

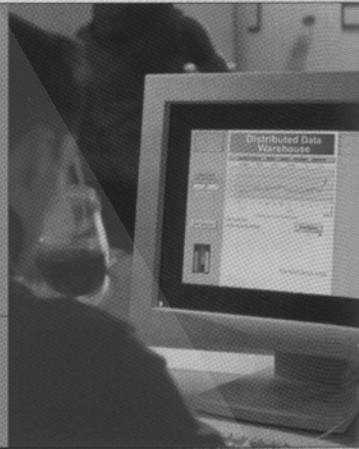
# 基于分布式数据仓库的政府采购统计分析设计

## The Statistics and Analysis of Government Procurement Based on Distributed Data Warehouse

**摘要:** 本文分析了政府采购统计的重要性,指出目前统计方法的不足。提出按照政府采购中心的层次结构构建一个分布式数据仓库统计分析系统,每个数据集市都支持独立的决策分析应用。给出数据仓库和数据集市的逻辑设计,以及用OLAP技术对政府采购信息进行多维统计分析的方法。

**关键词:** 政府采购统计; 分布式数据仓库; OLAP

刘 畅 (大连 东北财经大学 经济信息系 116023)



### 1 引言

随着我国政府采购的深入发展,采购范围和规模不断扩大,采购处理形式也从传统的手工处理逐渐转向基于信息技术和电子商务的电子采购。各地采购中心利用采购管理系统进行信息管理,在采购网站发布招投标信息,建立电子交易平台进行网上询价和竞标。

政府采购积累了很多重要的数据资源,对这些数据按采购预算、合同金额、节支、数量、品种、采购方式、原产地、地区分布、供应商等指标进行统计分析,可以全面了解财政支出的使用和有关政府政策目标的落实情况,加强政府采购基础管理,为政府部门经济政策的制定、修订和政府采购制度的完善提供决策依据。政府采购信息统计分析是政府采购的重要组成部分,是政府采购工作成果的综合反映。

目前各地政府采购的统计分析相对滞后,普遍的做法是各市、县区的政府采购中心利用信息管理软件进行采购预算与合同登

记,按月汇总统计采购数据、填写报表并上报到省级政府采购中心,各省每年对全省范围内的政府采购情况进行汇总统计和各项指标计算,将统计报表和分析报告上报到中央财政。这种做法导致各基层单位的统计分析工作量大,统计时间长、统计口径不易统一、重复工作多等问题,对各地上报数据进行汇总综合也很繁琐,往往上一年的数据到第二年的5、6月份才能汇总完毕。另外,政府采购信息系统是以数据库为中心的联机事务处理系统(OLTP),其设计是面向数据更新比较频繁的应用,如果在系统中进行统计分析,会降低系统的性能,数据的组织也不适合进行多维分析。

### 2 基于分布式数据仓库的政府采购统计分析

#### 2.1 数据仓库、数据集市和分布式数据仓库

数据仓库是一个面向主题、集成、稳

定的且随时间变化的数据集合,用于支持管理人员的决策。数据仓库需要对来自多个数据源的数据进行转换、清洗,确定装载策略。

数据集市是某一个部门所使用的小容量数据仓库,相对于数据仓库而言,具有信息少、费用低廉、实施时间短的特点。各部门可以在数据集市上建立自身的决策支持系统。

可以使用分布式数据仓库策略将数据集市集成到整个应用系统中。分布式数据仓库由分散在各地的自主数据集市和中央全局数据仓库组成,中心组织制定元数据的规则,分布式数据仓库对用户的查询是透明的。

