

联机分析处理技术(OLAP)与决策系统 开发工具 Light Ship

D&B 公司

早期决策支持系统面临着查询速度慢,不能满足随时产生的灵活查询需求,系统维护困难、效率低、柔性及可操作性差等一系列问题。而 OLAP(On-Line Analytical Processing)联机分析处理技术的出现导致了 DSS 系统的新局面,改变了传统企业决策速度慢、灵活性并、欠全面的决策过程,提供了灵活、快速、实时的联机报告分析,满足了不同阶层、不同重点、不同专业的需要,在机会出现时及时把握,共享决策成果。

Pilot 是一家从事 DSS 开发已有十多年历史的美国公司,是 OLAP 技术开发的先驱之一。其中客户/服务器体系下可视化决策系统开发工具 LightShip,是具有 OLAP 技术的典型产品。在商务智能工具(Business Intelligence Tools)领域里具有领先地位。该产品连续三年被美国 PC WORLD 评为 EIS 开发工具最佳产品。

一、什么是 OLAP

OLAP 描述的是一种多维数据服务,这种服务设计目的是保证分析员、经理和决策者针对特定问题,通过快速、一致、交互式的实时数据访问和分析。获得有创意的发现。OLAP 技术的基本特征是综合用户数据进行动态时实多维分析,提供给用户快速一致的查询响应速度,而不论数据库大小和数据结构复杂性。由于 OLAP 趋向采用“面向目标”的方式,揭开层层数据,直接提供针对管理问题的解决方案,因而 OLAP 更接近人类智能思维。OLAP 具有下列特性:

1.OLAP 是发展中的技术

由于 EIS(Executive Information Systems)和 DSS(Decision support Systems)的发展以及 Client/Server 和 Groupware 技术的涌现,因此 OLAP 也在不断的发展。

2.OLAP 是交互式的分析方法

OLAP 是由许多假设性的“what-if”和“why”数据模型组成的。是一种交互式分析方法。OLAP 应用采用“面向目标”方式,揭开层层数据,信息与分析并重,直接提供针对管理问题的答案,具有很高的实用性和可操作性,对于商业用户分析数据提供了以下功能:

EIS

DSS



Client/Server

Groupware

- 数据的切割功能(Slice and dice) 可以按照拥护需求提供给用户所需查询的那部分数据。
- 透视功能 按照数据层次管理结构从上层钻透到下层去取数据,以满足用户的查询需求。
- 寻觅功能 可按照用户特殊的需求去查找数据。
- 回翻功能 可追溯用户查找数据的全过程。

