

证券经纪业务数据仓库设计

朱晓武

(武汉大学计算机学院 430072)

随着证券市场的日益规范, 券商之间的竞争加剧。印花税的下调和佣金的浮动制使得券商经纪业务的竞争愈发激烈, 因而提高客户服务质量, 进行集中有效的管理监控, 提供及时准确的经营情况分析 and 趋势预测变得更加重要。券商在多年的经营过程中积累了大量的企业经营、客户资料、交易记录等数据, 利用这些数据进行决策分析是当务之急。可是实际情况是券商下属的营业部存在各个时期针对各个应用编制的不同的柜面系统, 运行在不同的平台之上采用不同的数据库管理系统 (DBMS) 和数据结构。要生成一个汇总的、跨越多个平台的综合统计报表非常困难, 通常是手工加上计算机处理, 不仅慢而且不准确。一些历史数据以光盘保存, 与现有系统完全脱离, 严重的造成数据资源浪费。券商对这些数据也无从下手分析于是对自身的经营状况越来越不了解, 对于加强管理和提高服务质量更是纸上谈兵。所以券商都在考虑向集中交易的方向转变, 建立由总部直接处理交易的大系统和相应的数据仓库。这种大型的集中式系统的确定为数据的集中管理提供了便利, 但是它所带来的风险也是很大的, 国内外证券信息技术工程师对“集中”还是“不集中”进行了大量的讨论。笔者根据当前国内证券业的现状, 提出了“风险分散, 数据集中”的数据仓库设计思想, 并进行实施取得良好的效果。



摘要: 本文针对券商经纪业务管理总部的需求, 根据“风险分散, 数据集中”的设计思想设计了相应的数据仓库, 着重介绍了数据仓库的网络结构、功能结构、模块实现, 并在此基础上提出了数据仓库的一些新的应用。

关键词: 数据仓库 证券 联机分析 数据挖掘 SQL Server2000

1 证券经纪业务数据仓库“总部—营业部模式”的网络结构

采用相同的平台, 把各个营业部的柜面系统统一。如现有较为常用的第三方软件商恒生、新利提供的柜面系统。营业部的柜面系统更多的要考虑导入历史数据的完整性, 更换后能把所有的历史数据导入。券商业务管理总部建立数据仓库, 使用数据仓库服务器双机热备份并设有数据预处理机对数据进行预先的清洁处理, 通过DDN专线与营业部相连, 另外还有ISDN备份, 在DDN通讯线路意外中断时切换到ISDN备份, 保证数据传输不中断。在营业部建立数据采集的接口, 实时取得数据。营业部内部支持TCP/IP和IPX/SPX协议, 总部到营业部采用TCP/IP协议。这样保证了数据集中, 能够统一管理, 同时又使得系统风险分散到各个营业部, 不会因为一个故障造成全部瘫痪。应用数据仓库的网络与营业部原有的网络并不冲突, 可以共存于同一个网络环境中。整个网络结构如图1。

2 证券经纪业务数据仓库的设计

2.1 券商的经纪业务及相关的业务伙伴

券商为客户提供证券经纪业务服务, 包括: 证券的代理买卖; 代理证券的还本付息和分红送股; 证券的代保管; 代理证券登记开户; 提供行情揭示和各种委托手段 (热自助、刷卡、网上交易、电话委托等); 信息服务和投资咨询等。与之发生业务的主要伙伴有证券交易所、证券登记结算公司、上市公司、信息提供商。其中证券交易所是最重要的合作伙伴, 为券商提供买卖交易、行情数据、上市公司的业绩信息等服务。券商必须在保证客户正常交易的前提下, 通过对客户属性特征、交易行为和资金信息的分析, 对客户进行细分, 有针对性的对客户实施One to one的服务, 提高客户服务质量, 保留老客户, 开拓新客户; 另一方面, 必须对现有市场信息进行全面、仔细的分析, 把握证券市场的走势, 为客户提供准确的咨询信息, 尽可能避免风险, 获得收益。

2.2 数据仓库的数据源及特点

证券公司的数据主要包括以下几大类: 证券相关的数据; 营业部的交易结算数据; 营业部的客户信息数据; 企业内部的财务数据。数据源为交易所、营业部 (交易系统、柜面系统、财务系统)、信息商和证券登记结算公司。其特点是: 数据处理流量大; 实时性强; 安全要求高。

2.3 数据仓库的系统功能结构

根据券商对系统的应用需求, 整个系统从功能的角度可以分为四个层次: 数据采集; 数据管理; 分析处理; 信息服务。系统功能结构图2。

(1) 数据获取: 从源数据库中获取交易数据、客户数据、财务数据和市场数据, 并进行清洁传输, 加入到总部数据仓库中, 然后采取增量更新的方法, 取得数据源中发生变化的数据。

