

# 在线数据分析系统 OLAP 在企业决策中的应用

杨晓波 (杭州浙江财经学院信息学院 310012)

**摘要:** 本文介绍了在线数据分析系统 OLAP 的体系结构, 系统实现原理和应用情况, 该系统采用分布式的三层体系结构, 接口之间遵循 CORBA 规范。实践表明, 该系统可以对业务数据进行很好的分析、展现, 是企业经营决策的有力工具。

**关键词:** 在线数据分析 体系结构 数据展现

## 1 引言

在线数据分析系统 OLAP 是基于数据仓库技术的分析决策工具, 系统通过分析企业的业务数据, 能方便地对各层面的信息进行统计分析, 提供灵活的图、表连动的方式直观地反映变化情况, 并通过数据向上钻取或向下钻取的方式分析客户业务收入以及业务增长或下降的原因, 为制定客户营销服务策略提供参考, 并可以预测未来的业务情况。

## 2 在线数据分析系统 OLAP 的体系结构

系统采用分布式的三层体系结构, 由

Database Server、Application Server 和 Explorer、Admin 来共同完成系统的功能。所有业务逻辑操作和数据库操作都是由 Application Server 来完成的, 客户端和服务器端之间的接口遵循 CORBA 规范, 从而具有良好的扩展性和稳固性, 其结构如图 1 所示。

从图 1 可以看出, App Server 是连接数据库和客户端的桥梁, 它是在线分析系统的核心部分, 集成了所有的业务逻辑和数据处理模块。Explorer 是供分析人员使用的简单的查询、分析、报表工具。Admin 是供系统管理员使用的工具, 用于配置查询主题。

以体系结构图为依据, 并将各部分进一步细化, 建立与之相关的软件体系, 具体结构如图 2 所示。

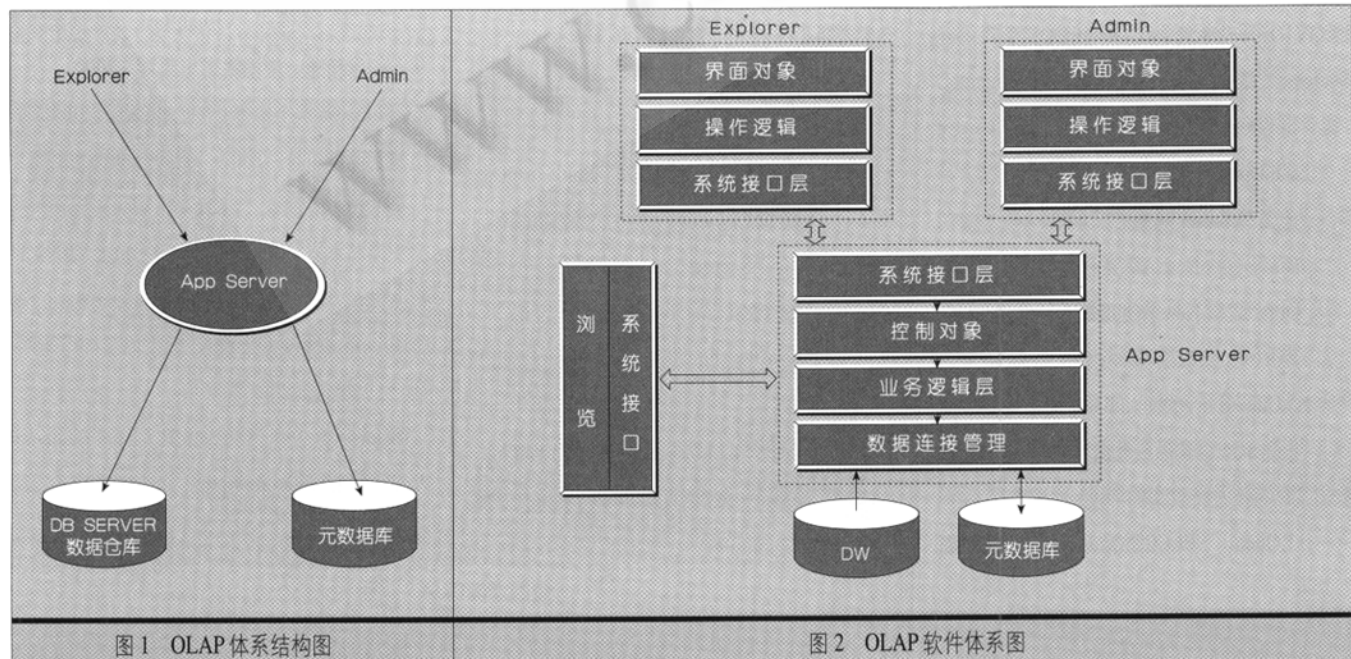
应用服务器(App Server)负责响应前台的请求, 应在 Admin 和 Explorer 运行前启动。