3.OLAP 是高可视化的

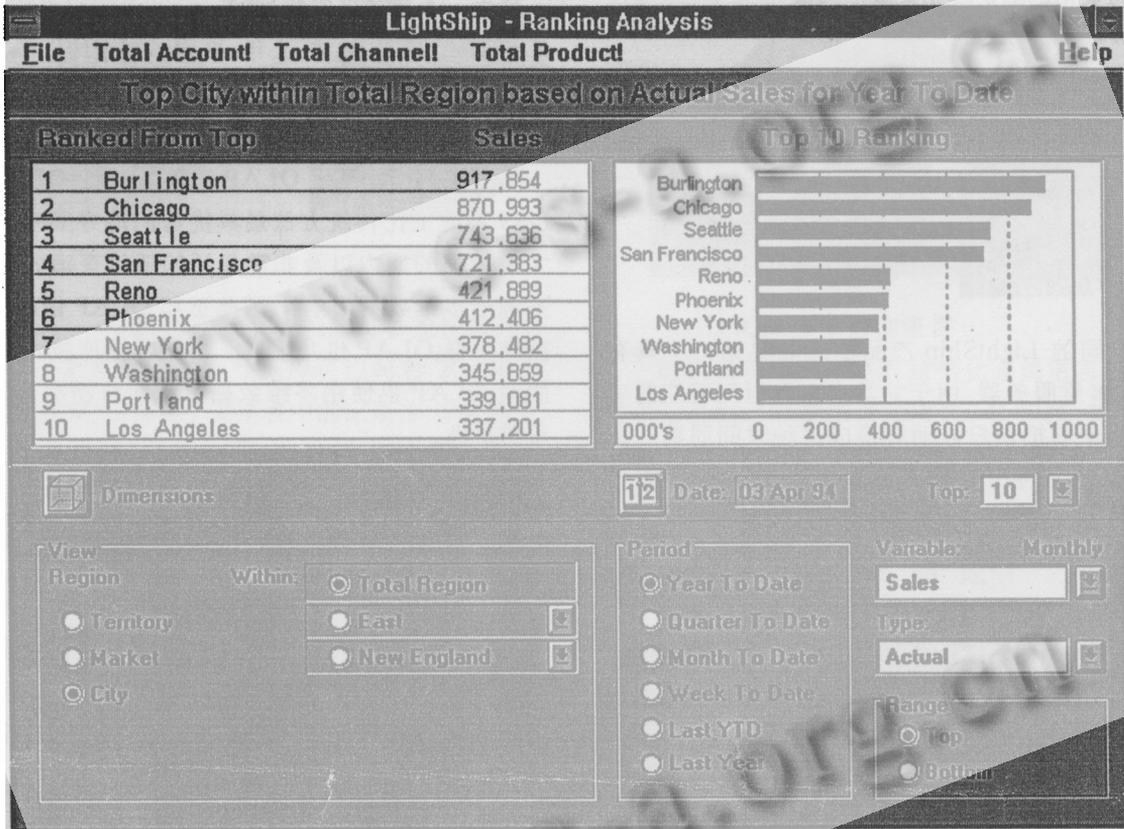
OLAP 是提供给商业用户去分析合成用户原有数据库中的数据,变为被管理者决策有用的信息,提供有创意的发现和和问题出现之前发现问题,掌握趋势,迅速做出反映。因此 OLAP 提供给最终用户去浏览和分析数据的界面必须是高可视化的。

4.OLAP 是多维的

- OLAP 使用多维数据库技术来提高查询速度。

由于 OLAP 要访问大量数据,并支持数据的不断累积,对大量的数据提供综合分析,同时还必须提供给用户快速一致的响应时间,因此 OLAP 一定要具有强大的综合数据能力。而关系数据库在表示数据间的多维关系上

,以及数据在磁盘上的物理存储方法和数据定位用的索引方法都与多维数据库有很大的不同,因此在综合数据的能力上多维数据库具有千倍于关系数据库的能力。同时,LOAP 提供给用户的分析界面也必须是多维视图,以使用户从多个角度全面分析了解企业状况。



• OLAP 使用多维数据库技术满足灵活查询需要。

OLAP 使用多维技术面向业务问题,多层次、多对多组织企业数据,使用多维数据库中的多维旋转功能,达到任意满足事先未定义的随机查询和分析。

5.OLAP 查询速度非常非常快

OLAP 采用多维技术提高查询速度。

OLAP 由于采用多维数据库技术,而多维数据库本身综合数据能力比关系数据库快很多,同时采用矩阵存储数据,并且对稀疏数据采取压缩等一系列优化算法,极

大地提高了查询速度。多维数据库在不同平面之间的运算采用向量运算,一步操作完成平面间的运算。因此,OLAP 有非常非常快的响应速度。

6.OLAP 动态数据分析

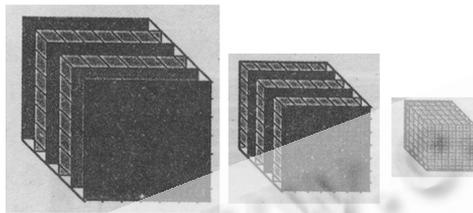
有效的数据分析处理需要使用不同的数据分析工具和数据模型。没有较多操作历史数据的数据分析被认为是静态数据分析。对历史数据需要进行广泛操作的数据分析被认为是动态分析。动态数据分析涉及到多维数组,并比在传统的 OLTP 环境下的分析更富有挑战性。通过多维数据库多次与分析主体打交道并更新之,动态数

