

OLAP 在供电行业的应用与研究

华院分析技术（上海）有限公司 咨询顾问 徐卫华

(同济大学经济与管理学院 上海, 200092)

[摘要] OLAP 技术是在企业界和软件开发界受到广泛关注的一个研究方向, 它代表为提高企业决策能力和运营能力而采用的一种方法、技术。本文研究的就是利用 OLAP 技术中的多维数据分析, 对供电行业的营销数据进行多层次、多角度、全方位的分析, 揭示电力市场和营销的内在规律。科学地、快速地指出企业存在的问题、隐患并能发现对于企业至关重要的变化趋势, 形成极具决策价值的战略信息。

[关键词] 联机分析, 分析处理, 决策支持, 分析建模

Application and Research on OLAP in Electric Power Industry

XU Weihua, HUANG Liping

(School of Economics & Management Tongji University SHANGHAI, 200092)

ABSTRACT: OLAP technology is a hot topic both in enterprises circle and software development area, which is a methodology and technology standing for the capability of enterprises' decision and operation. This paper studied how to carry out multi-level, multi-angle and all-around analysis for external and internal data to disclose inherent discipline of electricity market and marketing by using the OLAP technology. It can point out hidden problem and can discover important changing tendency for enterprise, then form the high-value strategic information.

KEYWORDS: OLAP, analyzing and processing, DSS, analyzing and modeling

1、引言

在当今竞争日趋激烈与信息瞬息万变的市场经济中, 几乎所有行业都面对着激烈的竞争, 正确及时的决策是企业生存与发展的最重要环节。在我国, 供电企业的信息化建设已经有了一定的基础, 但随着信息化建设的进一步深入, 如何有效地利用供电企业在多年用电营业管理过程中产生的大量数据和信息, 如何全方位、多角度、多层次地对大量已有数据进行分析, 从中提取制定供电企业发展战略和市场策略的信息, 来为企业的经济和决策服务就成了供电企业信息系统建设中要上新台阶所面临的重要课题。尽管供电企业在其信息系统中建立了庞大的数

数据库，但面对众多的数据，高层决策者仍然束手无策，仍然难以准确获取决策所需的信息。究其根本原因在于：数据并不等于所需的信息。拥有大量的数据并不等于可以及时准确地找到决策者想要的答案。如果说几十年前决策失误的原因是由于数据的匮乏。到了现在，则很可能是源于对数据的分析能力不够。那么怎样把大量的数据转换成动态的、可靠的、商用的信息以便于决策支持呢？基于数据仓库的 OLAP (On-Line Analytical Processing 联机分析处理) 正广泛地被公认为是最好的解决方案。

2、OLAP 技术优势及其实施步骤

OLAP 是使分析人员、管理人员或执行人员能从多种角度对从原始数据中转化出来的、能够真正为用户所理解的、并真实反映企业维特性的信息进行快速、一致、交互地存取，从而获得对数据的更深入了解的一类软件技术^[2]。

用户-分析人员总是从不同的角度来看待、衡量企业数据, 因此企业数据空间本身就是多维的。决策所需的数据总是与一些维数(如地区、时间)和不同级别(如部门、领域、地区和国家)的统计(或计算)有关, 可以说, 以多维数据为核心的 OLAP 是决策的强力支柱。OLAP 通过与用户现有数据的无缝连接, 通过快速的处理手段, 简洁的操作方法, 将数据展现在你的面前。拖动鼠标即可实现不同方位的数据透视, 配合数据的钻取技术, 你所关心和会引起你关注的信息尽览无余。查询、报表、分析、图表所有的工作变的简单、容易。真正摆脱了依赖信息技术人员“预制”报表的束缚, 把信息使用的管理者用智能化的工具“武装”了起来, 创造了一个轻松的数据分析环境, 数据的价值得到真正提升^[3]。

因为行业存在差异性, 以及数据分析的侧重点各不相同, 所以多维数据分析不存在必须遵循的严格的流程。但是, 千变万化不离其宗, 在进行电力营销数据多维分析的时候, 我们遵循了以下几个最基本的步骤^[6]:

