

# 基于 OLAP 技术的污染源在线监测数据分析系统\*

朱亚兴

(广东科学技术职业学院, 广东 珠海 519090)

**摘要:** 污染源在线监测数据分析系统基于污染源管理综合系统数据库提供的数据库, 经过数据清洗、转换, 加载到数据仓库中, 进行主题设计与多维建模, 采用 OLAP 技术实现海量数据展示、分析统计, 同时提供功能强大的共享查询、多维分析和污染源企业排口监测因子预警、事件自动提醒及各类数据报表统计等, 提高了减排工作信息化的水平和工作效率。

**关键词:** 联机分析处理; 数据仓库; 多维分析

中图分类号: TP393

文献标识码: A

文章编号: 1674-7720(2012)06-0004-03

## OLAP-based on-line monitoring and data analysis system of pollutant source

Zhu Yaxing

(Guangdong Vocational Institute of Science and Technology, Zhuhai 519090, China)

**Abstract:** The datasource of on-line monitoring and data analysis system of pollutant source is based on integrated pollutant management system database. After data cleaning, transformation, the data is loaded into the data warehouse. Thematic design and multi-dimensional modeling, mass data show using OLAP technology, statistical analysis, a powerful shared inquiry, multi-dimensional analysis and the notice and alarm of monitoring emission factor from enterprise pollutant source, automatic event reminders, various data statistical reports are provided. It also improves the information level of emission reduction and working efficiency.

**Key words:** OLAP; DW; multidimensional analysis

根据国家环保总局《关于制定主要污染物总量减排年度计划的通知》(环办[2007]33号)要求,及广东省人民政府《关于加强我省主要污染物总量控制工作的实施意见》(粤府办[2007]118号),对已建成在线监控设备的企业提出了考核要求。为加强减排工作信息化,提高工作效率,强化数据处理能力,完善在线监测数据库,同时结合“减排”工作的需要,需建立在线监控数据上报及联机数据分析系统。

### 1 联机分析系统研究

联机分析处理 OLAP(On Line Analytical Processing)的概念最早是由关系数据库之父 CODD E F 于 1993 年提出的。当时, CODD 认为联机事务处理 OLTP (On Line Transactional Processing) 已不能满足终端用户对数据库查询分析的需要,用户的决策分析需要对关系数据库进行大量计算才能得到结果,因此提出了多维数据库和多维分析的概念,即 OLAP<sup>[1-3]</sup>。OLAP 是数据仓库系统的主要应用,支持复杂的分析操作,为人们提供直观易懂的

查询结果<sup>[1]</sup>。

OLAP 服务器一般作为数据仓库系统的中间层和用户预定义的多维数据视图,对数据仓库的信息进行统计分析处理,为具有明确分析范围和分析要求的用户提供高性能的决策支持<sup>[2]</sup>。OLAP 技术通过基本多维分析操作为用户提供一幅幅直观的视图。常用的多维分析操作有钻取、切片、切块和旋转等。

### 2 系统需求及功能概述

污染源在线监测数据分析系统基于污染源管理综合系统数据库提供的数据库,要求将系统监控数据收集到分析数据库中,实时分析监测数据,即时发现数据异动发出报警,实现数据分析、发布上报和报警,同时提供功能强大的即席查询和统计分析能力,提高工作效率。

系统主要功能模块划分为:实时监测和异常报警、事件提醒及待办处理、数据分析、设备维护、数据同步和系统管理。

实时监测和异常报警功能模块包括对数百家重点企业进行实时排污监控,主要是排水和排气。每家排污企业根据其行业类别、生产工艺等不同设立了不同监测

《微型机与应用》2012年 第31卷 第6期

\* 基金项目:广东省自然科学基金项目(8151064007000004),珠海市科技计划项目(PC20082010)

软件天地 Software Technology

因子和排放范围,当企业的排放超出了设定的排放标准或产生其他异动状况,系统及时发出警报,并向值班人员发送短信提醒;发现设备运行异常时通知运营单位进行检修,填写设备异常报告等。

事件提醒及待办处理功能模块包括:当系统侦测发生异常和设备维护等事件时,系统自动提醒用户;由用户输入相应事件处理的日程安排。

数据统计分析功能包括对污染源企业、监测因子排放、现场巡视、企业生产异常事件、污染源排放、超标排放、掉电情况、超标率等统计分析等。

数据同步功能分为业务数据更新后同步和数据补登后数据同步。

系统管理功能模块包括用户管理、角色管理、权限分配和系统参数设置等功能。只能由系统管理员操作使用。

3 在线监测与分析系统数据仓库设计

3.1 数据仓库设计框架

数据仓库的设计为决策、分析及高效管理提供保障和服务,是系统在线监测与分析的基础与核心。

本系统的数据仓库设计总体框架如图1所示,按照数据获取层、数据存储层、数据展现层模式建设。其好处是层次分明,层次之间松耦合,保证整个系统的稳定运行,并方便维护和功能的扩展工作。

数据获取层 从污染源管理综合数据库中获取的数据作为数据源,经过必要的抽取、清洗、转换等处理后加载到数据仓库系统中,即 ETL (Extraction Transformation Loading)过程。本系统关于数据抽取和转换使用基于商业智能的开源工具 Kettle ETL 对多种不同数据源、不同格式的数据进行清洗、转换、加载等繁杂工作。