#### 2.2 基于分布式数据仓库的政府采购分析系统的设计

政府在采购中,采购数据存储在各个基层单位的采购中心,各省、自治州需要对全省范围内的采购数据进行分析,中央财政需要对全国范围内的采购数据进行分析。因此

可以构建一个主从式的分布式数据仓库, 各省、自治州将分散在本省各市区采购中心的数据集成到本省数据集中, 中央财政建立中心数据仓库, 其数据来源是其他省份数据集中的汇总数据, 中心数据仓库可以直接访问其他数据集市。各个数据集市可以通过采购专网与中心数据仓库连接。整个统计分析系统用基于Web的三层结构实现, 数据层为中心数据仓库和各个数据集市, 可以采用ORACLE 9i作为数据服务器。业务层用组件实现OLAP功能, 用户层用WEB页面实现可视化的统计分析与挖掘。

由于各地区的政府采购信息管理系统基本统一, 因此可采用统一的关系模式设计各省的数据集市, 以减少系统维护代价。在中心数据仓库存放一些公共的基础数据, 如供应商目录、商品目录等, 实现信息共享。

在每个数据集市上建立独立的统计分析子系统, 通过权限设置, 允许本省各市区采购中心访问数据集中与自己相关的数据并进行统计分析。政府采购分析系统结构如图1所示。

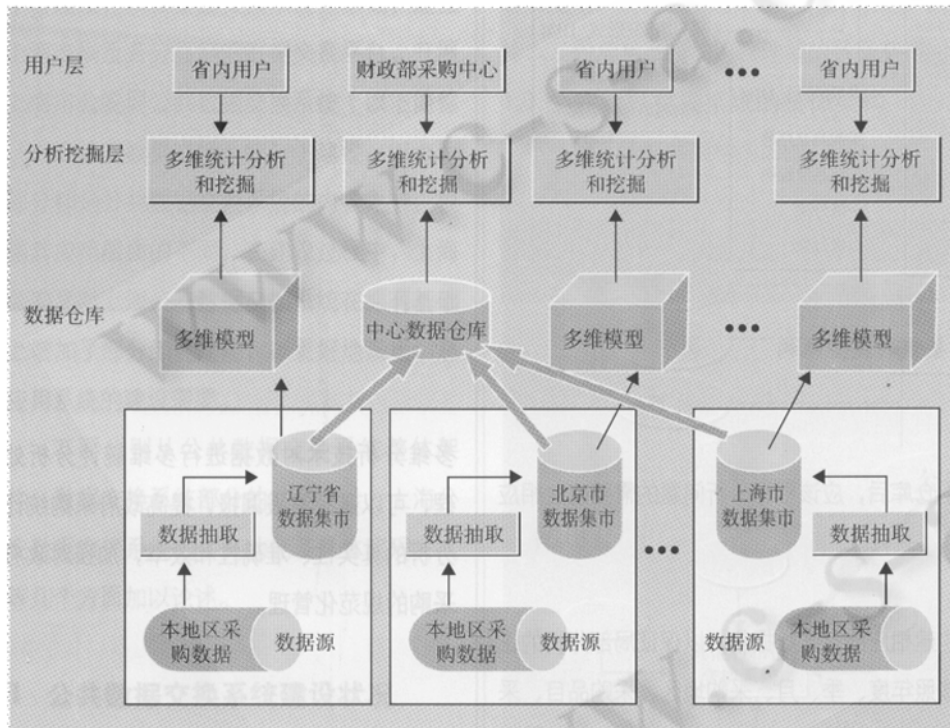


图1 政府采购分布式数据仓库统计分析系统体系结构

### 3 数据集市与数据仓库的逻辑设计

#### 3.1 主题的确立

政务采购统计涉及到供应商、采购品目、采购单位、采购合同等。从要求上报的统计报表来看, 政务采购统计主要对各地区政府采购项目的预算、合同金额及效益进行分析, 研究采购的效率和节支情况, 因此可以选择采购合同作为数据仓库和数据集市的主题, 使各级政府采购中心对采购预算的执行、项目个数、预算资金、减少财政支出金额、项目市场价效益比、政府采购预算的执行进度有全面的了解。