(2) 数据管理: 建立以客户、行情、财务等为主题的多维数据模型 (MOLAP), 对预计算的汇总数据快速索引, 建立渐增的更新机制, 即在增加新的数据到数据仓库时, 必须保护已有数据。

(3) 分析处理：进行指标分析、数据挖掘(Data Mining)，主要有客户分析、忠诚度分析、贡献率分析、证券分析、财务分析、经营状况分析。例如客户分析，可以结合地理分布信息，按照客户的操作习惯和距离营业部的远近，为其提供相应的现场交易和非现场交易。

(4) 信息服务：主要包括风险预警、个性化服务、信息咨询、配股分红提示等功能。

2.4 数据仓库的模块实现

在 Windows2000 server 操作系统下，采用 Microsoft SQL Server2000 和 SQL Analysis Services2000 实现中间数据库和数据仓库，该系统的模块如图 3。

(1) 数据采集模块：由于数据源的分布、异构、自治的特性，所以数据采集模块采用多种数据访问技术，定义数据仓库的数据规则，对数据源进行界定。不同的数据库系统，设置不同的源数据库的副本，实现跨平台、跨地区的从业务系统中抽取数据，在第一次全部导入数据仓库后，增量更新。在采集模块中，还有数据清洗，记录字段的重组、增补与原数据的一致性与完整性的检查，例如增加一些时间属性，增加后的数据必须单向传送到数据仓库，而不能回传原数据库。

(2) 中间数据库模块：采用 SQL Server2000 数据库管理系统作为中间DBMS，必须明确数据仓库的主题，全面考虑在构建数据仓库和数据立方体时所需的数据信息，设计所需数据与原数据库的关联关系，在保证提供数据仓库所需数据信息的前提下，尽量避免未更新数据的多次传输，减少营业部数据库的处理时间，以免影响正常交易。

(3) 数据传输模块：确定数据装入的次序，清除无效或错误的数据，对数据刷新。采用 Microsoft SQL Server2000 提供的数据库转换服务(DTS)实现。该服务提供 Oracle、Informix 等大型数据库，可通过 ODBC、OLEDB 接口，实现异种数据库同 SQL Server2000 之间的数据传输转换。通过 SQL Server2000 的数据转换服务建立数据仓库增量更新包，周期性的运行该增量更新包。方法一使用 DSO 控件编程周期性的调用；方法二设置调度包的属性周期。