- (1) 明确多维分析的具体内容, 越全面越好。这部分的主要工作就是明确供电企业用户的需求, 尤其是企业决策者的需求, 这是整个数据分析过程中最重要的一步。只有在明确需求的前提下才能提取相应的数据形成数据集用于分析建模。
- (2) 明确数据源存储方式。供电企业在其信息系统中建立了庞大的数据库, 但是我们主要关注的是电力营销数据, 因此必须了解我们所需要的数据存在于哪个数据库中, 以及存取形式、字段类型、字段含义等等。
- (3) 提取数据, 生成中间表。为了避免影响日常管理信息系统的正常使用, 分析所需要的数据应该和原始数据库隔离开。由于受条件的限制, 我们没有使用数据仓库, 而是把分析所需数据经过抽取、转换、清理等一系列工作后加载到新的数据库中。
- (4) 建立多维模型。根据需求选择相应的分析模型, 这是直接影响数据分析质量的关键一步, 因此需要熟练掌握各种统计分析模型, 并在此基础上按照

实际需求选择最适合本次分析的模型。

- (5) 选择或开发前端展示工具。为了更好的展示分析结果，根据实际需要，可以自主开发前端展示界面，也可以选择市面上业已成熟的展示工具。
- (6) 根据用户需求，定制当前报表和图形。数据分析的最终结果是以图形或者报表的形式展示出来的，可以选择直方图、折线图、饼图等形式。

3、应用 OLAP 技术分析电力营销数据

本次电力营销数据分析的主要目的是在现有数据中找出影响电量增长的主要因素这样一些信息，已经罗列的电量增长点考虑的因素包括：业扩（新装，增容等）、天气因素（温度）、国民经济因素、自然增长、政策因素等，并不旨在发现不可预测且有价值的结构，因此使用确定性分析。首先对影响电量的因素进行分类并利用 OLAP 中的有关方法进行有效性验证；然后给出每个分类中各个分块的电量增长量随时间变化的曲线；最后定位影响电量增长的主要因素。

在多维数据分析中，数据整理是最重要的一步，但是它也是经常被忽视的一步。事实上，将大量的业务数据应用于分析和统计原本是一个非常自然和自然的想法。但在实际的操作中，我们发现要获得有用的信息并非想象的那么容易，由于目前供电企业的应用系统基本上都是以事务处理过程为基础建立起来的，事务处理过程是面向工作流的，面向事务处理过程的应用系统产生的数据带来了它“先天的缺陷”^[5]。

(1) 蜘蛛网现象：在企业中，随着组织的发展和业务的开拓，由于各种应用的需要，根据各个事务处理过程开发了很多应用程序，这些应用程序彼此之间存在数据的交流。事务处理环境中数据的交流，会导致大量复杂、集成性很差的数据，形成一个蜘蛛网，主要表现在数据算法上有差异、多层次的数据抽取、没有统一的外部数据。

(2) 现有数据内容与结构存在一定程度上的不一致。由于种种原因，供电企业数据库中已有的数据在某种程度上存在不一致的问题，它会在数据分析的初级阶段花费较多时间来修正、整理数据，以符合分析的要求。

(3) 存在大量零数据：根据现有观察发现，所考察的度量值中存在大量的零电量数据项，造成数据库中大量信息的浪费，这些数据项将对分析过程产生一定的干扰，需要找出方法判断零电量的情况。我们的解决方法是寻找零电量比例，将它们暂时剔除，做零电量可能造成干扰的分析，如单变量频数分布分析等。

在数据建模方面，我们选择了 Microsoft SQL Server 2000 Analysis Services 中的 OLAP 多维数据集模型。Analysis Services 是微软公司数据仓库解决方案的核心，是进行联机分析处理和数据挖掘的一个中间层服务器。分析服务和数据透视表服务提供了对数据仓库的数据立方体及数据挖掘模型的设计、创建和管理等方法和工具，并且提供了客户端对这些 OLAP 和数据挖掘数据的访问机制。分析服务在数据仓库的事实表和维表的数据上创建了多维立方体，立方体

可包括汇总数据、复制数据，或者是直接从数据仓库中读取数据。在多维立方体的建立过程中，数值型数据也进行了一定的预聚集，以适应各种不同的查询方式^[4]。作为一个比较成熟的产品，微软 Analysis Services 有很多优势，比如与数据仓库紧密结合、对多种结构多维数据库的支持、开放式的体系以及支持基于 Web 的应用。同时，由于供电局的大部分软件应用程序都是微软公司的产品，为了更好的兼容，我们最终选择它作为后台分析服务器。

