

数据仓库辅助中央银行决策支持系统的构想

杨丽南 胡昌铸 (陕西财经学院 710061)

摘要:数据仓库是适应工商企业经营决策的需求,在数据库的基础上发展起来的信息技术,具有广阔的应用前景。本文对数据仓库在金融领域的应用,开发设计支持中央银行决策的数据仓库提出一些见解,与同行探讨。

关键词:数据仓库 决策支持系统

一、数据仓库一般理论

企业的营运系统每天产生大量的数据,有价值的信息就蕴藏其中,关键是采用合适的工具将其挖掘出来,数据仓库正是在这种背景下应运而生。

按照 W. H. Inmon 对数据仓库的定义,数据仓库是支持管理决策过程的,面向主题的,集成的,稳定的,不同时间的数据集。通俗地讲,数据仓库就是根据决策的需要按照不同的应用主题,如产品,供应商,客户等来组织数据,把分散在不同地域或不同部门有关该主题的信息集合在一起,经过格式转换,整理,加工等工作存储在数据仓库中,基本不再改变。然后采取适当的分析工具,如 OLAP,数据挖掘技术,用户查询工具等对数据仓库的数据进行相应操作,以辅助决策者进行决策。

1. 数据仓库的数据结构

数据仓库的数据结构由五块组成:历史性详细数据、当前详细数据、归纳总结信息、专业分析信息、数据仓库结构信息(元数据)。历史性详细数据和当前详细数据都存储的是原始信息,只是存储的时期不同,其中历史性详细数据可存储 5 至 10 年的信息,因此数据仓库的容量是很庞大的。归纳总结信息是在详细数据基础之上进行的粗加工,如按部门、产品、供应商等进行简单汇总,得到一些概要数据。专业分析信息是采用一些专用分析工具对其下层数据进行分析,如时间趋势分析,聚合分析,市场预测,数据挖掘等等,从而得到有助于决策的信息。最后一个是元数据,负责描述数据仓库的数据和环境。元数据又可分为管理元数据和用户元数据。管理元数据对“源数据及其内容,数据仓库主题,数据转换及各种操作信息的描述;用户元数据帮助用户查询信息,理解结果,了解数据仓库中的数据和组织”。元数据在整个数据仓库中占据核心位置,具有举足轻重的作用。

2. 数据仓库的数据组织方式

数据仓库的数据组织方式可分为虚拟存储方式、基于关系表的存储方式和多维数据库存储方式。虚拟存储方式是虚拟数据仓库的数据组织形式,它没有专门的数据仓库来存储数据,所有数据还是存放在源数据库当中,只有需要进行分析或查询之时,才依据一定的规则从源数据库中提取相应数据临时组成所需数据视图,以供用户分析或查询之用。这种方式花费少,较灵活,但对源数据要求较高,必须符合数据仓库对数据的格式、组织要求等。

基于关系表的存储方式是将数据仍存放在关系型数据库当中。在这些表间有维表和事实表之分,依据元数据的管理来构成多维数据库的形式。这种组织方式又可分为星型模型和雪花模型。星型模型是以事实表居中央,从每个维表中抽取维关键字构成事实表的关键字,因此事实表与维表是放射状连接。雪花模型是对星型模型的扩展,它为维表中增加了层次,使这个数据仓库的层次更加清晰,逻辑关系更加明白,但是同时增加了系统的复杂程度,降低了通用性。

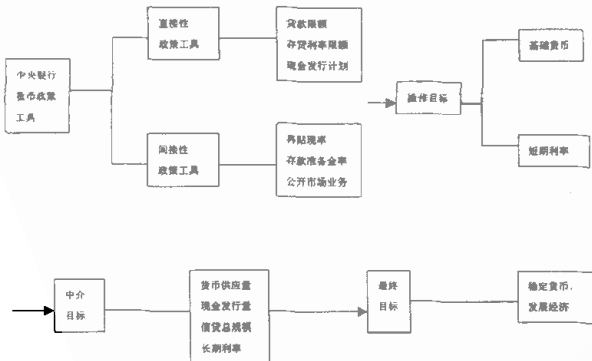
多维数据库存储方式是一种具体的数据仓库,按照预先定义的多维结构将数据组织起来。多维数据库可以简单地理解为将数据存放在一个 n 维数组当中,同时为了避免出现大量的稀疏矩阵,对维进行划分。那些必须存在的,对构建多维数据起关键作用的维称为稠密维,以它们为基础构建多维数据库,而那些只对部分数据具有存在意义的维称为稀疏维,并且只存储多维组合存在的数据,采用一个指针指向包括稀疏维的“多维体”。这样,就可有效解决稀疏矩阵问题。