据分析能够理解一个商业企业内部所发生的变化,可以用来对那些没有发现的特定的商业挑战指明选择方案,可以用来对未来战术方案的确定。

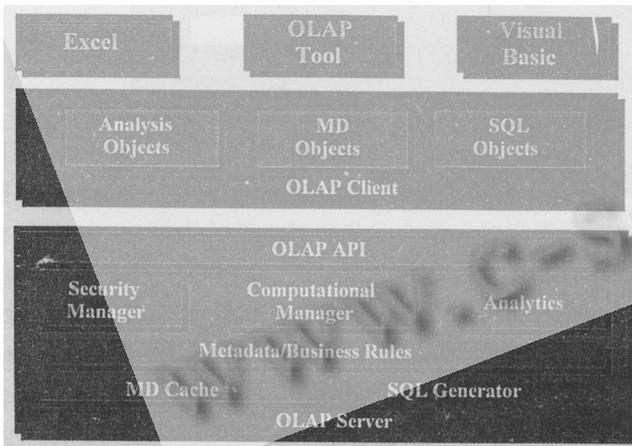


魔方效应...

任意灵活的满足事先未定义的随机查询和分析



Pilot 公司的 LightShip 产品是工业界第一个具有时间智能的多维服务器,由于它具有内置时间智能的功能,可以很方便的把用户自动转换成多种时间周期的数据,使得用户的各种分析都是具有时间趋势的智能分析。



7.OLAP 体系结构

一个 OLAP 体系结构是由前端的 OLAP 专用开发工具如 LightShip Professional 或其它语言的开发工具,如

Mircosoft 的 Excel、VB 等,客户端的分析工具和服务器上的数据处理部分组成。如上图所示:

二、如何用 OLAP 建立一个商务智能系统 BI(Business Intelligence)

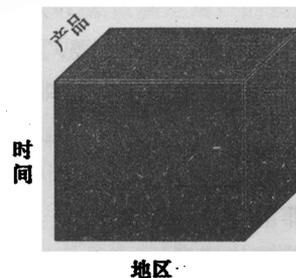
- 从用户原有的事物处理的应用系统中获取大量的用户数据
- 建立一个商务多维数据型
- 按预算、预测、趋势、去年、平均等制定计划
- 发现异常情况报告
- 用切割、钻透、寻找等方法来分析问题、发现问题

多维数据库基于 OLAP 的关系,由于 OLAP 在开发 DSS 系统上比传统方法最具优势的地方,即极强的综合大量数据的能力以及非常灵活的随机查询分析功能,都来源于多维数据库本身,因此正如 OLAP 依赖于关系数据库一样,OLAP 也几乎成了多维数据库技术的同义词,同时 OLAP 也使用多维数据驱动和自动综合技术达到可维护性。

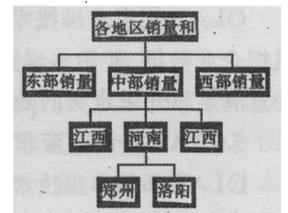
多维数据库技术具有以下基本技术特性:

- 多维数据 数据项之间存在着一对多和多对多关系的数据。
- 维中层次关系 在多维数据库中,每个维可建立多个 Hierarchies 的层次关系,由于具有这种层次关系,我们可以用“深入访问”的方法很方便的在各层级的详细数据之间漫游,这意味着用户只需点击屏幕上的一个数据项,系统就会自动调出下一层级的数据。

维的层次



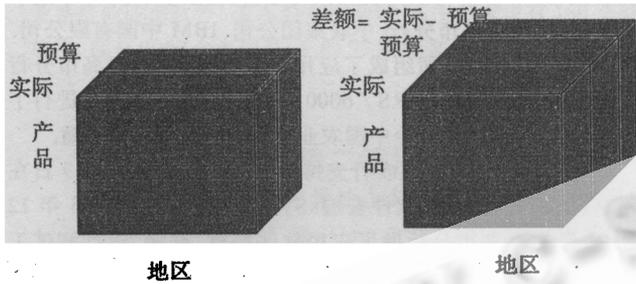
地区维的层次



- 向量运算 把数据组组成为阵列运算比把同样数据组成为关系表要快,而且容易综合。向量运算可把数据库中的数据按整个平面进行数学运算,在 LightShip 的

多维 OLAP 数据库服务中,上述向量运算可用一步操作来完成,而关系数据库不可能用一步操作完成平面间运算。因此在运算时间上,关系数据要比 OLAP 数据库多几个数量级。