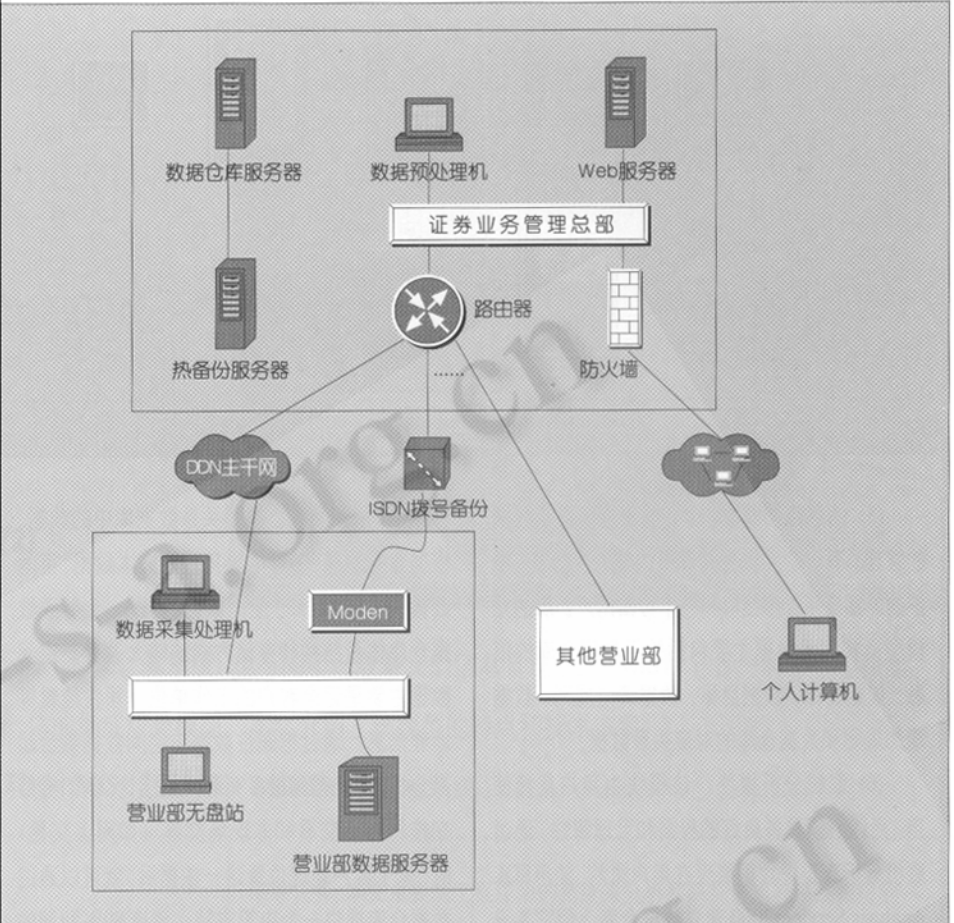


图1 券商数据仓库系统网络结构

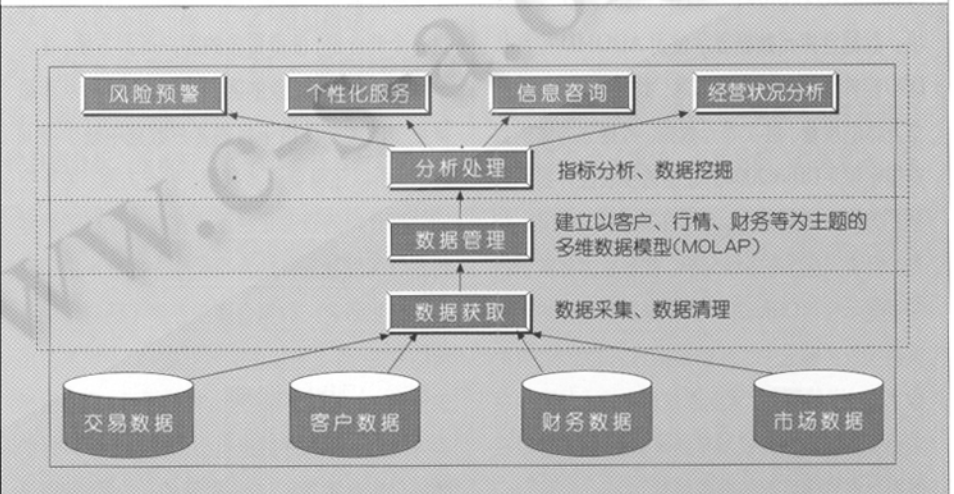


图2 系统功能结构图

(4) 数据仓库模块：使用星型构架来构建数据模型。采用 SQL Analysis Services2000 建立数据仓库，针对客户、交易、资金、行情、营业部经营状况等主题进行设计，如客户分析，主要从不同的角度为管理人员提供客户群体和重点客户

的信息，揭示客户群的行为特征，同时动态的反映客户的状态，评估客户的价值。这一主题从不同的角度细分，可以分为客户分类分析、客户交易分析、客户帐户分析、客户价值分析、客户地理分布分析。

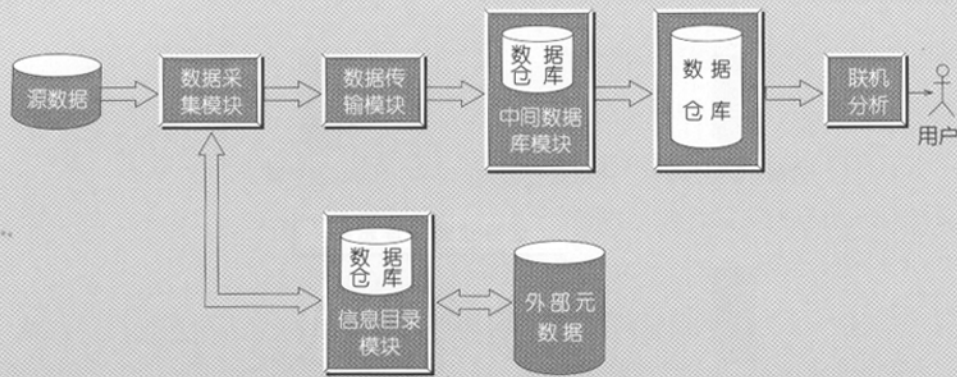


图3 数据仓库模型图

(5) 信息目录模块: 这一模块主要是元数据的管理, 有一个元数据管理员的输入、输出、维护的外部应用程序接口和一个可由用户扩展的元模型。技术元数据主要包括关于数据仓库的信息, 如数据源、数据目标、数据清洁规则、数据增加规则和数据库的对应关系数据。

