 所示。可以看出,部分母线越限时长严重不合理,与电网运行实际情况不符,经过排查,是由于当前电网调度控制主站系统中母线电压遥测限值设置不合理导致,经过重新设定限值,避免了此类告警对调度员的误导。由此可见,通过 DashBoard 软件的直观展示效果,可以快速定位电网调控主站和子站存在的问题。

同理,基于此模型,可设计断面、线路及主变的越限统计的数据展示页面。

3.2 遥信变位统计

本场景对遥信变位数量进行统计,可用于快速定位站端频繁误发遥信的设备,然后通过站端整改或在 D5000 系统中封锁遥信的方式,减轻误遥信对用户的影响。场景展示可以从“区域→厂站→设备/测点”的角度对变位次数进行数据钻取,以获得变位详细信息。

与上一个实例类似,传统的实现方法会带来较多的编码工作量,或消耗大量的系统资源进行实时统计。而采用 BI 则无需进行代码级开发,首先根据变位告警信息确定变位设备及发生时刻,然后有针对性地获取并统计变位信息,这样可以极大地缩小统计范围,保证统计效率,最终形成包含设备维度、时间维度及遥信类型维度的遥信变位统计模型。设备维度包含区域、厂站及具体设备 3 个自上而下的层级;时间维度包含年、月、日 3 个自上而下的层级;遥信类型维度只有一个层级,共有开关变位、刀闸变位、保护信号及测点遥信 4 种常规遥信类型。

此模型支持从时间维度进行“年→月→日→时”钻取,支持从设备维度进行“区域→厂站→设备”钻取,从遥信类型维度可以进行全部统计或对某类遥信进行统计,或在多个维度上进行切片、切块操作。

基于此模型设计了遥信变位统计数据展示页面,确定统计时段后,首先可以通过饼图概览指定时间段内各区域遥信变位次数的分布情况;点击饼图中的指定区域后,可以查看该区域内各厂站遥信变

位分布情况;点击相关厂站后,可以查看具体设备或测点的遥信变位次数;点击设备或测点后,可以看到变位次数在每天的分布情况。通过这种自上而下的数据钻取,可以迅速获取所需的信息。

具体数据如附录 A 图 A2 所示。可以看出,绝大多数遥信由全球定位系统(GPS)脉冲发送保护动作及复归信号导致,此类信号对电网调度控制运行的意义不大,可以采用 D5000 系统的告警抑制功能屏蔽此类告警,以减轻调度运行维护压力。屏蔽此类遥信后,还可以进一步检查哪些设备或测点会频繁发出无意义的变位信息。

3.3 其他

采用类似上面例子的方法,还可以对告警直传、相量测量通道中断情况、事故情况、数据跳变、数据不刷新、交流线路输电效率等场景设计相应模型并进行展示。

4 结语

本文将 BI 与电网调度控制系统有机地融合在一起,以 BI 多维观察、深度钻取的思想为指导,研发了适用于电网调度控制系统的 BI 软件模块。在现有告警及采样数据的基础上,设计了基于星形模式的统计建模方法,直接定位需要统计的异常数据,有效缩减统计数据量和时间;应用研发的 BI 软件模块实现了基于标准星形模型的 DashBoard 客户端软件,并实现模拟量越限、遥信变位等统计分析场景。本文的研究成果已应用于国家电网公司华北调控分中心 D5000 系统。

但是,本文的研究与开发还存在以下欠缺之处。

- 1)对 BI 的高级应用功能——数据挖掘尚未深入研究。
- 2)统计数据使用的是 D5000 系统分钟级采样数据,精度不高,如果使用时间序列数据库采样,统计精度可与实时统计相当。

基于本文的研究与开发可以看出,与电网调控系统现有辅助决策软件的实时应用不同,BI 应用不需要开发特定的算法,主要使用电网调度控制系统常规海量历史信息,定位日常调度监控过程中容易忽略的问题,例如可将告警和历史采样数据结合起来进行统计分析,有力地补充了 D5000 系统现有的统计功能。此外,借助 ETL 强大的数据处理能力,BI 还可以应用于一些非标准格式数据的统计分析。

附录见本刊网络版 (<http://www.aeps-info.com/aeps/ch/index.aspx>)。

参 考 文 献

[1] 陆瑞华,胡蓉,刘林林.电网一流调度智能数据平台设计与应用[J].华东电力,2013,41(1):77-80.
LU Ruihua, HU Rong, LIU Linlin. Design and application of power grid best dispatching intelligent data platform[J]. East China Electric Power, 2013, 41(1): 77-80.