向量运算



• **综合速度** 一个典型的关系数据库每秒能搜索几百条记录, 而一个典型的多维数据库可以每秒 20,000 到 30,000 个数据单元(相当于关系表中的记录)的综合速度, 甚至更快。

• **旋转数据立方体** 在一个多维数据库组织结构里, 一个理想的分析过程需要某维可与其它进行任意的组合, 也就是能够“旋转”你的数据立方体的视图。一个 n 维阵列就有 n(n-1)个视图, “旋转数据立方体”功能是多维数据报表的主要技术, 有时也称为“面和块切割”。

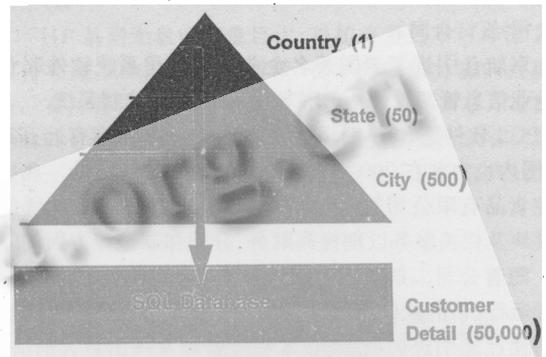
• **时间序列数据类型** 时间可能是多维数据库里最普遍的一维, 在 Pilot 公司的 LightShip 产品里, 不把时间作为一维, 而是采用时间序列数据类型, 即在每一个单元里, 都存储下诸如 10 年里每天的数据, 这就是时序数据类型思想, 这样做大大简化了应用开发和数据存储, 从开发人员角度来看, 时间序列的使用是多维 OLAP 产品中最大的一项时间节约措施。并不是每一个 OLAP 产品都具有时间序列数据的, Pilot 公司的 LightShip 是工业界第一个具有时间序列数据类型的产品。

• **稀疏数据** 在多维数据库的维数增加时, 其数据单元为空或零值的情况将增加, 即数据稀疏性。LightShip 数据库服务器支持稀疏数据处理, 能略过零元缺失的数据和重复的字符串。

• **钻进关系数据库** LightShip 多维数据库服务器, 支持穿过多维数据库低层去访问关系数据库。

• **安全性** LightShip 多维数据库服务器, 能保证和控制未授权用户对数据的访问, 以及在数据库备份恢复和维护方面的数据完整性。

“钻进”关系数据库(SQL)



“MIDSQL”多维查询语言

就象关系数据库具有结构化查询语言一样, 多维数据库也有自己的多维查询语言 MIDSQL, 类似于关系型 SQL 语言, 关系数据库的 SQL 是以字段为匹配查询的, 多维 SQL 是以维来匹配查询的。

三、结论

OLAP 数据库服务器技术的精华是快速、灵活的数据统计分析, 在今后一段时间内, 关系数据库将继续统治 OLAP 在线事务处理领域(因为需要逐条处理记录), 而 OLAP 服务器技术则更适合于客户 / 服务器模式下的数据报表和智能管理决策分析系统。高效和灵活的数据分析需要用多种方法统计数据的能力和按时间观察事务发展趋势的能力。OLAP 服务器可以和关系数据库和睦相处, 共同创造一个客户 / 服务器环境, 使用户能够按照所需要的形式迅速地得到数据。

• 读者来信摘登

<< 计算机系统应用 >> 杂志面向广大中高层开发人员和计算机工作者, 自发行以来, 就得到了社会各界的承认。以其实用, 印刷精美占用广告幅面少、经济等优点受到有识之士青睐。我在学校休息时间, 经常在电脑协会、自动化系或计算机系借阅此刊, 我们现在正在为一家建筑公司 < 国家三级企业 > 开发应用软件, 由于大量的数据分析, 我经常依靠贵刊登载的优秀文章中所介绍的方法试作。希望贵刊以后能沿着这些宗旨继续走下去。

(洛阳 陈新勇)