应用服务器启动后首先将元数据从元数据文件载入到元数据缓冲, 然后等待响应前台的请求; 当前台有请求时, CORBA 接口部分收到请求, 交给接口装卸部分, 由接口装卸部分识别前台的不同请求, 调用相应对象的方法完成响应。

配置管理器(Admin)负责大部分元数据的配置和管理工作, 包括数据源管理、维度管理、查询主题管理、权限管理。

配置管理器启动后首先请求登录到应用服务器, 然后向应用服务器请求获取元数据包并将元数据包导入本地元数据缓冲; 以后用户对元数据的操作都保存在本地元数据缓冲。当操作中有对业务数据库的访问要求时, 向应用服务器提出相应请求。当需要将元数据保存到后台时, 配置管理器向应用服务器提出保存元数据包的请求。最后用户退出程序时, 配置管理器向应用服务器请求注销。

前台展现管理(Explorer)负责将用户查询, 作为分析工程发出请求, 应用服务器从元数据库



中找到相应的配置结构,再从业务数据库中提取数据,打包发回 Explorer,进行前台展现。

### 3 系统实现原理

动态数据统计分析的数据模型采用星形结构,即所有数据信息由维表和事实数据表组成,其中维表存放各个维的可能值,事实数据表存储各种条件下的事实数据,这种模型设定整个数据仓库是一个多维的信息空间,空间中的每点都包含一定的数据信息,需要由维表中的信息来决定信息点,而由事实数据表来决定该点的数据信息是什么,如图3所示。

显示所有的维都是相互独立的,不会与其他维发生任何关系,而只与事实数据发生关系,而且必须与事实数据发生关系才能根据该维来取出数据,事实数据表与所有维表都存在一个主外键的关系,事实数据的确定就是根据各个维的键值到事实数据表中找到相应的事实数据,如图4所示,这种结构以事实数据表为中心组成了一个星形的网,故称为星形结构,显然要查找事实数据需要确定所有 Key 值(键值),而这些键值的确定就需要根据各种维度信息从维度表中来获取,如何获取各个维的键值是星形模型中重要也是最复杂的一部分内容。

一个查询主题(Query Subject)只拥有一个事实信息表,该事实信息表包含该查询主题查询空间中所有点的数据信息,事实信息表实际上由两部分组成,一部分是维信息,用来决定该信息点在查询空间中的位置,如上图中的 Key1-Key4;另一部分就是数据信息,即量(Measure),用来存储该点的数据信息,如上图中的 Fact1-Fact3。

一个查询主题可以拥有多个维度信息表,这些维度共同构成了该查询主题的查询空间,实际上每一个维度本身的信息又由维的层(Level)和

维的值(Value)组成了一个平面,如图5所示:

维中的信息可以根据各种原则分为不同的层次,如最常见的时间维可以分为年、半年、季度,月这么几个层次,每个层次又都有各自的值点,如年这一层可以有(1999、1998、1997、1996、1995...,1960)等值,在半年这一层上可以有(1999上半年、1999下半年、1998上半年、1998下半年,....,