3. 数据仓库和营运系统

数据仓库和营运系统在企业范围内有各自的地位和作用,不可相互取代。二者都可进行分析工作,但分析的侧重点不同,如下图。

	营运系统	数据仓库
商业分析	运作效率 ROI 和 ROA 资本管理	收入/利润增长 市场和顾客和增长 资产管理
金融分析	利润 应付(收)帐 运作参数	收入/利润 资产分析 产品周期 TQM
销售和市场营销	面向顾客 响应时间 重复的业务 维护与定点定货	面向市场 市场份额 分配渠道的策略 产品包装

数据仓库在制造业、零售业、银行业、航空业等都有广泛应用。利用数据仓库提供的多个应用程序可以支持一企业中的多种功能,如定点市场、交叉贸易、信息风险管理、商店布局、产品包装、供应商集成、保单管理等等。



二、中央银行数据仓库的应用

中央银行的政策、措施对整个国民经济都有重大影响。正因为如此,中央银行的举措也具备了很大的风险性。中央银行在进行宏观调控往往要通过有若干环节的传导机制。如下图,中央银行只修改最初的参数,其余的参数变化只能希望它们按预期发展,这就要求对传导机制中央参数的影响关系必须有准确、本质的了解和掌握;另一方面,中央银行操作的数据往往具有以小搏大的特征,如货币供应量。

$$M = mm * H$$

H: 高能货币(由中央银行控制)

M: 全社会货币供应量

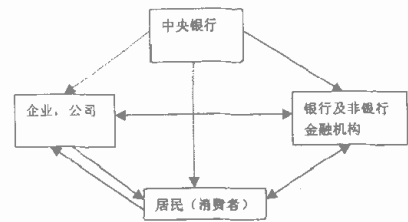
$$mm = (1 + cu) / (cu + re), \text{ 货币乘数}$$

cu: 公众现金 - 存款比率

re: 存款准备金率

通常 $mm > 1$, 如 $cu = 0.37, re = 0.10$, 则 $mm = 2.91$

所以在中央银行看来微小的变化可能会导致其他参数的剧烈变动,进而破坏了经济稳定,造成不利的经济波动,这就要求对经济生活中各种关系必须有较为精确的定量分析,才能将风险降到最低。那么,如何做到这一点呢?众所周知,经济生活就是由多个经济实体和管理机构相互作用构成的。如企业与企业,公司经济往来,企业,公司依赖银行进行资金往来,中央银行管理银行等等。可用下图简易表示:



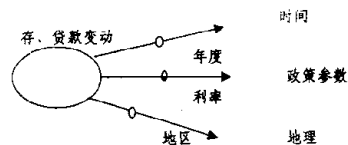
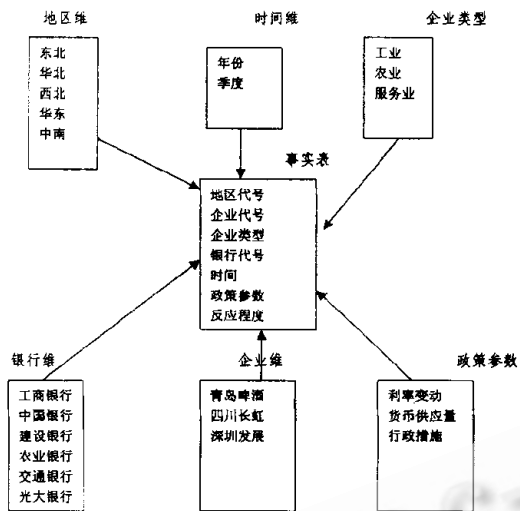
中央银行在制定政策前,必须深入了解各种实体之间的影响和制约关系,发现它们之间的传递参数,并尽可能进行准确的定量分析,以此作为决策制定的基础,而数据仓库在这一方面恰可以作出突出贡献。