#### 3.2 事实表和维度表设计

数据集市存放各地的采购数据, 而中心数据仓库存放元数据和基础数据。在数据集中, 采购事实表是采购合同表, 维度表有供应商表、采购品目表、采购形式、采购组织方式、时

间、采购地区、采购部门七个。其中采购事实表的事实数据有预算金额、预算内资金、预算外资金、自筹资金、采购合同金额、采购预算内金额、采购预算外金额、采购自筹资金、实际支付价、采购项目数、采购合同个数。采购方式有公开招标、邀请招标、竞争性谈判、询价和单一来源五种, 采购组织方式有集中采购、国管局采购、部门统一采购和单位分散采购四种。采购部门分为科教文、公检法、卫生、农林水及其他五类。采购地区分为中央、地方, 地方又分为省、市、县、区。采购品目分为货物、工程、服务三大类, 每类又有一级目录、二级目录等。维度表的设计采用合并维分层结构, 事实表和维度表通过外键连接。政府采购数据集市的结构如图2所示。

#### 3.3 粒度选择

数据集中的事实数据除了记录本省每个地区发生的每笔采购合同细节外, 为了提高数据仓库的查询和分析性能, 还应该存放按维度进行轻度汇总的综合统计表。如按月、年建立本省采购统计表如下:

综合统计表(序号, 年, 月, 采购方式序号, 采购组织形式序号, 采购品目序号, 采购部门序号, 预算金额、预算内资金、预算外资金、自筹资金、采购合同金额、采购预算内金额、采购预算外金额、采购自筹资金、实际支付价、采购项目数、采购合同个数, 节约资金, 预算资金节支率)。

其中: 节约资金=预算金额-实际支付价

预算资金节支率=预算资金节支额/项目预算中预算资金\*100%。

中心数据仓库可以直接使用这样的统计表。

#### 3.4 抽取策略

数据集中的数据来源于各基层政府采购中心的政府采购管理系统, 由于不同省份的采购管理系统可能运行在不同的平台上, 因此要分别设计数据抽取和转换的策略。对往年的历史数据编写专门的转换程序载入。