(6) 联机分析模块: 该模块由用户直接使用, 所以必须具备良好的界面和可理解性, 通过联机分析真正把数据展现在用户面前。使用SQL Analysis Services2000建立数据库, 但它并没有提供良好的界面及分析程序, 所以必须采用Client/Server和Browser/Server相结合, 实现该模块。数据库元数据显示使用ADOMD控件和DSO控件显示数据库信息、数据立方体、数据立方体的度量、层次等元数据。数据立方体的数据显示使用Office2000的数据透视表控件显示数据立方体对应的数据, 并可在其上实现钻取、切片、切块、过滤等多维分析。数据立方体数据的图表显示使用Office2000的数据图表控件显示数据透视表中被选中区域的数据对应的图表, 可为柱状图、饼状图等。营业部信息咨询人员可以将数据查询以网页方式发布到网站上, 客户可以在网页上看到个性化的服务信息。

3 证券经纪业务数据仓库的应用拓展

设计和建设证券经纪业务的数据仓库的过程是一个动态的过程, 它必须随着业务的更新而变化。所以在设计时首先要考虑到系统的可扩展性。数据仓库的应用拓展主要有以下几个方面:

(1) 数据源的多样性: 在上述构建的数据

仓库中, 所涉及的主要是关系数据库中的数据。根据业务的需求, 可能要拓展到文本数据、空间数据、时序数据库。空间数据的引入, 会导致数据仓库体系结构和存储方式的较大改变。空间数据主要是记录客户的工作单位、居住位置等地理信息, 通过把属性数据与空间数据结合起来分析, 可以得到很多有意义的结论。例如券商在推行现场交易和非现场交易(如网上交易)时, 有了客户地理位置这一依据后, 就可以通过数据仓库提供一个决策支持, 创造80%利润的客户居住地离营业部较远, 可为他们提供网上交易的服务; 或者在一个城市合适的地方建立一个营业部, 使得更多的客户得到方便, 在这一基础上, 提供细致的客户个性化服务, 从而赢得更多的客户。由于目前客户数据中缺少空间数据, 所以还需要逐步获取。另一方面, 关于空间数据仓库的研究正在逐步完善, 怎样把地理数据与属性数据存放在同一个数据库中还需要进一步研究。数据仓库中存储了证券市场行情的数据, 涉及到较多的时间戳, 对于股票价格的预测, 会用到趋势分析、相似搜索、周期分析等方法, 这种时序数据也必须提供一个较好的数据仓库存储方案, 定义时间的多粒度, 使得分析性能最优。

(2) 联机分析及数据挖掘: 仅仅只做联机分析还是不够的, 要提供给客户和业务管理总部更多的有用信息, 还必须用到数据挖掘。采用深层次的分析, 要使用一些数据挖掘工具。所以数据仓库必须为数据挖掘工具提供良好的接口, 采取无耦合、松耦合、半松耦合、紧耦合等几种方式

之一。使用MDX (Multidimensional Expressions) 建立自己的应用, 进行关联分析、聚类分析、孤立点分析、演变分析。

(3) 建立在证券经纪业务数据仓库之上的应用: 可以借助数据仓库提供的大量信息, 建立一个较为完备的CRM和ERP系统, 协助券商进行管理和客户服务, 提高自身的竞争力。数据仓库系统还可以与企业人力资源管理系统相连, 从客户的反馈中得出最受欢迎的股评师、服务人员等, 从而发现公司中的优秀人才。

4 结束语

数据仓库是一种新兴的数据管理和分析技术, 在银行、电信业已经得到了应用, 但在证券业的应用还不是很多, 其主要原因就是证券业务的特殊性, 就证券市场发展来看, 应用数据仓库势在必行, 本文提出的设计紧密与券商的经纪业务实际情况相结合, 由“业务”驱动“设计”避免了数据仓库流于形式, 无法使用。该数据仓库的实施避免了“集中交易”的风险和“分散交易”的弊病, 达到了“风险分散, 数据集中”的效果, 因此具有很强的现实意义。随着数据仓库技术的不断完善, 券商会建立一个更加完备的数据仓库, 从而有效的提高服务质量, 挖掘优秀客户, 加强管理和提高工作效率, 使其在证券市场中具备更强的竞争力。

参考文献

- 1 冯玉才, 数据库系统基础(第二版), 华中理工大学出版社, 1993。
- 2 张维明, 数据仓库原理与应用, 电子工业出版社, 2002。
- 3 Rob Hawthorne, Microsoft SQL Server 2000 Database Development from Scratch, 人民邮电出版社, 2001。
- 4 Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, 机械工业出版社, 2001。
- 5 Carl Nolan, Introduction to Multidimensional Expressions(MDX), Microsoft Corporation, 1999。
- 6 杰夫里.C.胡克, 华尔街证券分析, 经济科学出版社, 1999。