在对数据进行抽取、转换、清理、加载及选取分析工具后，开始进行关键的一步：分析建模。数据分析建模是以数据分析为目的的构建面向数据的模型的过程。简而言之，就是构建一个模型，然后通过模型组织并汇总数据库中的数据，以便能够利用联机分析和图形工具迅速对数据进行评估，给用户更直观的分析结果。其中核心度量值 (Measure) 为电量和金额，在此基础上，我们同时选择了户数和容量作为度量值。维 (Dimension) 将包括用作数据细分的各个标准，如时间、用户类型 (大、中、小)、业扩类型、政策因素等。分析方法选择了两种：同期相比和环比。建立后的多维模型架构如图 1 所示。

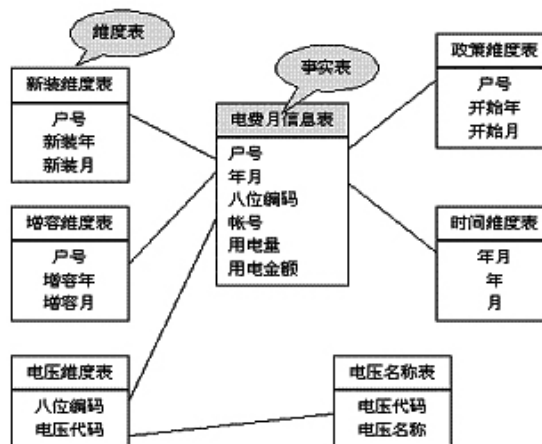


图 1 多维模型架构图

在数据展示方面，我们自行开发前端展示工具（见图 2）。前端展示工具与后台的 Analysis Server 相连，以多种形式提供前台的分析结果展示（条形图、线图、区域图等），给决策者直观的感受，从而补充 Analysis Server 在图形方面的不足。前端的客制化工具能够提供海量数据的处理、减少用户等待时间、实现即时响应，并使得分析结果的报表、图表实时输出。整个系统分为五个分析功能模块和一个系统功能模块，分别从不同角度查看数据分布趋势。它们是：

- 时间波动分析：大、中、小用户按照时间的总体波动趋势分析。可根据需要选择不同用户类型、分析指标、分析方法和时间跨度查看数据和曲线。
- 电压等级分析：大、中、小用户按照电压等级再次细分后的随时间的总体波动趋势分析。可根据需要选择不同用户类型、分析指标、分析方法和时间跨度查看数据和曲线。
- 业扩类别分析：大、中、小用户按照不同业扩类别和业扩发生时间挑选出的部分用户随时间的总体波动趋势分析。可根据需要选择不同用户类型、业扩类别、分析指标、分析方法和时间跨度查看数据和曲线。
- 政策因素分析：大、中、小用户中受政策因素影响的部分用户的随时间的总体波动趋势分析。可根据需要选择不同用户类型、分析指标、分析方法和时间跨度查看数据和曲线。