数据存储层 包括数据仓库和从属的数据集市。数据仓库是信息的逻辑和物理的存储。在数据仓库实施过

程中,对于某些主题业务的分析问题,按照主题采用数据集市的方式对数据进行进一步的组织。在中央数据仓库的基础上根据分析需求创建相应的从属数据集市。

数据展现层 通过对数据存储层的数据进行各种加工整理,将获得的数据以文字、报表、曲线和各种图形的方式,简便、快捷地展现出来。主要包括即席查询、联机分析处理、报表输出和信息结果发布等功能。

3.2 主题分析设计与多维建模

数据仓库中的数据都围绕相关主题组织,面向主题是在线监测与分析系统数据仓库区别于污染源管理综合数据库系统的主要特征。经过需求分析和与用户反复探讨最终确定了在线数据分析系统数据仓库的几个基本主题:污染源类型主题、异常事件类型主题、监控设备类型主题,并确定其公用类主题。如污染源类型主题包括所属行业、监测站类别、工艺流程、监测因子、企业规模、注册类型、管理级别、排放去向、污染物处置方法等信息。异常事件类型主题包括异常类型、监控类型、超标因子等信息。监控设备类包括设备类型、调试目的等信息。

公用类主题类型包括的维度有日期、时间和地区。如图2所示。

日期	定义:分析时间范围
日期	定义:分析时间范围
单位:年、季、月、日	
层次:年->季->月->日	
时间	定义:分析时间范围
单位:时、半小时、刻、五分、分、秒	
层次:时->半小时->刻->五分->分->秒	
地区	定义:分析所选择的地区
层次:地区->区县	

图2 公用类主题类型维度

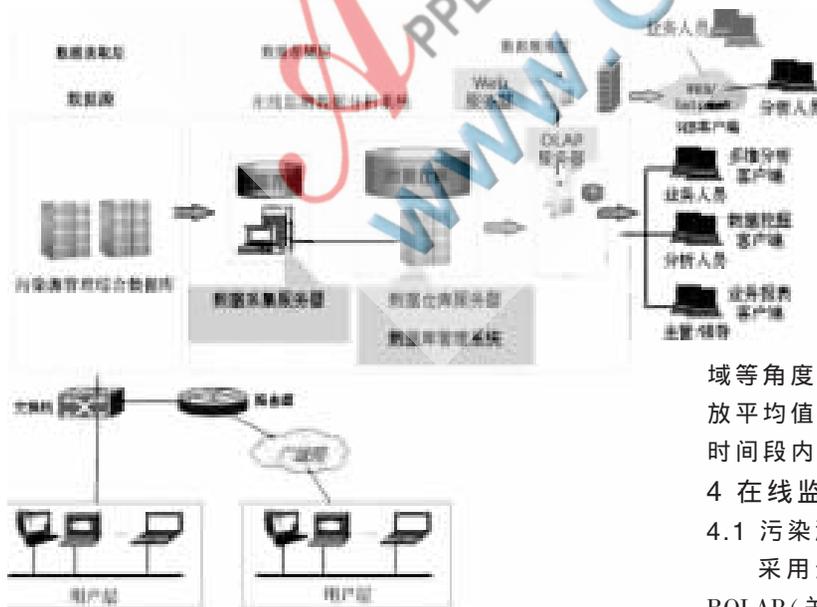


图1 数据仓库设计总体框架图

确定主题后,可以用多维模型进行描述。多维建模方式表现了属性间更为丰富的关系和灵活性。多维建模易于使用者进行多维分析<sup>[5-6]</sup>。在多维分析模型建设中主要存在两种模型:星型模型(Star Schema)和雪花模型(Snowflake Schema)。以污染源排污监控分析主题关系逻辑雪花型多维模型如图3所示。

此模型对应的系统分析目标和功能为:从时间段、污染类型、企业规模、行政区域等角度对排放因子进行分析,掌握各类排放因子的排放平均值、最大值、最小值和累计排放值,并可进行不同时间段内趋势和比较分析,了解排放值的变化趋势。

4 在线监测数据分析

4.1 污染源排放统计分析

采用企业级商务智能分析处理模块,引入主流的ROLAP(关系型联机分析处理系统)技术,通过 Mondrian



- [5] 黄晓颖,李亚芬,王普. 基于数据仓库的学科建设决策支持系统的设计[J]. 计算机工程与设计, 2010,31(23): 4995-4998.
- [6] 秦永平,王丽萍,孙庆. 基于数据仓库的突发公共卫生事件预警预报系统[J]. 计算机工程与设计, 2010,31(23): 3119-3122.
- [7] 王亚妮,李茜. 基于 OLAP 技术的社区卫生服务体系信息管理平台的设计与实现[C]. 2010 3rd International Conference on Power Electronics and Intelligent Transporta-

tion System,2010.

- [8] 陈元中. 基于聚类的 OLAP 多维分析查询推荐方法研究[J]. 计算机工程与设计, 2010,31(15): 3503-3505.  
(收稿日期:2011-12-28)

作者简介:

朱亚兴,女,1972年生,硕士、高级工程师,主要研究方向:企业信息化与大型数据库建模。