1960上半年、1960下半年)等值,依次类推,每个维值在不同的层面上都可以找到对应的值点。

### 4 在线数据分析系统 OLAP 实际应用

该系统核心模块采用C++语言开发,数据库采用 Oracle 数据库,以电话话费为例,介绍该系统的应用情况,应用服务器(App Server)首

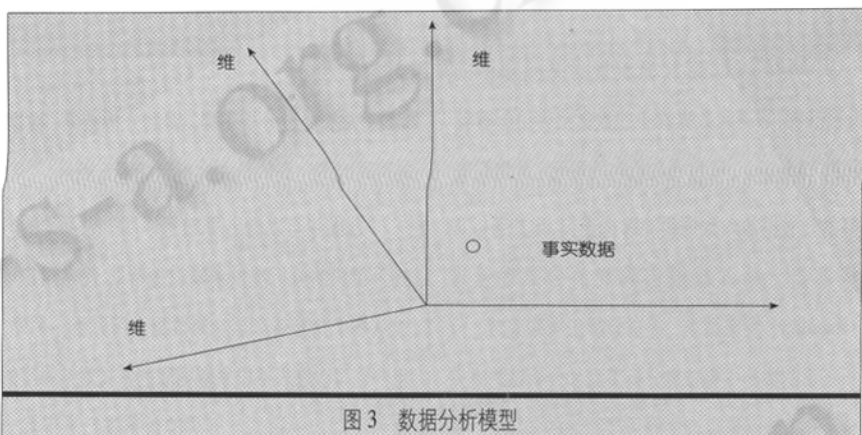


图3 数据分析模型

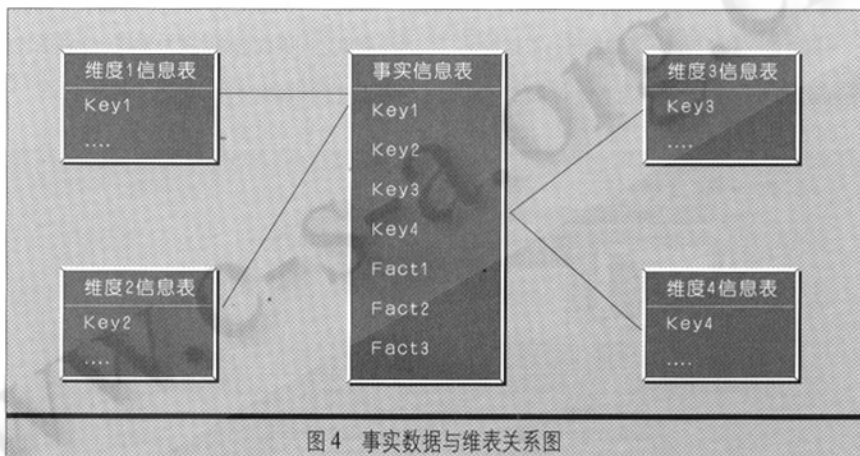


图4 事实数据与维表关系图

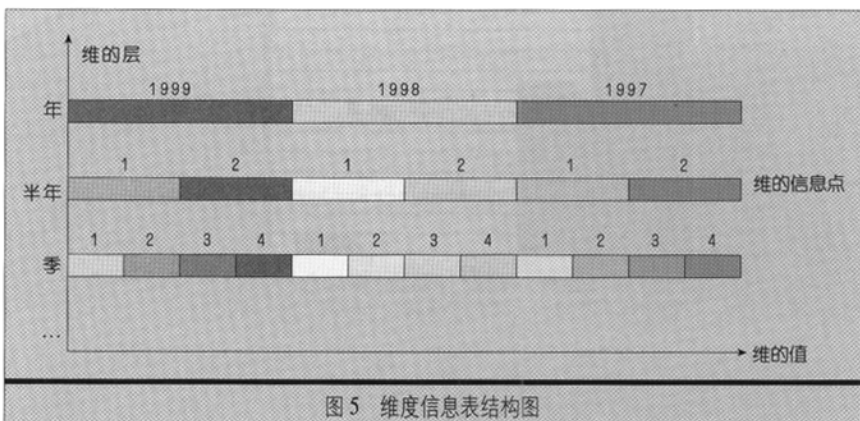


图5 维度信息表结构图



先被触发,它将元数据库中的数据上传给配置管理器(Admin),元数据在此进行配置,并将配置好的元数据放入数据缓冲等待查询。用户可以根据需要,输入权限密码向前台展现管理(Explorer)发出请求,提交分析工程来完成。分析工程的实施主要依靠分析节点和分析链路来完成,分析节点是查询主题所要查询的知识点,由分析工程派生出分析链路,依据分析链路将查询所得数据送入数据缓冲区,最终进行数据展现,所选数据为2001年上半年业务收入情况,数据