而对于日常运行产生的业务数据，如果使用的数据库是ORACLE，则可以利用ORACLE的数据转换程序进行数据抽取；对其他类型的数据库，还需编写转换程序。

可以按周或月为单位抽取新产生的政府采购业务数据，所抽取的数据通过专门的采购网传递到数据集市。

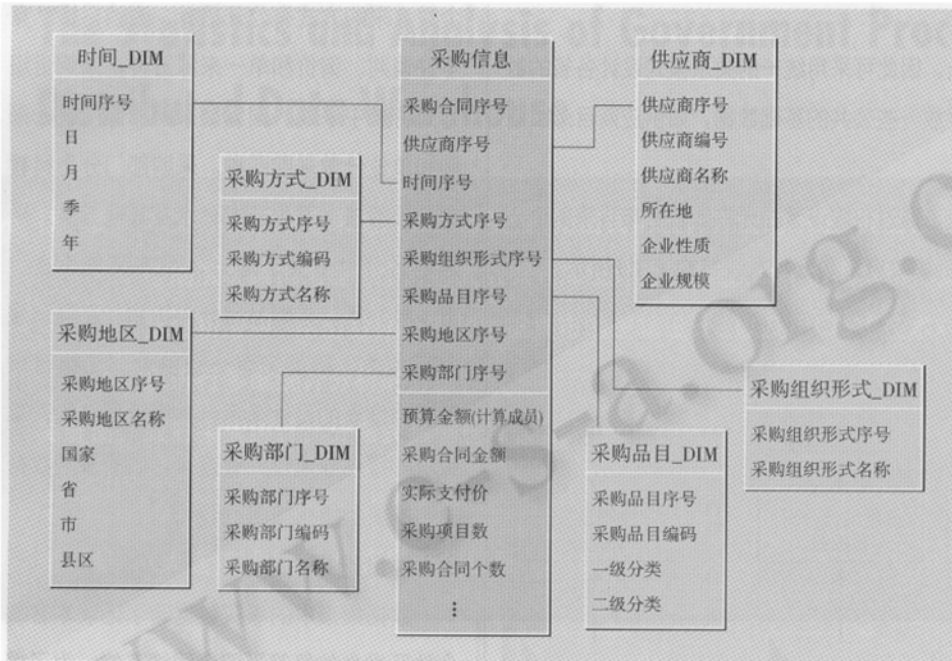


图2 政府采购数据集市的星型结构

#### 4 用OLAP进行政府采购的统计分析

数据仓库是进行决策分析的基础。在建立数据仓库后，应该根据分析问题的需要建立相应的多维数据模型，以便进行分析查询。

##### 4.1 创建多维数据模型

对于分布式数据仓库的每个数据集市，应该构建相应的多维数据集，以保证局部分析的正确执行。多维数据集也采用星型结构，在构建时按照年度、季、月、采购地区、采购品目、采购规模、采购方式、采购部门等建立共享维，度量数据包括预算金额、合同金额、项目数等，还应包括计算成员，如增加节约资金、预算资金节支率等度量。

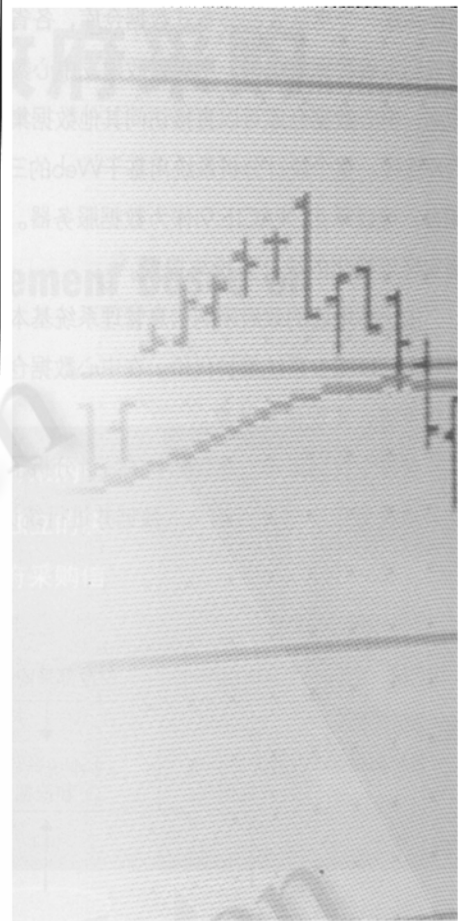
##### 4.2 用组件实现多维分析功能

在应用服务器层，可以采用ORACLE 9i提供的OLAP API接口访问数据仓库，OLAP API通过多维元数据模型MDM访问数据仓库中的多维数据，如事实度量、表示数据结构的维度。ORACLE 9i还提供了BI Beans组件，包括数据展现组件和获取操作数据的组件。针对政府采购统计的特点，在应用服务器层要建立相应的组件，用多维术语访问、操作、统计存储在数据仓库中的数据。

在用户端，可以采用基于Web浏览器的B/S模式，这种方式更适合非局域网环境。

#### 5 结束语

政府采购信息统计是政府采购工作的重要内容之一，也是制定有关经济政策、完善政府采购制度的重要参考和基本依据。按照政府采购中心的分级管理模式建立分布式数据仓库，利用



多维分析技术对数据进行多维统计分析处理，可以避免重报漏报，提高政府采购统计分析的真实性、准确性和效率，加强对政府采购的规范化管理。

#### 参考文献

- 1 Inmon W.H. Building the Data Warehouse [M], 王志海等译, 机械工业出版社, 1996.
- 2 王珊等, 数据仓库与联机分析处理[M], 科学出版社, 1999.
- 3 中国政府采购统计. <http://www.ccgp.gov.cn/web/tongji/tjzk.htm>. [EB/OL].