[2] 李秋丹,迟忠先,金妮,等.基于数据仓库的地区电网调度决策支持系统[J].电力系统自动化,2004,28(12):86-89.
LI Qiudan, CHI Zhongxian, JIN Ni, et al. Regional power dispatching decision support system based on data warehouse[J]. Automation of Electric Power Systems, 2004, 28(12): 86-89.

[3] 黄杰,孙莉.商业智能在电力公司决策支持系统的应用[J].计算机应用与软件,2010,27(1):221-223.
HUANG Jie, SUN Li. Research and application of business intelligence indecision support system for power company[J]. Computer Applications and Software, 2010, 27(1): 221-223.

[4] TURBAN E, SHARDA R, DELEN D, et al. Business intelligence: a managerial approach[M]. New Jersey, USA: Prentice Hall, 2010.

[5] INMON W H.数据仓库[M].王志海译.北京:机械工业出版社,2006.

[6] 王珊.数据仓库技术与联机分析处理[M].北京:科学出版社,1998.

[7] 闪鑫,戴则梅,曹路,等.针对电网 3/2 接线方式的设备风险在线辨识方法[J].电力系统自动化,2012,36(1):112-115.
SHAN Xin, DAI Zemei, CAO Lu, et al. An on-line identification of equipment risk for 3/2 circuit breaker connections[J]. Automation of Electric Power Systems, 2012, 36(1): 112-115.

[8] CHEN H C, CHIANG R H L, STOREY V C. Business intelligence and analytics: from big data to big impact[J]. MIS Quarterly, 2012, 36(4): 1156-1188.

[9] 张艳杰.BI 技术在国华电力企业信息门户系统中的应用[J].电力技术,2010,19(4):22-25.
ZHANG Yanjie. Application of BI technique in the enterprise information portal system of Guohua power[J]. Electric Power Technology, 2010, 19(4): 22-25.

[10] DUAN Lian, XU Lida. Business intelligence for enterprise systems: a survey[J]. IEEE Trans on Industrial Informatics, 2012, 8(3): 679-687.

宋 鑫(1981—),男,通信作者,高级工程师,主要研究方向:电网调度运行技术。E-mail: songxin@sgepri.sgcc.com.cn

郭 骏(1983—),男,高级工程师,主要研究方向:电网调度运行技术。E-mail: guo.jun@nc.sgcc.com.cn

尹寿焜(1984—),男,工程师,主要研究方向:电网调度运行技术。E-mail: yinshouyao@sgepri.sgcc.com.cn

(编辑 杨松迎)

(下转第 145 页 continued on page 145)

(上接第 96 页 continued from page 96)

Application of Business Intelligence in Historical Data Statistical Analysis of Power Dispatch and Control System

SONG Xin¹, GUO Jun², YIN Shouyao¹, ZHANG Yong², ZHANG Zhe², WANG Maohai²

(1. NARI Technology Co. Ltd., Nanjing 211106, China;

2. State Grid North China Power Dispatch and Control Center, Beijing 100053, China)

Abstract: In order to accurately find abnormal information on the power dispatch and control system operation, it is necessary to obtain key information from large amounts of historical data. From the perspective of business intelligence, a solution for massive data statistical analysis is proposed. First, the concept of business intelligence and the cause of difficulties in its popularization in the power dispatch and control system is explained. Then, a description is made of the essential core software modules for integrating business intelligence into the power dispatch and control system. Finally, the way business intelligence is integrated into a D5000 system based on the dispatch and control business is demonstrated.

This work is supported by State Grid North China Power Dispatch and Control Center.

Key words: smart grid power dispatch and control system; business intelligence; star model; extract-transform-load (ETL); DashBoard