配置好后,经过分析工程,得到分析结果,数据展现可以图、表的形式输出,如图6所示。

从图中可以看出,系统可以图、表互动的方  
式为用户的分析提供清晰快捷帮助,使得用户可以从各个角度各个层次对数据进行分析,另外,用户也可以在表格中采取掘取的方式对某维度的  
维值所对应的上层或下层的数据进行展现和分  
析。对于数据的相关性分析,可根据同一维度中  
两个维值之间的数据对比关系形成一个线图,以  
分析两者之间的相关性。

## 5 结束语

在线数据分析系统 OLAP 可以挖掘出业务数据中潜在的大量有用的信息,从而为企业的经营决策提供参考。数据可以进行动态查询,分析,并且方便、快捷的将查询分析结果展现给用户,如将这一系统加以很好利用,并融合数据挖掘技术,可使企业在激烈的市场竞争面前立于不败之地,有着很好的应用和发展前景。■

## 参考文献

- 1 B. Mobasher, N. Jain, E. Han, and J. Srivastava. Web mining: Pattern discovery from world wide web transactions. Technical Report TR 96-050, University of Minnesota, Dept. of Computer Science, Minneapolis, 1996.
- 2 O. R. Zaiane and J. Han. Resource and knowledge discovery in global information systems: A preliminary design and experiment. In Proc. of the First Int'l Conference on Knowledge Discovery and Data Mining, pages 331--336, Montreal, Quebec, 1995.
- 3 李德毅.《数据开采和知识发现研究的回顾与展望》. 产品与技术, 2000.01.03 第1期.

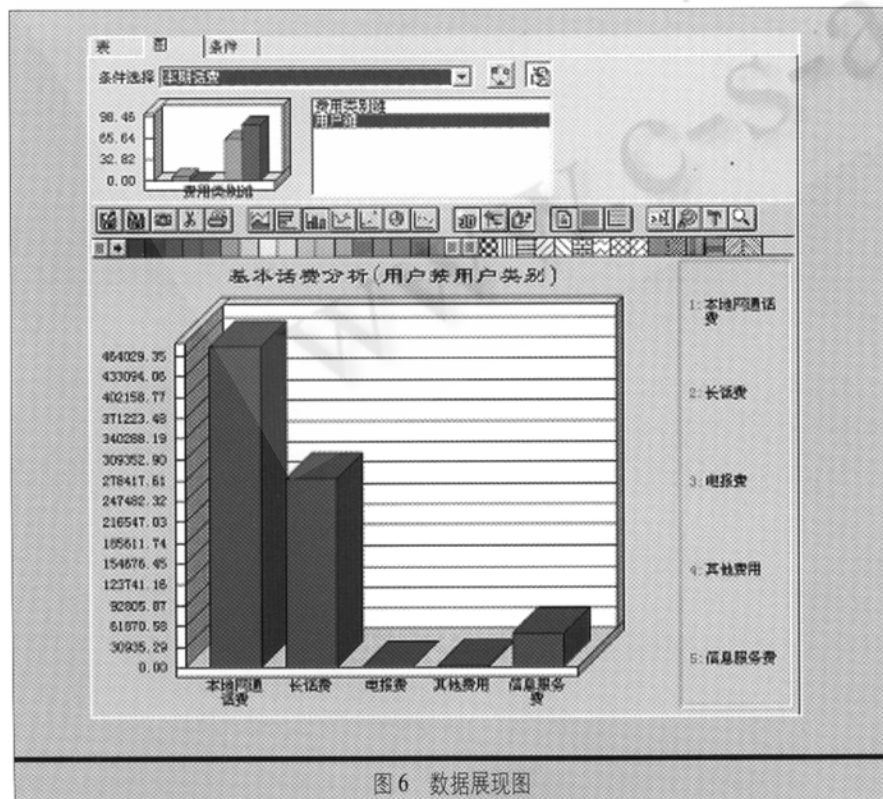


图6 数据展现图