依据前面对数据仓库的描述,商业企业数据仓库可以用来发现顾客购买模式,购买趋势。借鉴这种思想,中央银行可以为企业,公司,银行建立数据仓库,发现它们对政策的反应灵敏度,判别企业行为模式,银行行为模式,以此作为定量分析的基础,进而运行金融决策。

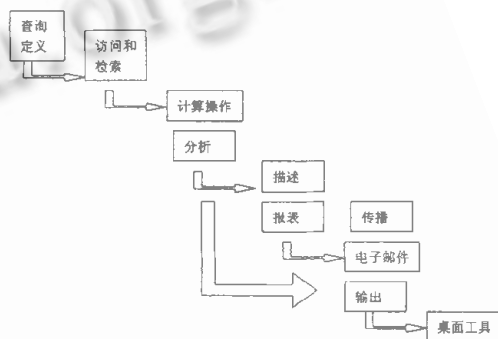
1. 中央银行数据仓库逻辑结构

中央银行数据仓库逻辑结构可由若干表组成:

(1) 可以采用在关系型数据库基础上增加维,构成星型模式或雪花模式。可分为两种表:事实表和维表。事实表可用来描述企业、公司、银行对政策参数的反应程度,维则选取对宏观调控有影响、有意义的划分。如可以选取全国若干有代表性的地区,在每个地区内选取有代表性的工业,农业,服务业等类型企业,参考模型如下:

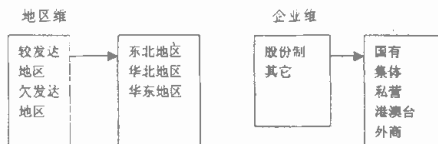


复杂查询必须经过以下过程:正式定义查询(用数据仓库可以识别的语言)、访问和检索数据、计算、操作和分析数据,最后显示结果。过程如图:



反应程度可依据不同的情况进行不同的定义。如,描述企业对利率的反应程度,可用利率变动引起的贷款变动;进出口企业存贷款及收益变化;银行可采用各类存款、贷款变动。

或者,还可采用雪花模式,对维增加层次。如下图:



地区维可依据中央银行的总行与分支机构设立的特点,如按照即将建立的区域银行的划分来建立地区维。

(2)建立企业和银行详细信息表。相应地,也可为银行设置详细信息表,如有必要可设置到分行一级。

(3)宏观经济、金融时间序列库。在该时间序列库中,可容纳宏观调控常用的指标,如工业生产总产值、农业生产总产值、进出口总额、各层次货币供应量、信贷规模等等。

(4)宏观经济文本库。在该库中容纳一些对国民经济有重大影响的的政策或举措。

2. 中央银行数据仓库的功能

中央银行数据仓库主要提供复杂查询,预测,模拟,鉴别模式。

(1)支持复杂查询。如在发达地区哪些企业在第三季度对利率调整反应较为灵敏;西北地区工商银行在98年份对降息引起的存款、贷款变动。在进行查询时,首先确定主题域,选择查询维,构成一个查询面,如:

(2)鉴别企业行为模式和银行行为模式。现在有许多制造商提供数据仓库解决方案,从数据仓库诸多表中发现企业行为模式和银行行为模式,如哪些企业对宏观调控反应灵敏,是按国家预期方向变动还是相反。又如银行行为模式,国家对贴现率、存款准备金率进行调整,银行的存、贷款变动如何,变动幅度是否与中央银行预测的基本相符,变动过程如何,是滞与中央银行所掌握的一致。鉴别模式用来发现中央银行与商业银行,与企业的直接或间接影响、制约关系,传导机制,为定性分析与定量分析完美的结合打下基础。

(3)预测和模拟功能。在前两种功能的基础上,辅助于数学模型,一方面可以预测企业或银行的变化趋势,可能会采取的措施,依据此可以有的放矢的采取相应的大调或微调措施。

参考文献

- [1] 赖福军、周婷:《数据仓库及其相关技术》,《软件世界》1997年2月
- [2] [美]Harjinder S. GILL等著《数据仓库—客户/服务器计算指南》,清华大学出版社1997年
- [3] 梁小民《高级宏观经济学教程》北京大学出版社

(来稿时间:1998年11月)