- 总年度分析：按照年度，汇总前面各模块的分析数据，合并显示各因素为电量/电费带来的大中小用户总增长。可根据需要选择不同分析指标、年度查看数据和曲线。

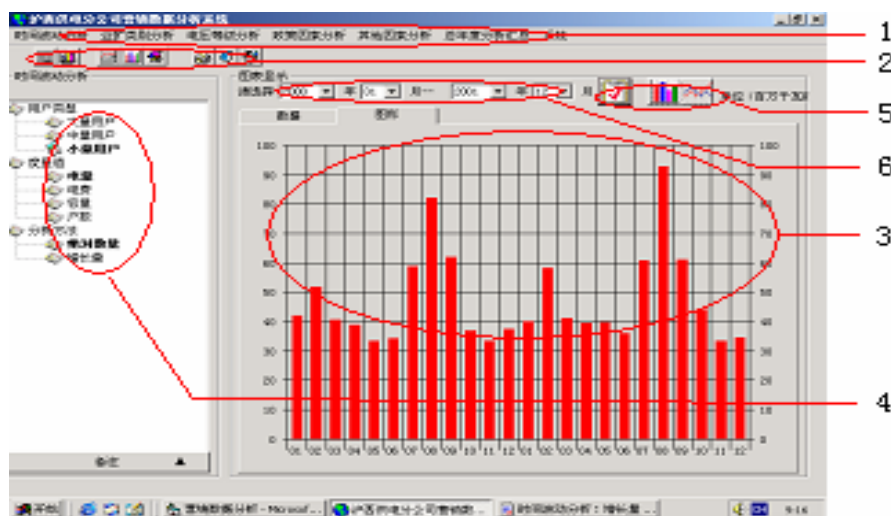


图 2：多维数据前台展示系统

界面各部分说明如下：

- 1——功能菜单：时间波动分析、业扩类别分析、电压等级分析、政策因素分析、其他因素分析（预留）、总年度汇总分析、系统；
- 2——工具栏 1：表格显示、图形显示、折线图、柱状图、饼图、打印、帮助、退出；
- 3——数据图标显示区域：两页标签，数据和图形；
- 4——查询条件选择：根据不同功能模块变更选择范围；
- 5——工具栏 2：刷新、柱状图、折线图；
- 6——时间选择：开始年、月，结束年、月，根据不同功能模块变更选择范围。

4、基于 OLAP 技术的决策对于供电企业的影响力

基于多维数据分析的决策在供电行业应用后，对企业带来如下益处：

(1)减少决策的盲目性：以往决策，可能通过市场调查或其他分析，推出不同的新业务或新的优惠政策。但在作出某项决定前是否有强有力的数据支持？该政策推出后是否对用户有吸引力？在该政策出台一段时间后，对该政策带来的用户行为是否作过评估，该政策是否有改进余地？比如供电局推出的居民分时表，用电高峰时期电费为每小时 0.61 元，而在晚上 12 点到第二天早上 6 点为每小时 0.30 元，推行分时表对用户用电量的增加有多大影响？该政策是否给供电局带来收益？设置优惠时间点是否能有改进？改进后的优惠政策能给供电局带来更多的收益？这些问题都能通过这次数据分析得到满意的回答。

(2)有效利用信息资源：供电企业拥有自己的运营支撑系统，他们包括业扩报装、电量电费、抄表核收、表计管理、用电监查、用电综合等，所有这些系统都有自己的信息，目前，这些信息都是企业进行决策的宝贵信息，但他们分散在

不同的系统中，信息之间的相关性难以进行分析，会导致决策的不全面性。把这些信息集合起来，真正做到信息向知识、知识向效益的转换。

(3)提高市场反映能力：伴随着电力企业走向市场的需要及商业化运营的要求，电力市场开始由“电力”生产为导向的卖方市场向“客户”服务为导向的买方市场转变，哪些服务会减少客户流失并增加电费收入，将是供电行业最关心的问题。及时发现用户行为与优惠政策、个性化服务之间的关系，并及时推出相应政策，是提高客户服务水平、增加营业收入的关键。基于多维数据分析的决策是达到该目的，提高市场反映能力的一条重要途径^[1]。

5、结束语

课题成果为供电企业管理层构建了多角度的绩效观察视角，利用交互式的前台展示界面实现实时分析；售电量增长构成的结论为供电企业绩效评估提供了有效依据，同时提供了企业资源更有效配置的可能性。事实表明，利用 OLAP 技术对电力营销数据进行多层次、多角度、全方位的分析，能帮助供电企业深入、全面地分析业务数据，从各种不同的角度审视业务的经营情况，揭示电力市场和营销的内在规律，科学地、快速地指出存在的问题、隐患并能发现对于企业至关重要的变化趋势，形成极具决策价值的战略信息，使企业的决策者、高层领导、部门主管乃至关键业务的技术骨干，始终能看清方向，把握大局，从容应对竞争复杂，快速多变的现实世界。

参考文献

1. Efrem G MaIIach. 决策支持与数据仓库系统[M]. 北京：电子工业出版社，2000
2. 王珊. 数据仓库技术与联机分析处理[M]. 北京：科学出版社，1998
3. Microsoft Corporation, 分析服务 [M], 北京：清华大学出版社，2000
4. (美)CIAUDE SEIDMAN , SQL SERVER 2000 数据挖掘技术指南[M], 北京：机械工业出版社，2002，1
5. [美]Erik Thomsen/George Spofford, Microsoft OLAP 解决方案[M], 北京：人民邮电出版社，2000，09
6. SQL SERVER 开发指南--OLAP(联机分析处理) [M], 北京：电子工业出版社，2001，11

作者简介：

- 1、徐卫华，女，同济大学管理信息系统专业硕士研究生，江苏省海门市人，1979年10月生。主要研究方向：数据仓库和数据分析。